WORKING PAPER

Understanding Medical and
Demographic Trends with MEDDAS

Martin Rusnak
Sergei Scherbov

April 1988
WP-88-23

IIASA
International Institute
for Applied Systems Analysis

Understanding Medical and Demographic Trends with MEDDAS

Martin Rusnak
Sergei Scherbov

Paper presented at the WHO Consultation on Efficacy and Effectiveness of the Integrated Programme for Community Health in Noncommunicable Diseases, held in Geneva, December 14-17, 1987.

Foreword

The rapid growth in the availability of microcomputers, and the vast and growing pool of available medical-demographic data, together underscore the usefulness of software tools which make the analyst's job easier. MEDDAS is one in a series of software packages designed to aid in one or more aspects of demographic research. Although the program development was motivated by applications in standard demographic time-series, and in health and disease statistics, the program could prove useful in other applications as well.

Douglas Wolf
Deputy Leader
Population Program

Acknowledgements

Contents

Understanding Medical and
Demographic Trends with MEDDAS

Martin Rusnak*, Sergei Scherbov**

INTRODUCTION

When the World Health Organization announced its strategy under the slogan "Health for all up to the year 2000", scientists and health care managers tried to develop efforts for fulfilling it. Among the most challenging factors to be faced are:

- a projected increase in the proportion of elderly persons in the populations of developed countries;

- technological innovations which will lead towards improvement in treatment and increase in productivity;

- growth in costs of all kinds of health care services.

As the economic pressure on health care systems is substantial and growing from year to year, the question of budget allocation is becoming a top priority. Economic recession increases demands on other aspects of social care, such as antismoking campaigns, preventive investigations, and AIDS prevention. The growing proportion of elderly persons in the population also adds to the economic pressure by increasing the amounts paid out in old-age pensions.

These conflicting pressures—the desire of the decision makers to restrain government expenditure as well as the desire of health professionals and the public to defend and expand their health services—may lead to a period of instability and uncertainty in health care. Planning based on the understanding of current trends and on projections of the likely impact of options will probably be of substantial help in reducing this uncertainty and instability.

However, standard health statistics are of a static type, which do not represent trends over time. Without understanding the dynamics of health-related data, one may not be able to make rational decisions in such a complex system as health care. More-

*Research Institute for Medical Bionics, Jedlova 6, 883 46 Bratislava, Czechoslovakia
**Population Program, IIASA, A-2361 Laxenburg, Austria

over, the combined effects of health and various demographic factors (as mentioned above) in a population complicates the overall process. A promising way out of this situation is the use of computers in processing health and demographic data. Unfortunately this is limited by the lack of available software. There are many different packages treating one or another side of the problem. However, a complex, integrated, user-friendly package for medical and demographical data processing is hard to find. The number of papers on different time-series analysis on health care data and the attention they attract, as well as a book from the World Health Organization[1] on different health-related projections has supported the design of MEDDAS.

Based on our experiences with data processing in medicine and demography, a first version of MEDDAS (**ME**dical and **D**emographic **D**ata **A**nalytical **S**ystem) has been prepared. It consists of three parts: data retrieval, estimation, and charting routines. The structure of data files allows the storing of time-series and retrieving of various formats of data for processing. The estimation routines are entirely devoted toward smoothing and extrapolation. Several different methods have been employed in this part of the package.

From a user's point of view, MEDDAS plays a crucial role in its charting capabilities. Medical doctors as well as decision makers may dislike reading statistical tables, but they can easily and quickly understand and interpret different types of curves. That is why we consider charting to be so important in human-machine interface and why we pay much attention to it.

There should be no doubt why a PC was chosen as a computer for this particular system implementation. The first motive was to put all the systems on a medical, professional, or decision maker's table. The graphics possibilities were a second reason. The availability of supporting software was another reason.

The examples attached demonstrate MEDDAS's ability to treat cause-specific mortality data, although it might be used for other time-series as well, for example hospital beds, number of physicians, newborns, and so on. The range of application areas seems to be quite wide.

---

[1]World Health Organization (1986) *Health Projections in Europe: Methods and Applications.* Copenhagen: WHO Regional Office for Europe. 306 pages.

AN OVERVIEW OF MEDDAS

As was already mentioned the acronym MEDDAS stands for MEdical and Demographic Data Analytical System. The words medical and demographic were chosen because of the main aim—application in medicine and demography. However, only a few changes (if any) are required to use it with other data as well. The only condition to be fulfilled is the form of data: the time-series. The general structure of MEDDAS is displayed in Figure 1.
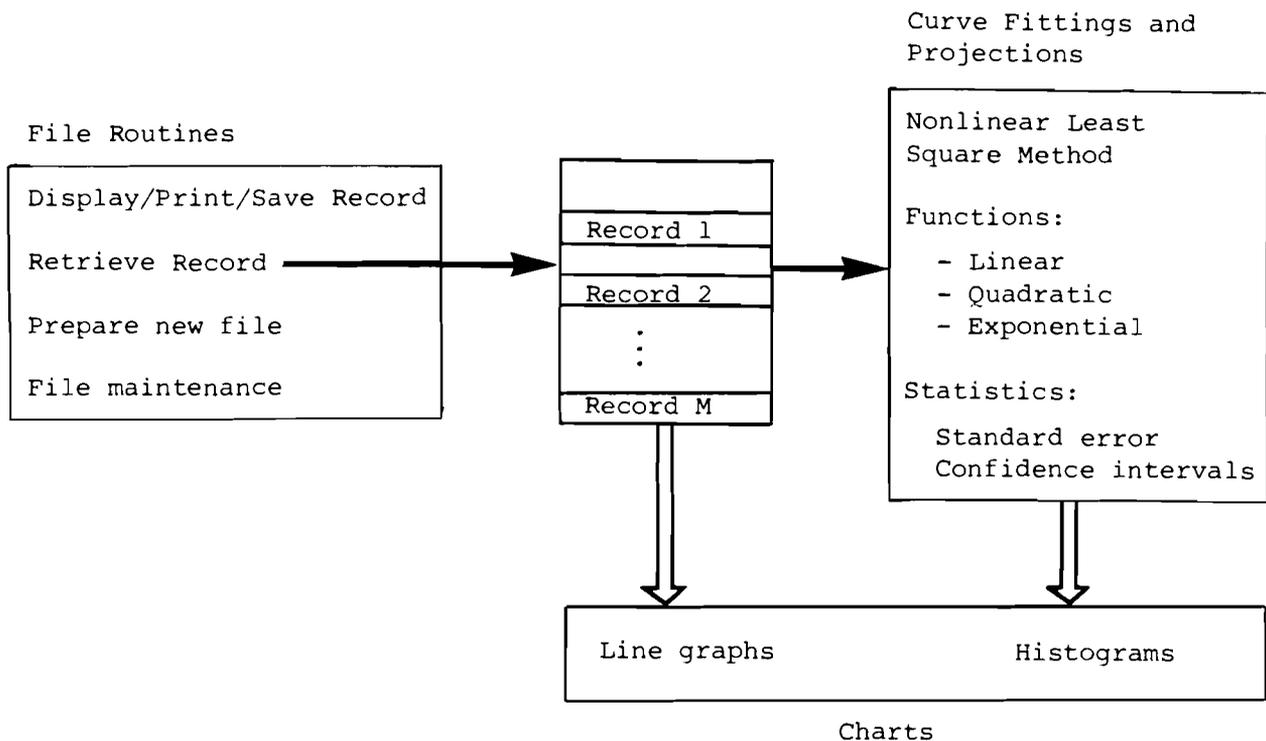
Figure 1. General structure of MEDDAS.

The overall design of the system allows for its interactive use by someone with only a basic understanding of computers. Extensive knowledge of the data being processed is essential for high quality results. Some knowledge of estimation and extrapolation techniques is required as well, as an aid to interpreting the data. Based on charts produced by the system, a notion of trends can be acquired. Basic statistics and confidence intervals are supplied as well. MEDDAS is designed for the implementation on personal computers. It makes use of various features supported by IBM PC computers and compatibles. It supports different types of graphic boards. It requires at least 512k bytes of RAM memory, and the availability of a hard disk is strongly recommended.

Data Maintenance

This part of MEDDAS allows for storing, retrieving, and editing of time-series data.

Data structure. The general structure of one record is comprised of two parts: data identification and the data itself. The first part contains 11 fields of either character or numeric type. The user is free to assign these fields according to his desire, except for three which are mandatory: FIRSTYEAR, STEP, and ENTRIES. They are essential for the time-series description. The first one, as the name suggests, stores the first year (or any time unit) of a data set. The second one records the number of years (or other time units) between two consecutive data fields. The ENTRIES field stores the total number of values being kept in one time-series. Examples on different types of assignments are in Figure 2.

First example

COMMENT	SEX	COUNTRY	TYPE	ICD	AGE
10 CHAR I	6 CHAR I	4 INTEGER I	4 INTEGER I	4 INTEGER I	4 INTEGER

SCENARIO	ENTRIES	FIRSTYEAR	STEP	DATATYPE	
I INTEGER 4 I	INTEGER 2 I	INTEGER 4 I	INTEGER 1 I	INTEGER 1 I	

Second example

COMMENT	SEX	COUNTRY	AGE	ICD	DESCRIPT 1
10 CHAR I	6 CHAR I	INTEGER 4 I	INTEGER 4 I	INTEGER 4 I	INTEGER 4

DESCRIPT 2	ENTRIES	FIRSTYEAR	STEP	DESCRIPT 3	
I INTEGER 4 I	INTEGER 2 I	INTEGER 4 I	INTEGER 1 I	INTEGER 1 I	

Figure 2.   Examples of different identification fields.

The second part of a record consists of fields with time-series data. The maximal format used for each of these fields is a 19–digit number with up to 15 decimal places. The maximal number of data fields depends on the number of identification fields and the total number of fields allowed by the system—128 fields. The number of data fields is:

$$\text{max number of data fields} = 128 - \text{number of ID fields} \quad .$$

If there is a necessity to use more data than is allowed in one file, the data should be split into two or more files. The maximum number allowed in the system should be less then 10^{35}. By convenience the number 10^{35} is used to indicate missing values throughout the system. The process of data entry is described below.

The part of the file with identification usually contains codes for various items. Their meaning is stored in the file CATALOG. It sums all the codes together with their interpretation and labels for easier orientation. Any integer number with a maximum of five digits could be a code. The label is one character code, where 'a' stands for age, 'c' for country, and 'd' for disease code by ICD (International Classification of Diseases). The system is based on the tenth revision of the ICD. Rates are labeled with 'r' and type of data with 'i' (for index). When specifying the conditions for data to be retrieved, it is necessary to enter the precise codes. The code values are displayed in Table 1 for age codes, Table 2 for country codes, Table 3 for ICD codes, Table 4 for rate codes, and Table 5 for index codes.

Table 1.   Example of codes for age groups.

CODE	LABEL	LEGEND
46	a	0–4
47	a	5–9
48	a	10–14
49	a	15–19
50	a	20–24
51	a	25–29
52	a	30–34
53	a	35–39
54	a	40–44
55	a	45–49
56	a	50–54
57	a	55–59
58	a	60–64
59	a	65–69
60	a	70–74
61	a	75–79
62	a	80+

The process of retrieving data based on menus ends with the creation of an ASCII file. The expression 'communication file' is used throughout this paper for this particular file. It serves as intermediate storage for later use by various programs and contains all the information necessary for graphic or estimation routines. The advantage of using a

Table 2. Example of codes for selected countries.

CODE	LABEL	COMMENT	CODE	LABEL	COMMENT
1	c	world total	23	c	sweden
2	c	more developed regions	24	c	united kingdom
3	c	less developed regions	25	c	south europe
4	c	canada	26	c	albania
5	c	usa	27	c	greece
6	c	china	28	c	italy
7	c	japan	29	c	malta
8	c	india	30	c	portugal
9	c	europe	31	c	spain
10	c	east europe	32	c	yugoslavia
11	c	bulgaria	33	c	west europe
12	c	cssr	34	c	austria
13	c	german democratic republic	35	c	belgium
14	c	hungary	36	c	france
15	c	poland	37	c	german federal republic
16	c	romania	38	c	luxembourg
17	c	north europe	39	c	netherlands
18	c	denmark	40	c	switzerland
19	c	finland	41	c	australia & new zealand
20	c	iceland	42	c	australia
21	c	ireland	43	c	new zealand
22	c	norway	44	c	ussr

file of this kind is that the user does not need to search through the whole file each time he wants to make several different analyses of the same data. After the session is over, he is free to delete it or to save it for later use. Its structure can be seen in Table 6.

The number in the first row specifies the type of graphic card used. The most widespread types of cards are covered, including Enhanced Graphic Adapter and Hercules Monochrome Adapter. The title and subtitle of the chart of maximum 30 characters long are contained in the subsequent two rows. The quantity of different time-series retrieved is noted in the fourth row. The number of values in these time-series is stored next. Rows 6 and 7 contain a legend of x-axes (30 characters long) and values of x. The next two rows contain legend for y data and y data itself. Such pairs of y data could be repeated, but not more than six pairs. The number of pairs should be in accordance with the statement in the fourth row.

Table 3. Example of codes for disease categories.

CODE	LABEL	COMMENT
134	d	mal.neo.buccal cav.,pharynges,oesophagus,stomach
135	d	mal.neo.larynx,trachea,bronchi,lung
136	d	mal.neo all forms
137	d	rheumatic fever
138	d	heart dis.-chronic rheumatic
139	d	heart-arteriosclerotic & degenerative
140	d	hypertensive disease
141	d	cardiovascular disease
142	d	cardiovasc d.& vasc.lesions affecting CNS
143	d	bronchitis all forms
144	d	respiratory dis. all forms
145	d	stomach ulcer
146	d	duodenum ulcer
147	d	peptic ulcer
148	d	cirrhosis of liver
149	d	bones and joints-all diseases
150	d	accidents-motor vehicle
151	d	accidents-all forms
152	d	homicide and injury not in war
234	d	ulcer of stomach,duodenum and peptic

Curve Fitting

Having specified the communication file, the process of curve fitting begins with a selection of data to be processed. It is based on legends for y-axis data. To get a chance to see the data before estimation, the possibility of displaying them is added. Without knowledge of the data shape the next step will be hazardous. After having seen a chart with raw data, the appropriate function could be specified. The current menu contains three functions to be fitted: linear, quadratic, and exponential. The algorithm for curve fitting is based upon a derivative-free nonlinear least-square algorithm (Scherbov and Golubkov 1986).[2] Results of the effectiveness testing of this particular algorithm compared with other standard ones proved that it requires fewer function iterations, while precision is almost the same. This makes it advantageous for applications of the present sort.

The process of estimation can be started immediately afterwards. The user specifies from a menu how many time points he wants to process and the step of function calculation. Finally, the charting and printing of results begins. Two forms of result displays are available: 1) input data with a fitted curve, or 2) both of these curves with confidence intervals added. The title and subtitle, as well as x-axes legend are to be found in a

[2]Scherbov, S. and V. Golubkov (1986) *Derivative-free Gauss-Newton-like Algorithm for Parameter Estimation.* WP-86-63. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Table 4. Example of codes for rates.

CODE	LABEL	COMMENT
95	r	dependency
96	r	child/woman
97	r	sex
103	r	growth expectancy %
104	r	growth comp %
105	r	natural increase
106	r	birth crude
107	r	death crude
110	r	reproduction-gross
111	r	reproduction-net
112	r	fertility-total
113	r	fertility-gen
118	r	migrants-rate
119	r	fertility 15–19
120	r	fertility 20–24
121	r	fertility 25–29
122	r	fertility 30–34
123	r	fertility 35–39
124	r	fertility 40–44
125	r	fertility 45–49
126	r	sex-at birth
127	r	mortality infant

Table 5. Example of codes for index.

CODE	LABEL	COMMENT
45	i	population
83	i	population
84	i	population
85	i	population
99	i	population-urban
100	i	land area–1980
101	i	population average
108	i	birth number
109	i	death number
114	i	life expectancy
115	i	life expectancy
116	i	life expectancy
117	i	migrants-total
128	i	prevalence

chart. The legend appears on the next screen. The possibility of manual zooming is implemented as well. The user can adjust the chart according to his requirements. The results can be printed to a disk file as an ASCII file to be used with more sophisticated charting programs.

Table 6. Items in a communication file.

ROW	CONTENTS	FORMAT
1	type of terminal	INTEGER
2	title of chart	STRING
3	subtitle of chart	STRING
4	number of time series retrieved	INTEGER
5	number of time points	INTEGER
6	x - axes legend	STRING
7	x - axes data	INTEGER
8	y - axes legend	STRING
9	y - axis data	FLOAT
10	y - axis legend	STRING
11	y - axis data	FLOAT
12	:	
	:	
	:	

Except for estimation, there is the opportunity to prepare line charts of several curves on one chart. The markers enable the user to distinguish among them. The communication file for this type of chart is identical to the previous one. A histogram is another possibility for data display; the procedure for preparing one is identical with that just described.

TECHNICAL SUPPORT

MEDDAS was designed for IBM PC and compatible computers with an MS DOS operating system. XT or AT types are to be used with a minimum of 512 kbytes of RAM. A hard disk is not essential, but is recommended because a file's magnitude demands several floppy disks to be used for programs only. Graphics adapters of several different types are supported. Table 7 shows the adapter characteristics supported.

A hard copy of graphic results can be obtained using the Print Screen routine. To print graphics on a printer, AUTOEXEC.BAT should contain the DOS command GRAPHICS. Throughout MEDDAS extensive use of the ANSII device driver is used. If any problems with cursor control occur, CONFIG.SYS should be checked for ANSII.SYS installable device driver. [For exhaustive MS DOS description, consult the appropriate manual such as the following one: DOS 3.2 User's Reference (Microsoft 1986).][3]

---

[3]Microsoft (1986) *DOS 3.2 User's Reference*. Microsoft Corporation.

Table 7. Characteristics of adapters supported by MEDDAS.

ADAPTER	RESOLUTION	COLORS
Color Graphic (CGA)	320*200	4
Color Graphic (CGA)	640*200	2
Hercules	720*340	2
Enhanced Graphic (EGA)	320*200	16
Enhanced Graphic (EGA)	640*200	16
Enhanced Graphic (EGA)	640*350	2
Enhanced Graphic (EGA)	640*350	16

The system is based on routines written partly in dBASE III Plus programming language, compiled by the Clipper compiler into executable files. The programmer's convenience in using dBASE III Plus is paid for by increased memory requirements and disk space occupied by executable files, compared with FORTRAN or other compiler files. The second part of MEDDAS is entirely written in RM FORTRAN. The graphic routines were written in RM FORTRAN as well, with the help of NO LIMITS library. The user who would prefer to use a more sophisticated graphic package can prepare an ASCII file with the resulting data and feed them into his system.

If a hard disk is available, all programs and files must be saved in one directory. If only two floppy drives are available, one of them must contain a floppy with programs and the other a floppy with data. If any problems occur check for the following files on the program disk: ID.DAT, DIALOG.DIA, CATALOG.DBF, and MOD.DBF. In case of unintentional deletion of any one of these files, the system could start misbehaving.

A MEDDAS SESSION

Although the user is controlled throughout the MEDDAS operation by menus, a description of steps seems to be in order.

To invoke the program from the hard disk type 'MEDDAS' followed by 'ENTER'.[4] Before starting for the first time, check the current directory for presence of all files from distribution floppies. In case a hard disk is not available the main program can not be used and the user should key in the names of individual programs (see Table 8 for a list of the program names).

---

[4]On some keyboards the key for ENTER could be labeled 'CR' or 'RETURN'.

Table 8. List of MEDDAS executable files names.

FILE	ACTIVITY
MEDDAS	main
IMPORT	import data from ASCII file
DISPLAY	display selected data to screen,printer or disk
FILES	file maintenance routines
RETRIEVE	retrieve selected data and prepare communication file
WHOEST	fit curve and chart results
LINE	chart several curves
HISTOGRAM	chart histogram

The first menu contains a summary of possible actions (Figure 3).

```
Options:

     0. Return to DOS
     1. Import data from ASCII file
     2. Display/Print/Save records
     3. Retrieve records for processing
     4. Fit/Forecast time series
     5. Chart line graph
     6. Chart histogram
     7. Display catalog
     8. File maintenance routines


Enter option number:0
```

Figure 3. The main menu.

The first option in the menu is to exit from MEDDAS to DOS. A zero convention for exiting from the current menu to the previous one is applied throughout the system. Option 1 allows for importing data from the ASCII file. The option 2 calls procedures for displaying selected data to the screen or printer or saving them on the disk. Invoking option 3 the communication file containing retrieved data is to be defined. Option 4 performs curve fitting and forecasting, as well as the charting of results. Line charts of retrieved data could be obtained by the means of option 5. A histogram (option 6) is another possibility for data visualization. A list of CATALOG codes is displayed with option 7. The last option (8) covers a variety of file maintenance functions, such as delete, copy, or rename file.

Option 1

To input new data into MEDDAS data file this option is required. The procedure consists of several steps.

Record definition. One record consists of two parts: descriptors and data fields. There are 11 descriptors required for one record. Eight of them are user-defined and three are defined by the system itself (Figure 4).

DESCRIPTOR 1	DESCRIPTOR 2	....	DESCRIPTORS	ENTRIES	FIRSTYEAR	STEP	DATA1	DATA2	....

    User defined descriptors   System defined descriptors   Data fields
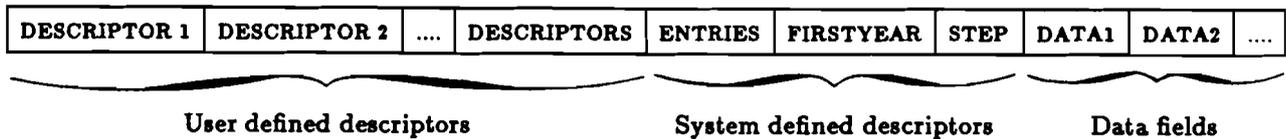
Figure 4. Structure of record in MEDDAS.

The program starts with descriptor fields definition. All eight descriptors must be defined, even when fewer are enough. In the latter case, dummy descriptors are used. The descriptor definition consists of name, type, and length specification. The name is to be fitted into the highlighted area on the screen. It must start with a letter—either lower or upper case letters are allowed. The descriptor might be either of a character or a numeric type. 'C' is entered to define a character type, 'N' stands for numeric one. The length in bytes is to be specified afterwards. When numeric fields with decimal digits are required, one byte for the decimal point must be added to the total number of digits. In this particular case a number of decimal places is to be specified as well. When dummy descriptors are required, the name specified must be 'DUMMY'. Then the field type will automatically be character and field length one byte.

When all eight descriptors are defined, the system creates three fields:

    ENTRIES – numeric type, length 3 bytes, no decimal places
    FIRSTYEAR – numeric type, length 4 bytes, no decimal places
    STEP   – numeric type, length 1 byte, no decimal places.

The contents of descriptive fields must be in agreement with codes defined in MEDDAS. If new codes are to be introduced, the MEDDAS catalog file must be updated; otherwise no meaningful help during query specification can be provided.

Data fields definition. A maximum of 117 data fields are currently allowed. They should all be of one type: numeric. The user has to specify the number of data entries for his current application. The field width in bytes must be specified afterwards. Decimal points are to be added as one byte to the total number of bytes. Finally, the number of decimal places should be specified as well.

After all queries are answered, an empty file could be created under the specified name. There is no need to specify the file extension, because the system automatically adds '.DBF' extension to the name.

Importing data. The data file structure is defined and it is possible to import data from the ASCII file. For this purpose, data are to be prepared in the form accepted by MEDDAS. The record structure in an ASCII file must agree with the MEDDAS record structure. Data are to be separated from one another by either a comma or a blank. When data are not available, two or more commas could appear in a sequence. When no data is used as a delimiter, no missing data are allowed. Otherwise, no correct transfer into MEDDAS could be guaranteed.

Assume the MEDDAS data file structure as shown in Figure 5; then, to import data from the ASCII file, the data organization must follow the structure of the MEDDAS data file (see Figure 6). When everything has been correctly prepared, data are imputed into the MEDDAS data file and the main menu appears.

Option 2

The program starts with several queries. The first one is the type of graphic adapter used. If Color Graphic Adapter (CGA) is used, no changes are required. CGA is considered a default by the system. If some other type of adapter is in use the exact type is to be specified. To select the default option (CGA) 'ENTER' is to be pressed. In any other case the number of the desired option is to be pressed. In this case 'ENTER' is not to be pressed afterwards.

The next screen displays a list of available files. The name of the desired file is to be keyed into the highlighted row on the screen. The default name extension in the data file is '.dbf'. The extension need not be specified if the file name has this extension. Keys DEL, INS, or keys with horizontal arrows are to be used for editing text inside the highlighted area. The use of wild cards or question marks is not supported. The name of a specified file is to be followed by 'ENTER'. The range of available data with the query for using all of them is displayed afterwards. In the highlighted square the default answer 'Y' is displayed. In case all the available data are to be displayed only 'ENTER' is to be

```
fdispl struct
Structure for database: C:cernay.dbf
Number of data records:      12
Date of last update    : 03/18/88
Field   Field Name   Type        Width      Dec
    1    TOWN         Character      10
    2    DISTRICT     Character       4
    3    TYPE         Numeric         4
    4    ICD          Numeric         4
    5    DUMMY5       Numeric         4
    6    DUMMY6       Numeric         4
    7    DUMMY7       Numeric         4
    8    ENTRIES      Numeric         2
    9    FIRSTYEAR    Numeric         4
   10    STEP         Numeric         1
   11    DUMMY11      Numeric         1
   12    V1           Numeric         8        2
   13    V2           Numeric         8        2
   14    V3           Numeric         8        2
   15    V4           Numeric         8        2
   16    V5           Numeric         8        2
   17    V6           Numeric         8        2
   18    V7           Numeric         8        2
   19    V8           Numeric         8        2
   20    V9           Numeric         8        2
   21    V10          Numeric         8        2
   22    V11          Numeric         8        2
   23    V12          Numeric         8        2
   24    V13          Numeric         8        2
   25    V14          Numeric         8        2
** Total **                         155
```

Figure 5.   Example of MEDDAS data file structure.

pressed. If the selection of fields is desired either upper or lower case 'N' is to be keyed in, not followed by 'ENTER'. A new screen is displayed immediately with field numbers followed by the value computed as $(STEP * \text{field number} - 1)$ and added to the FIRST-YEAR value. As for example the value of FIRSTYEAR is 1950, the numbers 1951, 1952, 1953, etc. appear on the screen up to the number specified by the field ENTRIES. Desired values are to be specified in the highlighted area using field numbers in three ways:

1)   'ENTER' stands for all available data fields;

2)   two field numbers separated by '-' standing for 'from- to';

3)   field numbers separated by commas or blanks enumerate desired ones.

An example in Figure 7 documents specification of data for years 1960 up to the year 1985.

The data fields are not necessarily described by years. Any other time-series specification is allowed, such as months, days, hours, or seconds. It takes just a few seconds for data fields to be initiated and a new screen to appear.

```
BANSKA B,SS,1,    ,2960,2836,7,1972,1, ,1E35,2847,2997,3021,3070,3059,3126,3087,2874,29
03,2886,2891,2960,2836
BANSKA B,SS,2,    ,    ,  ,7,1972,1, ,22.8,19.7,25.7,18.2,25.7,19.0,24.3,18.8,21.9,22
.0,21.5,20.8,10.0,21.5
BANSKA B,SS,3,    ,    ,  ,7,1972,1, ,15.1,15.8,16.0,11.6,21.8,11.8,18.9,11.0,16.4,13
.8,15.6,14.9,10.3,17.3
LIPT M  ,SS,1,    ,    ,  ,7,1972,1, ,12.3,15.5,15.0,12.5,13.5,10.0,16.9,17.4,16.9,19
.5,9.8,10.5,10.4,11.2
LIPT M  ,SS,2,    ,    ,  ,7,1972,1, ,20.4,18.9,20.0,16.3,18.7,17.7,21.9,21.8,22.6,24
.2,12.4,17.0,14.0,13.9
LIPT M  ,SS,3,    ,    ,  ,7,1972,1, ,12.3,15.5,15.0,12.5,13.5,10.0,16.9,17.4,16.9,19
.5,9.8,10.5,10.4,11.2
STREDSS ,CI,1,    ,    ,  ,7,1972,1, ,26889,28238,29467,29552,30354,30581,30704,30646
,28762,28461,28135,27826,27677,27565
STREDSS ,CI,2,    ,    ,  ,7,1972,1, ,22.1,21.8,21.4,21.6,22.3,21.6,22.7,20.2,20.9,20
.3,18.3,19.9,17.1,16.3
STREDSS ,CL,3,    ,    ,  ,7,1972,1, ,15.2,15.9,15.5,15.8,16.6,14.5,17.0,14.0,14.9,14
.3,12.7,13.5,11.2,12.5
BRATISL ,CI,1,    ,    ,  ,7,1972,1, ,5217,5735,6320,6667,7018,7343,7402,7413,7053,67
94,6646,6423,6390,6377
BRATISL ,CL,2,    ,    ,  ,7,1972,1, ,31.4,33.3,25.8,27.3,30.8,19.7,23.3,22.2,20.1,17
.1,18.7,16.3,16.6,16.2
BRATISL ,CL,3,    ,    ,  ,7,1972,1, ,24.2,27.2,21.2,22.3,26.8,17.2,18.2,16.4,13.6,14
.2,13.6,12.0,13.3,17.2
ZAPADOS ,CL,1,    ,    ,  ,7,1972,1, ,28480,30304,31801,31963,32240,32321,32059,32704
,30015,29560,29124,28681,27760,27232,28480,24.1,17.9
ZAPADOS ,CL,2,    ,    ,  ,7,1972,1, ,23.0,21.5,20.5,23.1,20.0,18.4,19.4,19.5,16.1,1
.0,13.6,9.7
ZAPADOS ,CL,3,    ,    ,  ,7,1972,1, ,14.2,17.6,16.3,15.2,18.0,14.8,14.0,14.5,13.8,12
.1,12.5,9.7,11.2,10.2
VYCHODS ,CL,1,    ,    ,  ,7,1972,1, ,27248,28530,29966,29454,30145,29509,29956,29975
,29270,28475,28713,28458,28071,28340
VYCHODS ,CL,2,    ,    ,  ,7,1972,1, ,28.8,27.9,24.0,28.6,26.8,23.5,24.4,22.9,22.5,22
.1,18.4,19.4,19.2,18.0
VYCHODS ,CL,3,    ,    ,  ,7,1972,1, ,18.6,17.2,16.3,19.5,17.1,15.0,16.2,13.9,13.2,13
.5,11.7,12.2,11.1,10.7
```

Figure 6.   Example of an ASCII data file to be imported into MEDDAS.

1	1950	2	1951	3	1952	4	1953	5	1954
6	1955	7	1956	8	1957	9	1958	10	1959
11	1960	12	1961	13	1962	14	1963	15	1964
16	1965	17	1966	18	1967	19	1968	20	1969
21	1970	22	1971	23	1972	24	1973	25	1974
26	1975	27	1976	28	1977	29	1978	30	1979
31	1980	32	1981	33	1982	34	1983	35	1984
36	1985	37	1986	38	1987	39	1988	40	1989
41	1990	42	1991	43	1992	44	1993	45	1994
46	1995	47	1996	48	1997	49	1998	50	1999

```
Specify field numbers( use comma to delimit field numbers).:
    (CR for all fields, or for FROM - TO )

              11 - 36
```

Figure 7.   Screen with available data fields and selection of data for the years 1960-1985.

The list of 11 identification fields as well as the appropriate field numbers is displayed next. Any combination of field numbers might be specified in the highlighted row. These fields are to be used for query specification. The user can select appropriate

rows in this way (Figure 8). Specification of fields ENTRIES, STEP, and FIRSTYEAR is useless, while the information they contain is the same for all data of the same record.

```
Use following symbols

symbol                        meaning

    ?                         help
    <                         less   than
    >                         greater than
    =                         equal
   < >                        not equal
   < =                        less than or equal
   > =                        greater than or equal
    $                         contained within

Specify condition for field AGE   :
```

Figure 8.   Query definition, initial display.

'ENTER' is to be pressed after all desired fields are specified. A sequence of screens for selected identification fields follows. Each display contains a list of available symbols for condition specification. The relation between two expressions is to be specified using logical symbols 'AND' or 'OR' (Figure 9). Pressing 'ENTER' will loop the program once more, so it is better to avoid it.

```
Use following symbols

symbol                        meaning

    ?                         help
    <                         less   than
    >                         greater than
    =                         equal
   < >                        not equal
   < =                        less than or equal
   > =                        greater than or equal
    $                         contained within

Specify condition for field ICD  :   = 139
Specify logical operator (AND/OR) :
```

Figure 9.   Query specification screen.

If the value of an item is not known in advance, help is available. A question mark will invoke help text for fields having the help text defined in advance (Figure 10).

```
                    Numeric field, 4 bytes wide.

                    It containes code for disease as listed :
    179      134 mal.neo.buccal cav.,pharyngs,oesophagus,stomach
    180      135 mal.neo.larynx,trachea,bronchi,lung
    181      136 mal.neo all forms
    182      137 rheumatic fever
    183      138 heart dis.-chronic rheumatic
    184      139 heart-arteriosclerotic & degenerative
    185      140 hypertensive disease
    186      141 cardiovascular disease
    187      142 cardiovasc d.& vasc.lesions affecting CNS
    188      143 bronchitis all forms
    189      144 respiratory dis. all forms
    190      145 stomach ulcer
    191      146 duodenum ulcer
    192      147 peptic ulcer
    193      148 cirrhosis of liver
    194      149 bones and joints-all diseases
    195      150 accidents-motor vehicle
    196      151 accidents-all forms
    197      152 homicide & injury not in war
    234      234 ulcer of stomach,duodenum and peptic
Press any key to continue...
```

Figure 10. Example on help for disease codes.

The user is free to define his own keywords and codes in the file CATALOG. The dBASE III Plus commands MODIFY, EDIT, or BROWSE are to be used for such a purpose while dBASE is on and CATALOG is in use.

The next step following query specification is selection of means for data display (Figure 11). Retrieved data are sent either to the screen, printer, or disk as an ASCII file. When the screen or printer has been selected data are displayed there, either with or without identification codes. Figure 12 displays an example of a print-out of two records along with appropriate identification codes.

When '0' is keyed in, the program returns to the main menu. After the third option was chosen ('save data to disk') the name of the file to save the data in is requested and data are saved to the disk. It is possible to make use of all three options consecutively. Keying in 0 finishes the display option from the first menu.

```
            Options:

            0. return to previous menu

            1. display data

            2. print data

            3. save data to disc


            Enter number:   0
```

Figure 11. Menu containing three ways to display data.

```
    Identification :          male        34   55   141      0      0 50 1950 1 0
    Data :    0.000760    0.000840    0.000830      0.000740      0.000800          0.00
 0800     0.000810    0.000900    0.000980      0.000810    0.001000
    Identification :          female      34   55   141      0      0 50 1950 1 0
    Data :    0.000270    0.000290    0.000290      0.000280      0.000250          0.00
 0230     0.000240    0.000260    0.000230      0.000230    0.000290
```

Figure 12. Example of printed record along with identification data.

Option 3

Selection of this option leads to preparing a data file for later use. The data specification consists of the same steps as in option 2:

- specify screen

- specify database file

- specify columns with required data

- specify query using identification codes.

The text for title and subtitle is requested afterwards. Both will be displayed on charts. A maximum of 30 characters are allowed (the highlighted area has exactly 30 characters). 'ENTER' is to be pressed at the end of each text, unless exactly 30 characters were typed in. The process of searching through the database begins with specification of the x-axis

legend. The *y*-axis legend specification is facilitated by a list of codes being displayed (Figure 13).

```
                              1. current y - data:
            COMMENT =   ;
            SEX = female;
            COUNTRY =   14;
            AGE =   56;
            ICD =   234;
            DESCRIPT1 = 0;
            DESCRIPT2 = 0;
            ENTRIES = 50;
            FIRSTYEAR = 1950;
            STEP = 1;
            DESCRIPT3 = 0;




            Specify legend for y - axes:
```

Figure 13. The *y*-axis legend specification.

The process of data retrieval begins afterward. Both the *y*-legend specification and data retrieval repeats according to the number of records satisfying the query condition. The last step is to define the name of the file which will keep retrieved records and texts of description for later use—the so-called communication file. The data are written there and the main menu reappears on the screen.

Option 4

This option allows for data visualization, curve fitting, and projecting. It starts with specifying the name for the file with retrieved data. 'ENTER' is to be pressed after the file name is keyed in. A menu containing five items is then displayed (Figure 14).

Options are sorted according to the main menu. The curve-fitting procedure starts with the first option. Legends for all *y*-axis data sets are displayed. The number preceding the legend followed by 'ENTER' is to be used for data set specification. The procedure branches now. The first of two possibilities is to display the original data in the form of a connected line chart. In that particular case option 5 is to be chosen. The next menu (Figure 15) contains an option for charting (option 2). Its selection opens the next menu where the second item will draw the original data as a connected line chart. As

```
ESTIMATION MENU


O - exit to DOS
1 - select data for estimation
2 - chose the estimated function
3 - start estimation
4 - results presentation and projection
```

Figure 14. Menu for option 4.

```
PREPARATION OF DATA TO PLOT


O - return
1 - draw
2 - print data
3 - projection
```

Figure 15. Results presentation menu.

soon as charting is finished, pressing 'ENTER' loads the next page with information on the curve legend. To return to the first menu where the process has been started, 'CTRL-Z' or 'F6' key is to be pressed.[5] Zero is to be answered for all but the last menu to back track to the first one.

The second possibility in curve-fitting main menu is to follow the sequence of options to the second one. Three functions for fitting the curve are available: linear, quadratic, and exponential. Curves are specified by an analytical equation along with the text description (Figure 16).

The next natural step is to perform the curve fitting. Results of this process are displayed as an equation with enumerated coefficients and corresponding standard deviation (Figure 17).

This is intended to be used for comparing different fitted functions. The fourth and last step invokes the menu for result presentation and projections. It offers drawing, printing, and projecting. The first option draws the results of curve fitting. The next one would print fitted data to the printer. The last one serves for projecting trends. The menu invoked by the first option offers four different types of data display (Figure 18).

---

[5]CTRL-Z is sometimes indicated as ^Z instead. In the text of this paper CTRL-Z is used.

```
                        ESTIMATION MENU


        1 - y = a + b*x              - linear

        2 - y = a + b*x + c*x^2      - quadratic

        3 - y = a*exp(b*x) + c       - exponential


        Specify function (1/2/3):
```

Figure 16. Fitted curve equation.

```
                ESTIMATION MENU


   y = 0.2180E-04  + -0.7206E-06 *  (x-x0) +   0.7009E-07 *(x-x0)^2
   standard deviation = 0.6223E-05
```

Figure 17. Display of equation fitted to data.

```
        PLOTTING MENU


   0 - return
   1 - manual scaling (by default automatic scaling)
   2 - draw original data
   3 - draw original data, fitted curve and
       confidence intervals.
   4 - draw original data and fitted curve.
```

Figure 18. Menu with charting options.

Manual scaling is to be used in the first option specification. It enables designing of a chart according the user's desire. The maximal and minimal value for y-data is to be specified. If manual scaling is not desired, automatic scaling is the default. Any of three options following the first one are to be used afterwards. Option 3 draws a connected graph of original data (Figure 19).

Figure 19. Connected graph of data retrieved (male mortality from cirrhosis, Czechoslovakia, age 60–65).

The next option allows the drawing of original data with the fitted curve and 95% confidence intervals. The last option displays a fitted curve along with original data without confidence intervals. Data for mortality rates for cirrhosis in Czechoslovakia based on data from the World Health Organization are displayed in Figures 20 and 21.

In all of these cases a hard copy is available using the Print-Screen routine. The following sequence is applied throughout the charting process: 'ENTER' stands for switching between the page with charts and the page with the legend. An example of a legend display is shown in Figure 22.

'CTRL-Z' or 'F6' followed by 'ENTER' stops switching and returns to the menu with charting options.

Choosing a projection in the menu for different ways of result presentation will start a sequence of queries. The first step is for the starting point for projection. The next one specifies the last point up to the one where the projection is to be estimated. Finally, the step for projection is to be specified. It is common to specify the step of one unit, but the user is free to employ any other value. Forecasting future values, as well as estimating past ones, is supported. Charting of results is identical to the previous description. Figure 23 displays the projection of cirrhosis mortality for males aged 60–65 from the year 1960 to 2000 for Czechoslovakia.
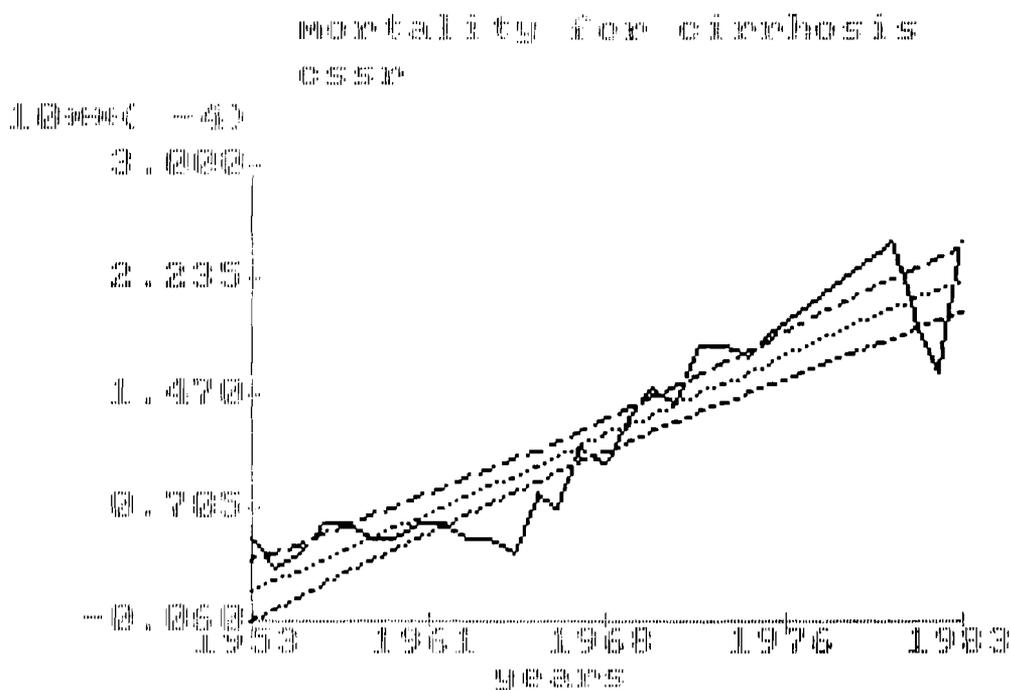
Figure 20. Display of cirrhosis mortality data with linear function fitted and with 95% confidence intervals.

Option 5

This option is very simple. It performs line graphs of all (maximum 10) records retrieved. Lines differ from each other in graphic design. Print-Screen is supported as well. Pressing 'ENTER' after the chart is drawn displays lines with a comment. To stop charting 'F6' or 'CTRL-Z' keys are to be used.

Option 6

This option charts histograms from the retrieved data. Usage is similar to the previous one. An example is shown in Figure 24.

Option 7

This displays contents of the CATALOG file. This file contains a list of codes along with corresponding comments to facilitate the interpretation of identification codes. Specification of labels enables the user to display selected codes.

Figure 21. Display of cirrhosis mortality data with linear function fitted without confidence intervals.

Figure 22. Legend which accompanies the previous figure.
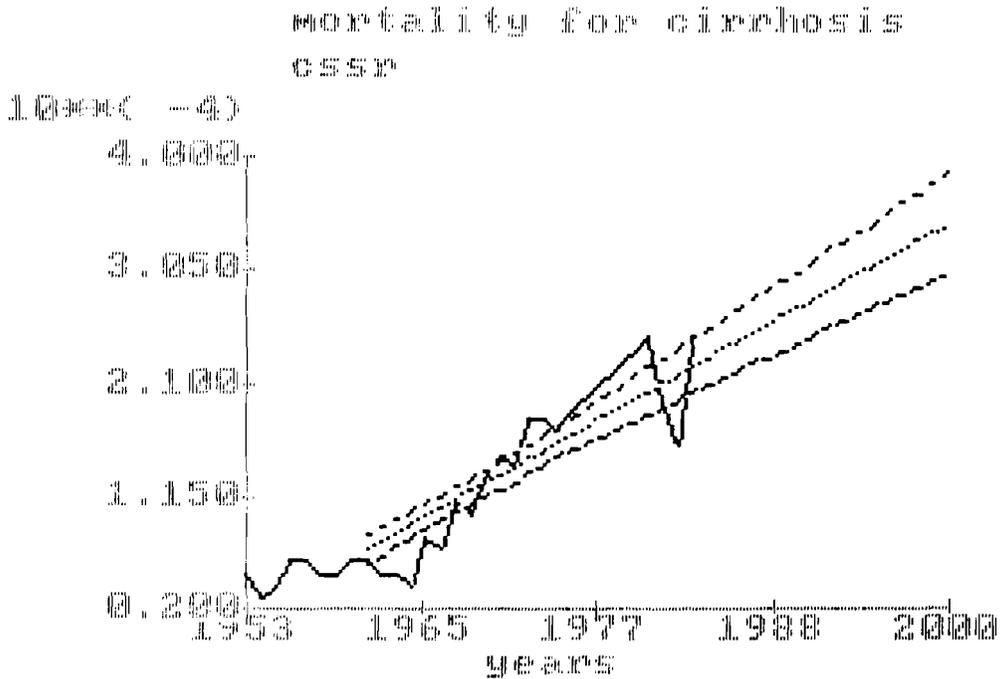
Figure 23. Projection of cirrhosis mortality in male for Czechoslovakia starting from year 1960 to 2000.

Option 8

This option offers several file maintenance routines (Figure 25).

The file maintenance routines are designed to perform several tasks. The first one is to copy the contents of one file to the other. For this purpose the name of the file to be copied and the destination file is to be specified. The use of wild cards is not supported in this option. If no file name has been specified, the menu appears again. After the copy has been made, the new file name along with its size is displayed. The next option allows file deletion. After specifying the full file name, the list of files in the current directory is displayed. The third option is to rename the file. It starts with a listing of files in the current directory, followed with the file name specification query. Two file names are to be specified: one for the original file and the second for a new one. The last option (number 5) is to be used when new data input is desired. While this option currently exists only in a preliminary version, we do not intend to describe it. Zero is used to exit from the current menu.
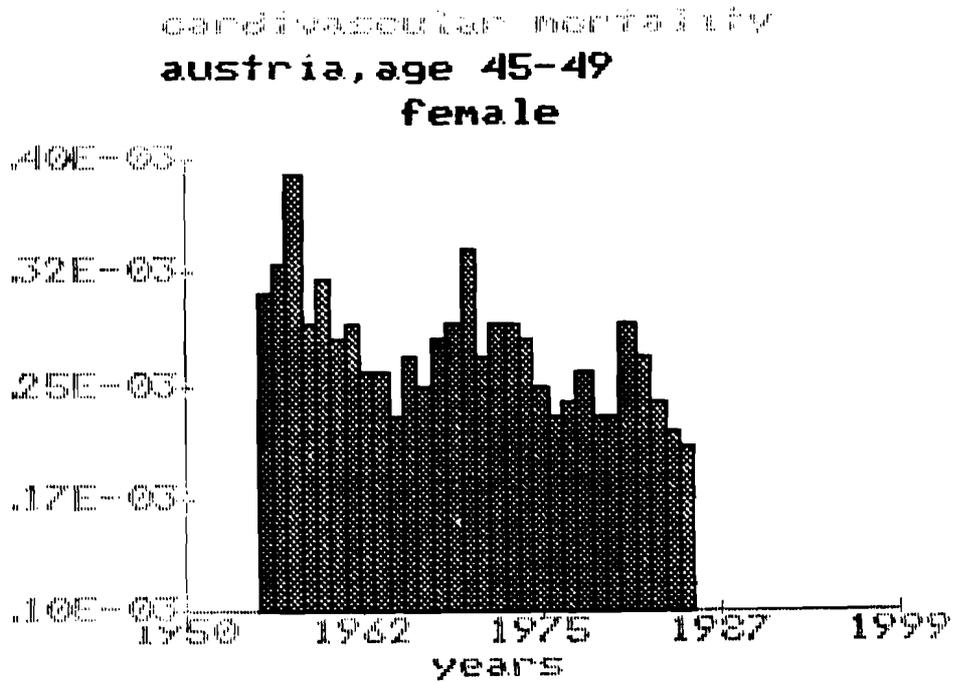
Figure 24. Statistical data for cardiovascular mortality for females in Austria aged 45-49.

File maintenance menu:

    0. exit

    1. Copy file

    2. Erase file

    3. Rename file

    4. Create new data file

    Enter number :

Figure 25. File maintenance routine menu.

PLANS FOR FUTURE DEVELOPMENTS

We intend to proceed in developing the system at least in two directions. The first one is to add new functions for fitting curve. Splines and some other regression models will be implemented in the near future. More exhaustive statistical analysis is planned to be used for evaluation of results. Possibilities of two or more trend comparisons are to be added as well. The ability to transform variables is lacking in MEDDAS and will be added in the future.

CONCLUSIONS

We have begun to develop a system for use in medicine and demography. The result is called MEDDAS. Both authors hope that the fruitful collaboration among scientists from various countries will profit from IIASA's work, and proceed in the future at even higher rates. MEDDAS might also play a tiny role in that direction.

Recent Working Papers Produced in
IIASA's Population Program

Copies may be obtained at a cost of US$ 5.00 each from IIASA's
Publications Department.

WP-86-01, *Exploratory Analysis of the Umea Data at IIASA* by Arno Kitts. January 1986.

WP-86-02, *Increasing Returns to Scale in Heterogeneous Populations* by Robin Cowan. January 1986.

WP-86-03, *Notes on the Effects of Cohort Size on Intergenerational Transfer* by Robin Cowan. January 1986.

WP-86-06, *A Simulation Study of the Conditional Gaussian Diffusion Process Model of Survival Analysis* by Fernando Rajulton and Anatoli Yashin. February 1986.

WP-86-09, *The Two Demographic Transitions of Finland* by Wolfgang Lutz. February 1986.

WP-86-19, *The Division of Labor for Society's Reproduction: On the Concentration of Childbearing and Rearing in Austria* by Wolfgang Lutz and James Vaupel. April 1986.

WP-86-29, *Dialog System for Modeling Multidimensional Demographic Processes* by S. Scherbov, A. Yashin, and V. Grechucha. June 1986.

WP-86-34, *Culture, Religion and Fertility: A Global View* by W. Lutz. July 1986.

WP-86-37, *The LEXIS Computer Program for Creating Shaded Contour Maps of Demographic Surfaces* by B. Gambill, J. Vaupel, and A. Yashin. August 1986.

WP-86-53, *Population Models Analysis Program (POPMAN)* by A. Lewandowska. October 1986.

WP-86-59, *Cancer Rates over Age, Time and Place: Insights from Stochastic Models of Heterogeneous Populations* by J. Vaupel and A. Yashin. October 1986.

WP-86-60, *Heterogeneity in Composite Link Models* by C. Vanderhoeft. October 1986.

WP-86-63, *Derivative-free Gauss-Newton-like Algorithm for Parameter Estimation* by S. Scherbov and V. Golubkov. November 1986.

WP-86-69, *Modelling Kinship with LISP — A Two-Sex Model of Kin-Counts* by J. Bartlema and L. Winkelbauer. November 1986.

WP-86-74, *Computation of Multi-State Models using GAUSS, A Matrix Based Programming Language* by A. Foster and N. Keyfitz. December 1986.

WP-86-76, *Structural Minimization of Risk on Estimation of Heterogeneity Distributions* by A. Michalski and A. Yashin. December 1986.

WP-86-77, *A Note on Random Intensities and Conditional Survival Functions* by A. Yashin and E. Arjas. December 1986.

WP-86-78, *Cause Specific Mortality in Japan: Contour Maps Approach* by B. Gambill, A. Yashin, J. Vaupel, Z. Nanjo, and T. Shigematsu. December 1986.

WP-86-81, *Kinship and Family Support in Aging Societies* by D. Wolf. December 1986.

WP-87-12, *Comparative Anatomy of Fertility Trends: The Aging of the Baby Boom* by W. Lutz and A. Yashin. January 1987.

WP-87-13, *Using the INLOGIT Program to Interpret and Present the Results of Logistic Regressions* by D. Wolf. January 1987.

WP-87-46, *The Multistate Life Table with Duration-Dependence* by D. Wolf. May 1987.

WP-87-51, *The Concentration of Reproduction: A Global Perspective* by W. Lutz. June 1987.

WP-87-58, *A Simple Model for the Statistical Analysis of Large Arrays of Mortality Data: Rectangular vs. Diagonal Structure* by J. Wilmoth and G. Caselli. June 1987.

WP-87-59, *Sibling Dependences in Branching Populations* by P. Broberg. June 1987.

WP-87-87, *The Living Arrangements and Familial Contacts of the Elderly in Japan* by K. Hirosima. September 1987.

WP-87-92, *The Demographic Discontinuity of the 1940s* by N. Keyfitz. September 1987.

WP-87-104, *A Random-Effects Logit Model for Panel Data* by D. Wolf. October 1987.

WP-87-116, *Some Demographic Aspects of Aging in the German Democratic Republic* by T. Büttner, W. Lutz, and W. Speigner. November 1987.

WP-88-10, *On the Concentration of Childbearing in China, 1955–1981* by W. Lutz. February 1988.

WP-88-13, *Beyond "The Average American Family": U.S. Cohort Parity Distributions and Fertility Concentration* by M. King and W. Lutz. March 1988.