The Israeli Society for the Prevention of Heart Attacks is conducting a clinical trial involving over 37,000 patients at 18 medical centers during the screening period, of whom 3,000 are selected for a five to seven year follow-up period. Some of the 8 different types of forms may appear only once per patient, while others may appear as many as 42 times. We also store results of up to 19 laboratory tests made on blood samples taken at the various visits. The main index during the screening period is a logbook number. After inclusion, the main index is the patient's randomization number. There are seven reporting systems in the project: 1) quality assurance; 2) project administration; 3) form flow management; 4) finance; 5) medication control; 6) adverse event reporting; 7) statistical analysis.

The data entry and reporting systems, including graphic output, are written in SAS @ System, Release 6.07 on a SUN SPARCstation 1 TM under UNIX TM with OpenWindows. To maintain the hierarchical structure, each form type is stored in a different data set. These are merged as required by patient number and visit date. Some of the new features in Release 6.07 that reduce development and run time are: flexibility in ARRAY's, IN, WHERE in procedures, NODUPKEY, CLASS in MEANS, and PROC PHREG.

The database contains 8 different types of forms. Some forms are to be completed only once per patient, while others may appear as many as 42 times over the 5-7 year follow-up period. In addition it contains results of up to 19 laboratory tests made on blood samples taken at the various visits.

The main index of the database for all records pertaining to the screening period is a logbook number, unique within each medical center. The main index for all records pertaining to the follow-up period is the patient's randomization number, which determines if he/she is receiving the study drug or placebo.

There are 7 major reporting systems in the project, all written in the SAS System Release 6.07:

1. Quality assurance
2. Project administration
3. Form flow management
4. Finance
5. Medication control
6. Adverse event reporting
7. Statistical analysis
The Society ran another large clinical trial (SPRINT - over 5,000 patients) during the years 1981-1985. The reporting systems in that trial were also written in SAS. But the configuration at that time was SAS Version 5.18 running on a Data General MV4000 under AOS/VS. The BIP Project is written in the SAS Release 6.07 running on a SUN SPARCstation under UNIX with OpenWindows. All of the above-mentioned reporting systems were written and are maintained by a staff of 3 full-time programmers. The initial reporting systems were written and de-bugged within a few months. The data entry system was written in 8 person-months.

The computer center is located at the study Coordinating Center, in the Neufeld Cardiac Research Institute of the Sheba Medical Center in Tel Hashomer. Forms are sent by messenger service from the 18 medical centers to the coordinating center, where they are entered into the computer by a staff of medical secretaries. Weekly reports are transmitted by FAX from the medical centers and are likewise entered into the computer.

Blood samples are sent by messenger service from the medical centers to the central project laboratory located at Wolfson Medical Center in Holon. Results are sent by file transfer, via telephone lines, from the PC at the central lab to the PC at the coordinating center, and from the PC to the SUN on the local ETHERNET TM network.

**PATIENT AND FORM FLOW**

a. The Logbook

Before the project began in May 1990, each of the 18 medical centers began to compile a logbook of potential participants from among its patients. Eligible patients have previous myocardial infarction (MI) (0.5 to 5 years prior to admission) and/or coronary insufficiency (stable angina pectoris, AP) during the 2 years preceding admission and who do not satisfy the exclusion criteria. The patient’s logbook number is a serial number within the medical center. During the screening period, each center added names to the logbook. The logbook now contains 37,583 names.

b. Screening

During the almost three years screening period (May 1990 - January 1993), each center examined patients from the logbook, at an average rate of about 30 per center per month. At the first visit, the physician tried to determine if the patient was obviously ineligible. If not, a blood sample was drawn and sent to a central laboratory for examination. If the lipid values were within certain ranges, the patient was invited for a 2nd visit, during which he/she met with a dietician and received a tailor-made diet. Two months later the patient returned for a second blood test. If the results were within certain ranges, the patient returned for a 4th visit, during which the physician made further examinations and reached a decision regarding the eligibility of the patient to be included in the study. Then the patient had to decide and to sign an informed consent.

The major form being filled out during this time is the Eligibility Form. It contains a different part for each of the four visits and is sent to the Coordinating Center as soon as a decision is made that the patient is included (at visit 4) or that the patient is not included (at any of the first 4 visits). If the patient is included, then an Admission Form is also filled out. At that time the patient is assigned a randomization number. At the conclusion of the screening period 3123 patients had been included in the study.
c. Follow-up

At visit 4, if the patient is included, the patient is given a box bearing his/her randomization number and containing 80 tablets. Every two months thereafter, until February 1997, the patient is to return the unused medication and receive a new box. At each visit a Drug Compliance Form is filled out.

Every 4 months, the patient undergoes a thorough physical examination and a Medical Follow-up Form is filled out. At any time, the patient may report changes in his/her medical condition and then an Adverse Event Form is filled out. If the patient reached a study endpoint (mortality or non-fatal myocardial infarct), then a Critical Events Form is filled out. If the patient withdraws at any time for any reason, then a Permanent Withdrawal Form is filled out.

Every week these forms are sent by messenger service from the 18 medical centers to the coordinating center. In addition, every week, a list of patients who arrived for each visit is sent by FAX to the coordinating center.

Blood samples are sent by messenger service from the medical centers to the central laboratory. Once a week, a file containing the lab results, is sent by telephone from the PC in the lab to a PC at the coordinating center. A SAS program checks the results (against the patient's age and past history) and outputs a detailed lab report that is sent to the patient's medical center.

THE DATA SETS

a. The Weekly Reports

Each week, each center faxes a list of the patients who were seen that week, and the visit number. These lists are kept in two SAS data sets - one indexed by the logbook number (visits 1-4) and one indexed by the randomization number (visits 5-46).

Each week during the screening period, each center faxes a list of the patients who were included that week. This is kept in a third SAS data set.

Each week, the central laboratory faxes a table containing the number of blood samples analyzed from each medical center, by visit number. This list is kept in a fourth SAS data set.

These data sets are used by the project management to monitor the progress of the medical centers. They are also used as the basis for monthly payments to the medical centers.

b. The Office

Every form that arrives at the coordinating center is recorded. This information is kept in a separate SAS data set for each of the 8 types of forms. These records contain the following information: patient I.D., date form arrived at coordinating center, date given to data entry operator, date received from data entry operator, date returned to medical center (in case data entry program discovered a problem in the form), data received from medical center, date given to data entry operator, and date received from data entry operator. In addition, other information may also be recorded for specific forms.

These data sets are used to keep track of the form flow as well as to plan and monitor the work schedule of the office staff and the data entry operators.

c. The Forms

Each form is entered in its entirety into a separate SAS data set for each of the 8 types of forms. These data sets comprise the database itself.
The hierarchical structure is preserved by storing each type of form in a different data set. These sets may be merged by patient I.D. number, and, where appropriate, by visit number. For the adverse event forms and the critical event forms, which are ordered by date rather than visit number, data sets are merged by patient I.D. number and RETAIN, ARRAY, FIRST and LAST. are used as necessary to generate one record per patient.

d. The Lab Results

All lab results are kept in one of two SAS data sets, one for visits 1 and 3 (indexed by logbook number and visit number) and one for all subsequent visits (indexed by randomization number and visit number).

e. Traceability

All records in all data sets contain the name of the data entry operator and the date of first entry of the record as well as the name and date of the most recent modification to the record.

f. Analysis

All data sets appear twice on the disk - one copy for data entry and modification, which is continually updated, and one copy for data analysis, which is updated once a week. There are two reasons for this duplication. First, the heavy statistical analysis programs do not lock the data sets, thus they do not add to the response time for data entry. Second, when a programmer debugging a program obtains different results from the previous run, he/she can be sure it is because of a change in the program and not a change in the data.

THE SEVEN REPORTING SYSTEMS

a. Quality Assurance

The programs that read the data from the data entry screens perform error checks of three types: within the field, among fields on the same form, among fields on different forms. We look for impossible and unlikely values and combinations of values. We also check that all previous forms for the patient are already in the computer.

In addition, quality assurance programs were written to check aggregate phenomenon, such as unusual frequency distributions, drifts in mean values, and standard deviations (for continuous variables) and in frequency distribution (for discrete variables) and inter-center variability. These reports are given to the monitors-nurses who visit the medical centers regularly to ascertain that the study protocol is being adhered to, and who assist the centers in remediating problems that are discovered.

b. Project Administration

The coordinating center management receives weekly and monthly reports, by medical center, showing the progress in recruiting new patients and the follow-up of existing patients. Unusual phenomena, such as large numbers of patients who missed a visit, are flagged, as on quality control charts.

Management also receives numerical summaries of the error reports given to the monitors. The monitors receive lists of names. The managers receive numbers of patients in each center, according to error or problem type.

In addition, management can request ad-hoc reports as answers to specific queries.

Other reports indicate the locations of forms that have arrived at the coordinating center but are not yet in the computer.
The coordinating center management receives information on the number of forms that should have arrived, number that actually arrived and the number that entered the computer. This latter is also given by data entry operator.

c. Form Flow Management

This information is used for planning work schedules and budgeting office time and data entry time.

d. Finance

The centers are paid according to the number of patients they examine. The weekly reports from the centers and from the central laboratory form the basis for these monthly payments. When the forms are finally entered into the computer, they are used as an audit against the original numbers.

e. Medical Control

Each box of study medication is labelled with the name of the medical center, the randomization number of the patient, and the visit number. The boxes are ordered from the manufacturer according to the planned visit schedule of each patient. The boxes arrive at the coordinating center and are stored there until they are sent to the medical centers. When the box is given to the patient, a compliance form is filled out. The same form also contains the number of tablets that the patient returned from the previous box. These leftover tablets are returned to the coordinating center. If for some reason the visit did not take place, or if the patient is temporarily withdrawn from the medication, this is indicated on the compliance form, and the entire box is returned to the coordinating center. In case of permanent withdrawal, all remaining boxes are returned to the coordinating center. Periodically, the returned tablets are destroyed by a special committee, in accordance with Israeli law. All of these transactions are recorded in a SAS data set, and SAS programs print progress reports and legal documents.

f. Adverse Event Reporting

Periodically, a tape is sent to the manufacturer containing information on all reported adverse events, all compliance forms (including dates of temporary and permanent withdrawal), all concurrent medications (including dates of administration), and all non-blinded laboratory values.

These data are added to the manufacturer's database in compliance with local and international laws. In addition, these data are passed on to the International Review Board of the project, who analyze the data after decoding the randomization number. On the basis of these analyses, they may stop the study prematurely.

g. Statistical Analysis

The coordinating center produces many statistical analyses and graphs from the data set, which, meanwhile do not depend on decoding the randomization number. These reports are presented regularly to the senior investigators as well as to the International Review Board. They are based on more sophisticated procedures in SAS/STAT such as GLM, NLIN, LIFETEST, PHREG as well as the 2- and 3-dimensional procedures in SAS/GRAPH.

DATA ENTRY

In the past, data entry was performed using SIR on a Data General MV4000. Every week, the SIR database was transferred to the SUN workstation where the SAS database was built. All of the above reports are run on that SUN. This year, a decision was made to replace the Data General with a second SUN.
workstation. We also decided to write the entire data entry system from scratch in SCL with PROC FSEDIT of SAS/FSP @. We began in November 1992 and finished in April 1993. Altogether it took about 8 person-months.

A screen was designed for each form in each of the four systems—weekly reports, the office, the forms, and the lab results. For each screen a series of logical checks is made: checks in the contents of each field; cross-checks among fields in the same data set; cross-checks with fields in other data sets.

ADVANTAGES OF VERSION 6.07

There are several new features in SAS Version 6.07 that were unavailable in Version 5.18. Some of these have made our programming much easier and have greatly reduced program development and debugging time.

a. ARRAY
Assigning initial values ARRAY A Al-A4 (100,200,300,400);
Retaining all variables in an array; RETAIN A;

b. IN
IF CITY IN ('AFULA', 'HAIFA', 'SAFED'); Similarly for WHERE

c. CLASS in PROC MEANS
The data set need not be sorted.

d. FREQ in PROC GCHART
Prints the frequency above each bar.

e. PROC PHREG
The procedure for survival curves with independent variables.

This had previously been unsupported, existing only in the SUGI library and unavailable under UNIX.

f. Interactive SAS with Windows

SAS under UNIX takes full advantage of the OpenWindows, which reduces debugging time considerably.

It is possible to check data set contents, including labels and formats, in separate windows.

It is possible to run various versions of a program under development. Work data sets are not erased and it is possible to restart at any point.

CONCLUSION

SAS 6.07 has enabled us to build and maintain all of the reporting systems necessary to manage and maintain a large long-term multicenter clinical trial, involving over 37,000 patients. These systems were built and debugged in a very short time due to the power of SAS 6.07.

In the future, we hope to use SAS/AF @ to permit the project managers to run (and build) SAS programs themselves.

SAS, SAS/AF, SAS/FSP, SAS/GRAPH, and SAS/STAT are registered trademarks or trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration.

Other brand and product names are trademarks of their respective companies.