BEN E. AMSLER

SELECTION AND PERFORMANCE ASSESSMENT OF AUTOMATED PATIENT MONITORING SYSTEMS FOR THE JOHNS HOPKINS ADULT INTENSIVE CARE UNITS

APL activities involved in the definition, selection, implementation, and evaluation of computer-assisted systems for patient monitoring at the Johns Hopkins Hospital are described. Attention is focused on the evaluation and refinement of an experimental computer algorithm used to detect and alarm automatically when critical changes occur in patient electrocardiogram patterns.

BACKGROUND

At the Johns Hopkins Hospital (JHH), two adult intensive care unit areas are provided to accommodate the increasingly complex procedures and technology now routinely employed in the care of acutely ill patients. One of these areas is assigned for the exclusive care of at-risk nonsurgical patients and is designated the Medical Intensive Care Unit (ICU). (The other area, the Surgical Intensive Care Unit, is reserved for the care of critical post-surgical patients.) The ICU area includes 20 individual patient rooms together with a centrally located nurses' station and other facilities needed to accommodate the required range of specialized functions. In addition, special space and facilities (e.g., isolated electrical power panel, oxygen supply, patient monitoring signal lines) are provided in each patient room to accommodate the latest advances in bedside life support and patient monitoring equipment.

A major segment of the ICU is further designated the Coronary Care Unit, with facilities and staff dedicated to the care and treatment of acute coronary patients. The remainder of the beds are reserved for those noncoronary patients who require special care beyond that afforded by the regular inpatient facilities.

The use of electrocardiogram (ECG) signals in the monitoring of at-risk patients has been standard practice for several years. More recently, invasive monitoring of both arterial and venous blood pressure via special pressure-measuring catheters has also become commonplace. Today, critical-care bedside monitoring equipment typically comprises one or more special signal conditioner/amplifier units, together with a multichannel oscilloscope for the display of ECG's, selected blood pressures, or other physiological waveforms.

To allow more nearly continuous on-line monitoring of all patients, most modern intensive care installations also include one or more central monitor stations at which signals derived from a number of patient beds are displayed simultaneously. With this arrangement, the physiological signals (e.g., ECG waveform, blood pressure, temperature) being collected at several bedsides can be monitored by a single nurse or specially trained technician. The attending clinical staff can then be alerted immediately and action can be taken in response to the occurrence of critical life-threatening events such as asystole (the cessation of heart action) and ventricular fibrillation (noncoordinated quivering action of the heart), as well as to less critical — but potentially significant — indicators of deterioration in patient status, such as intermittent arrhythmias (irregular variations in ECG waveform and/or rhythm) or rapid changes in blood pressure.

Initially, central patient monitoring at JHH and most other institutions employed only visual observation of ECG display oscilloscopes by an assigned staff member to detect patient ECG arrhythmias for action by the clinical staff. However, experience in central monitoring gradually brought into question the reliability of arrhythmia detection based only on visual monitoring of the ECG displays. At JHH, this suspicion was supported by informal observations of central monitoring activities during 1975 and 1976.

At that time, it was concluded that the deficiency resulted from a combination of two factors. First, standard "nonpersistent" ECG display oscilloscopes were then being used at both the central station and bedside. This type of display was generally believed to be satisfactory for most routine bedside activities, where attention was necessarily focused on a single display. However, adequacy for simultaneous monitoring of a number of ECG traces at the central station was seriously questioned.

More generally, it was concluded that when many ECG traces were being monitored simultaneously, the level of attention required to detect visually occasional transient changes in individual ECG's could not be maintained for any appreciable length of time.

The task of intently watching an array of ECG traces is both boring and tiring; thus, monitor watch fatigue becomes a limiting factor in the visual detection of critical ECG changes. Some type of attention-getting, automated alarm system is required to alert the monitor watch and focus attention on the particular ECG trace of immediate concern.

It is of interest that during the intervening years since these observations were made at JHH, more rigorous studies have been performed at other institutions to determine the limitations inherent in centralized ECG monitoring via the visual observation of a group of ECG displays.[1,2] In tests of this type, miniature portable tape recorders are often used to record ECG signals continuously for extended periods of time on selected patients whose signals are also among those being visually monitored at the central station. After recording, the tapes are analyzed and results are compared with the central station monitoring observations. The results of these studies have been consistent with the initial observations made in the JHH Coronary Care Unit CCU. One recent study[3] reports that even when using persistent "memory-type" oscilloscope displays, episodes of "serious arrhythmias" remained undetected in 56% of the patients. This same study also notes that results did not improve even when more than one person was assigned to the monitor watch function.

SECOND-GENERATION ADULT INTENSIVE CARE PATIENT MONITOR SYSTEM: DEFINITION/PROCUREMENT

By mid-1976, the need to upgrade the existing JHH adult intensive care monitor systems was generally recognized. However, no general agreement had been achieved regarding either the clinical needs to be met or the operating features to be included in the upgrading. Accordingly, in late 1976, a special task team was established at JHH to accomplish the following:

1. Develop functional requirements and objectives of the system that reflect the needs of the various clinical specialties involved (e.g., cardiology, surgery, respiratory care, etc.), together with the practical considerations and techniques involved in the day-to-day use of the existing patient monitor system;
2. Translate the identified functional requirements into practical system concepts and proven technical alternatives;
3. Review alternative system implementation approaches (e.g., procure a new system or modify and upgrade the existing system) and recommend an appropriate course of action; and
4. Review commercially available components and systems and obtain proposals for system implementation from competent and experienced vendors.

Overall leadership and direction of the task team was provided by the Chairman of the Department of Anesthesiology. Regular membership included the cardiologist then assigned as Director of the CCU, the supervisor of the Clinical Engineering Services Group at JHH, and two APL staff members. In addition, several staff physicians and nurses experienced in the use of the existing monitoring equipment participated on an *ad hoc* basis as required. Responsibility for the coordination of technically oriented activities and for the identification of important trade-offs between clinical needs and technical constraints was specifically assigned to the APL members of the team.

During the initial stages of the effort, team members, together with representatives of the affected JHH clinical areas, visited other clinical facilities where some of the latest monitoring equipment and computer aids were in use. Application and utilization of the systems in the ICU environment were observed, and user comments were solicited to help identify clinically useful concepts. In addition, team members consulted extensively with various members of the JHH clinical staff to better define the differences in monitoring priorities as a function of intensive care area (e.g., coronary versus surgical), user responsibilities (e.g., staff physician versus nurse), and patient characteristics.

The potential value of computerized detection of ECG arrhythmias had been recognized from the outset. But, at the same time, the question of accuracy in general — and especially the potential for false alarming by the automated ECG arrhythmia detection process — had also been repeatedly highlighted.

In this connection, efforts to develop computerized algorithms for the purpose had been under way for several years and were regularly reported in the literature. Thus, the difficulty of simultaneously providing an acceptably high likelihood of cardiac arrhythmia detection and, at the same time, achieving a sufficiently low probability of false alarming on nonsignificant changes in ECG patterns caused by patient motion or variations in lead placement was well known. Notwithstanding the importance of this trade-off, no concrete data could be found that defined just how well a particular computer algorithm or commercially available system would perform in a specific clinical environment. Such quantitative results as could be found were usually based on laboratory-type tests involving taped ECG segments of varying (and usually undefined) complexity and length. To complicate matters further, anecdotally reported assessments by users of then-installed automated systems ranged from glowing testimonials to complete denunciations of automated arrhythmia detection as an impractical or worthless concept. In summary, while detections of transient arrhythmias based solely on visually monitoring centralized ECG waveform displays had been found to be deficient, the capabilities of existing automated systems were also known to be less than perfect. Sole reliance on human observers for the detection of transient events

resulted in a relatively high incidence of "false negative" reports (i.e., transient arrhythmias that should have been reported but were not), even though the human observer is an excellent judge of whether a given ECG waveform does or does not represent a clinically significant event. On the other hand, although most automated systems were believed to be less prone to false negative errors, experience had shown that the systems also produced a much higher incidence of "false positive" alarms (i.e., cases in which normal ECG beats were erroneously labeled as arrhythmias). Thus, an appropriate combination of the human and automated arrhythmia monitoring functions could be expected to improve materially the overall likelihood of reliable arrhythmia detection.

Ultimately, on the basis of discussion with ICU staff members at three different installations where automated ECG arrhythmia detection and alarming systems had been in use for some time, it was tentatively concluded that, if suitably configured and properly used, current arrhythmia detection systems could provide major assistance in the care of those patients who were suffering from severe cardiac arrhythmia problems. However, enough negative reports had also circulated to cause concern about specific performance limitations of certain systems and to bring into serious question the optimistic performance claims of some commercial system vendor representatives.

By early 1977, a set of characteristics and constraints for the upgraded patient monitor system for adult intensive care had been identified and set forth for final review and comment by all affected parties. In this document, desired system features were identified as the following:

1. Automated ECG arrhythmia detection;
2. Automated storage of detected arrhythmic ECG waveforms and the capability to display them selectively on command to permit rapid retrospective review of chosen segments;
3. A capability for automatic display and on-command printing of trend plots of key patient status indicators such as heart rate, level of ECG arrhythmic activity, and blood pressure;
4. Simplified front panel controls and displays of bedside monitoring units;
5. The incorporation of memory-type oscilloscopes at the central stations; and
6. Automatic provision of hard-copy records of detected ECG arrhythmias for examination and validation by the monitor watch.

Concurrent with the assessment of clinical needs and identification of functional requirements, the technical features of currently marketed monitoring systems and hardware were also reviewed. That effort soon established that the newly defined objectives and constraints of the monitor system could not be met via any straightforward modification or expansion of the existing JHH systems (which by that time reflected the technology of about 10 to 15 years earlier). At best, only selected components of the system (e.g., the bedside oscilloscopes and some of the recorders) could be retained if their interfacing with other more up-to-date components was ultimately found to be practical.

A Request for Proposal (RFP) that reflected the defined clinical needs was prepared and distributed to all interested vendors. In order to focus on already existing and validated design concepts, desired monitoring features were defined, to the extent possible, in terms of specific functions and capabilities already being marketed by at least one system vendor. A special requirement for formal testing of systems as part of the preselection evaluation effort, using an existing JHH-prepared ECG arrhythmia test tape, was also defined as a precondition for consideration of any candidate system.

Preliminary responses to the RFP were obtained from several vendors who were then invited to meet with the task team to review in more detail the functional, physical, and interface requirements that would have to be met by any system installed in the JHH adult ICU areas. Formal proposals were received from five vendors for the design and installation of patient monitoring systems. These proposals were reviewed both for clinical adequacy and for technical suitability, after which the list of contenders was reduced to two.

During the final evaluation period, one day was spent at each of the two vendor facilities to further review system hardware design features, observe fabrication techniques, assess vendor facilities and customer support capabilities, etc., and to gain some "hands-on" experience in the use of both bedside and central station controls and displays. Formal tests of the automated ECG arrhythmia detection capabilities were also performed during each visit, using the JHH-prepared ECG test tape. Test results for each system were retained for subsequent review and comparison.

The characteristics and features provided by the two systems were acceptable, on the whole, although each exhibited specific advantages and disadvantages in certain areas. Demonstrated arrhythmia system performance using the test tape was found, for the most part, to be comparable to — and generally in accordance with — expectations based on vendor publications and other literature.

However, the arrhythmia detection capabilities differed in one significant regard. In particular, the tests showed that one of the systems was not able to detect consistently even critical ECG arrhythmias if they were immediately preceded by a burst of "noise" resulting from patient motion, loose ECG electrodes, etc. This deficiency was deemed unacceptable by the involved JHH clinical personnel. In response, the vendor indicated his awareness of the problem and pointed out that the development of a new "improved" computer algorithm to eliminate the problem was well along and that its early incorporation for trials and evaluation at selected clinical

sites was planned. Although the algorithm had not yet been finalized for use in a multibed clinical environment, the ECG tape tests were repeated using an experimental version that was then undergoing laboratory tests. The resulting tests did indeed verify that the unacceptable characteristic had been eliminated without otherwise affecting the ability to properly detect ECG arrhythmias as reflected in the test tape. With this change, there existed no observable difference in arrhythmia detection performance of the remaining two candidate systems; final selection of the vendor and the associated patient monitoring system was therefore reduced to a point-by-point comparison of other system design characteristics, vendor resources, system cost, etc.

During the proposal review and evaluation effort, design alternatives unique to each system were identified and presented to members of the intensive care clinical staff to establish the relative desirability and utility of each. In the course of this effort, it became clear that the system procurement cost could not be held within the limits prescribed by JHH management without eliminating some features deemed desirable by one or another segment of the involved clinical staff. This, in turn, refocused attention on the need to retain components of the existing system wherever possible and to identify all nonessential features of each system that could be deleted without degrading essential functions in the process. As a result of that effort, composite concepts for the two systems were defined to allow maximum use of some components of the existing system. The system concept, as it was finally provided to the competing vendors, included special bedside equipment enclosures, signal interconnection modules, and specially configured central station enclosures to be designed and provided by APL. In addition, less costly alternative configurations of vendor-supplied equipment were also identified to facilitate the final comparison of capabilities for the two proposed configurations as a function of system cost.

In the final analysis, it was concluded that, if the newly developed but as yet unmarketed ECG arrhythmia-detection algorithm were included, the system as finally proposed by that vendor was preferred for central patient monitoring in the JHH adult ICU's. The selected vendor was notified, and technical specifications were prepared for inclusion in the system procurement contract. In this specification, all electrical and mechanical interfaces between vendor-supplied system components and other elements (e.g., the bedside oscilloscopes and patient signal connections) that would be separately supplied were defined. The schedule for delivery and installation of various segments of the system was also delineated, along with the special system-level acceptance tests that would be performed at JHH as a condition for acceptance of the delivered system. Details of that specification were negotiated with responsible vendor representatives, after which the specification was jointly approved by JHH and the vendor as the basis for system procurement.

The principal features of the system, as ultimately configured, are illustrated in Fig. 1, and the physical arrangement of the central station is shown in Fig. 2. Automated data processing for both the medical and surgical ICU systems is provided by remotely located minicomputers in combination with 15-megabyte random-access disk units on which all alarmed-on ECG segments are stored for selective recall, evaluation, and copying at the central station via a special keyboard terminal and printer.

Conversion of the bedside monitoring equipment took place during the first and second quarters of

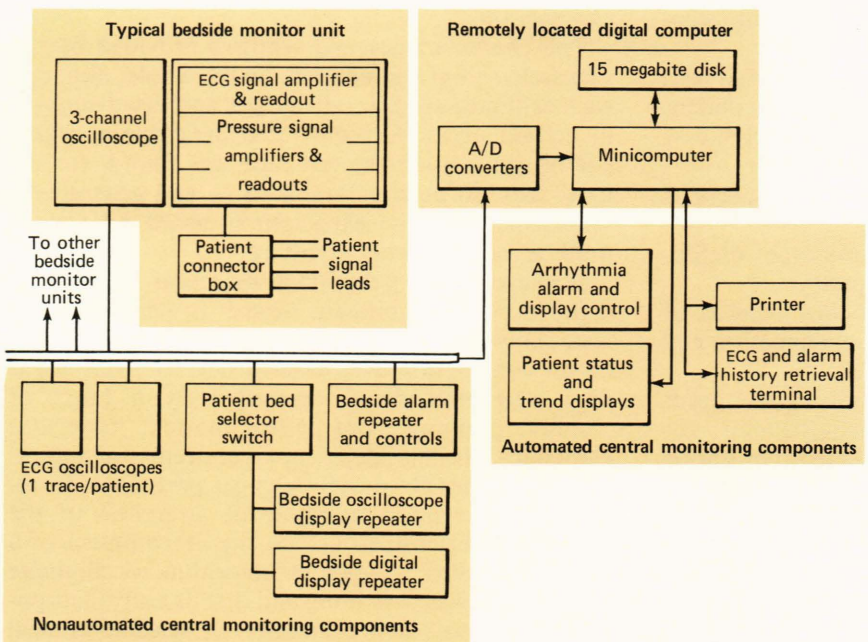

Fig. 1—Patient monitoring systems for the adult ICU's include standard bedside monitoring units, standard nonautomated central monitoring components, and special minicomputer based systems that perform automated detection of patient ECG arrhythmia and pressure computations and provide patient status and trend displays.

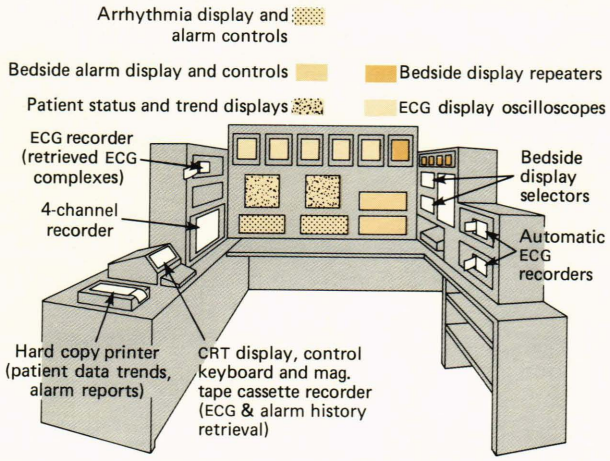

Fig. 2—Standard central monitoring and automated functional components are grouped together in a central monitor station display console.

1978; the central stations and associated data processing equipment were delivered in the second quarter of the year. Performance acceptance tests were performed on all hardware components prior to their integration into the bedside units and central stations.

After integration and checkout of the central station equipment, special system acceptance tests were performed as prescribed in the procurement specification, using the ECG arrhythmia test tape. They were carried out to ensure that all automated functions were in accordance with the specifications and that the ECG arrhythmia detection performance was consistent with results obtained during the earlier tests of the developmental model. During those tests, several software errors were detected and corrected by the vendor.

Finally, special interfacing cables and adaptors were devised to facilitate direct "plug-compatible" substitution of the new central stations in place of the existing units. This was accomplished, and the systems became operational in July 1978.

EVALUATION OF THE NEW ARRHYTHMIA MONITORING SYSTEM

During completion of the procurement specification, it became apparent that the new software package slated for incorporation into the JHH systems represented a larger departure from the version then being marketed than was envisioned by the task team prior to system selection. This was due not only to the new arrhythmia detection software included in the JHH system but also to a number of as yet unmarketed features that had been included by the vendor to meet other JHH-defined objectives. These included automatic storage and recall of alarmed-on ECG complexes and the capability for automated computation of blood pressures of selected patients at the central station.

Thus, both the automated system software and some of the associated hardware included new design features that would be undergoing initial clinical trials in the JHH intensive care areas. It was therefore decided that, at least during an initial 6- to 12-month period following personnel training and system shakedown, data would be collected both to validate new system functions and to quantify the accuracy of ECG arrhythmia detection actually achieved by the system under realistic ICU conditions.

As noted previously, during the preselection review of various arrhythmia detection systems, few quantitative data were found that could be used to define the actual clinical performance of arrhythmia monitoring systems then being marketed. The general lack of operational data resulted at least partially to the fact that in most — if not all — automated arrhythmia monitoring applications, no clinical staff member is exclusively assigned to monitor and manually validate as true or false all system-generated ECG arrhythmia alarms, even though the capability for such validation is provided in many arrhythmia monitoring systems. On the contrary, it appears that more often the system alarms are simply noted by the nearest member of the on-duty staff and clinical action is taken as appropriate. Thus, manual resetting of the alarms occurs irregularly at best, and long-term collection of data needed to establish alarm detection accuracy is simply not feasible under those conditions.

On the other hand, at JHH a member of the intensive care staff had always been assigned to provide a continuous, 24-hour per day monitor watch at each central station. Moreover, it was mandated that at least initially this special assignment would be continued after installation of the new central monitoring systems. With the new systems, however, the monitor watch would be expected to check visually the ECG tracing or oscilloscope waveform associated with each system alarm and manually identify the alarm as "true" or "false." In this way, the ECG arrhythmia trend plots provided by the new system for each patient could be made to reflect more accurately only true and not false arrhythmia detections. Thus, the JHH installation provided the opportunity to gather data for refinement and validation of new system functions. More significantly, it also provided a unique capability to collect the long-term quantitative data needed to establish the accuracy of a typical, automated, ECG arrhythmia detection system in a real clinical environment without the addition or reassignment of personnel for this purpose.

Toward that end, the capability to record automatically each occurrence and type of arrhythmia alarm for each patient, together with the associated manual reset condition (i.e., true or false), was incorporated by the vendor as a special feature of the JHH system.

Admittedly the ability to capture the arrhythmia alarms and reset data does not provide any measure of the "false-negative" alarm performance, i.e., alarms that should have occurred and did not. On the other hand, ECG tape tests run prior to system acceptance and using the new ECG arrhythmia detection

software indicated that, at least with this particular system, false-negative alarms occurred so seldom that they did not represent a major consideration in the initial assessment of system performance. For the most part, the validity of that initial, if limited, observation has since been borne out. In particular, a few instances of a false-negative alarm response were noted by the staff very early in the evaluation effort. However, in all cases the application problem (e.g., inappropriate ECG lead placement) or design limitation leading to these occasional "missed alarms" was identified and corrected. Since that time, no significant incidents of this type have been reported either by the monitor watch or by other involved clinical personnel. Accordingly, attention has since been focused almost exclusively on the isolation and elimination of sources of false-positive alarms that interfere with the monitor watch function and detract from the utility of the system as a clinical tool.

It had been anticipated that the planned system shakedown and training phase would be completed by about October 1978, and that routine operation and data collection would be initiated immediately thereafter. However, before the end of September, complaints received from the clinical staff had risen to a level that left no doubt that all was not well. Although most of the automated functions performed as specified (and those few that did not were quickly corrected by the vendor), performance of the ECG arrhythmia detection and alarm system was found to be below expectations, with the number of false alarms far exceeding either the vendor's predictions or test results obtained during the ECG tape tests performed prior to installation in the intensive care areas. In fact, the arrhythmia detection function was deemed to be more of a hindrance than a help by just those CCU staff users who had been envisioned as the primary beneficiaries of the automated detection feature.

A sampling of performance results during October confirmed that the false alarm rate was, indeed, high enough to seriously impair system use in the detection of real ECG arrhythmias, with the ratio of false-positive-to-total alarms ranging from about 0.65 to 0.9. Therefore, rigorous procedures were inaugurated in an effort to quantify more precisely the performance of the system under realistic operating conditions, to identify contributing factors not properly reflected in the ECG test tape, and, finally, to correct the causes of the unacceptably large number of false alarms then occurring.

During the last quarter of 1978, arrhythmia alarm performance data were collected continuously along with selected segments of patient ECG data for subsequent off-line study. Using these data, it was soon found that a significant fraction of the erroneous alarms was caused by excessive patient motion and other signal artifacts that occurred much more often and were of much greater severity than in the ECG test recordings. Fortunately, the desirability of minimizing errors resulting from that effect had earlier been recognized by the system vendor, and preliminary efforts leading to a software change to alleviate the problem were already under way even though the severity of the problem was not then fully recognized. Therefore, an intensive effort to incorporate the change was initiated by the vendor. This resulted in an important revision to the arrhythmia detection algorithm that sharply reduced the level of false alarming without simultaneously causing any apparent reduction in arrhythmia detection sensitivity. The change was completed and ultimately incorporated as a revision to the computer software program in January 1979.

Data collected over the next two months demonstrated that the false alarm rate had been reduced by about a factor of two. More importantly, involved staff members indicated that the observed rates, while still higher than desired for certain types of patients, were on the average no longer high enough to preclude use of the system in the monitoring of patient ECG arrhythmias.

Thus, with the incorporation and validation of that essential modification to the operational software, the long-term collection of data was finally initiated in April 1979 to assess the performance of the automated system features and their impact during routine system use in support of the care and treatment of critical-care patients.

As noted earlier, the capability to record automatically each arrhythmia-system detection alarm, together with the associated manual true or false response by the assigned monitor watch, had already been incorporated as an integral part of the operational system software. Initially, the results were available only as printed summaries. With the initiation of the long-term evaluation, an additional capability was included to record the alarm summaries on digital tape cassettes for subsequent processing and analysis. Special data processing software was developed at APL to facilitate the selective retrieval and analysis of the alarm data as a function of time period, care area, frequency of specific classes of arrhythmia alarms, length of patient stay, etc.

During the 18 months after completion of the APL data analysis software program, JHH alarm performance data were continuously collected and accumulated in the data base. Results were regularly reviewed and updated throughout the data collection effort to

1. Provide a realistic assessment of the arrhythmia alarm performance achieved in a typical intensive care environment;
2. Identify areas where specific improvement in performance could materially enhance overall system utility as an adjunct to patient care; and
3. Validate the specific results achieved by any new software revisions or other changes introduced during the evaluation period and identify the overall performance impact of such changes.

Since the initiation of the effort, analysis of the collected alarm data in combination with suggestions and other information provided by the clinical staff (e.g., selected ECG strips and taped ECG segments for observed "problem cases") resulted in a number of improvements to the arrhythmia algorithm. These, together with some changes to improve display and control formats and to enhance the capability for automatic storage, recall, and editing of alarmed-on ECG complexes, were reflected in a series of software revisions that have been introduced at intervals of one to three months. Although some of these revisions significantly changed system operational features or arrhythmia detection characteristics, others involved only minor changes to correct specific software bugs observed during use. However, in all cases, significant new features were introduced singly to allow a meaningful "before and after" comparison of results.

Since installation of the systems at JHH, several improved operational aids that were originally included only at JHH and one other experimental location were announced and made available by the vendor as standard options for inclusion in currently marketed systems. These include the pressure analysis and the ECG storage, recall, and edit capabilities. On the other hand, the new detection algorithm has so far been incorporated for evaluation only at JHH and, for a brief period, at one other location; the arrhythmia detection program has not yet been released for general use.

Throughout the evaluation effort, APL data analyses and vendor software refinement activities were directed almost exclusively toward the elimination of sources of erroneous ECG arrhythmia alarms that interfered with the monitor watch function and that otherwise detracted from system usefulness as a tool for the evaluation of patient status. As this effort progressed, data analyses were also focused more and more on the CCU area, where arrhythmia detection and related features are of special interest and potential value in patient care.

SYSTEM PERFORMANCE

As the study progressed, it became increasingly evident that most false alarms were being produced by just a small fraction of the patient population. This is illustrated in Fig. 3 in terms of the relationship between the fraction of false alarms and the cumulative fraction of patients responsible for their production. For example with this particular software revision (Number 9, covering the period May 2 to September 4, 1979), 50% of the patients produced only about 5% of all false alarms. Conversely, about half of the false alarms resulted from only about 7% of the patients.

As shown in Fig. 3, a similar relationship was also evidenced for the fraction of alarms reset by the monitor watch as "true." However, unlike the false-alarm distribution, the larger true-alarm numbers do not result from limitations in the arrhythmia detection algorithm. Rather, they are regarded as indicating clinically significant variability in patient ECG patterns and reflect the system effectiveness in properly identifying this variability.

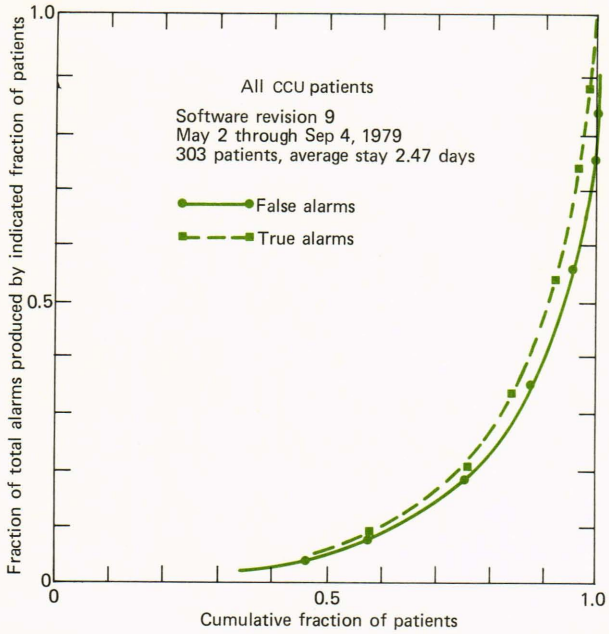

Fig. 3—The majority of patient ECG arrhythmia alarms were produced by a small but highly variable segment of the patient population.

True- and false-alarm relationships of this type were more or less expected on the basis of past experience; however, the wide disparity between the fraction of both true and false alarms and the fraction of patients producing these alarms had not been anticipated. Accordingly, efforts were initiated within the unit to identify the specific patients producing most of the observed alarms.

As a result of these initial observations, alarm data were compiled separately in subsequent analyses both for those patients who exhibited average alarm rates of less than 50 false alarms per day and for patients who exhibited less than 50 true alarms per day. Using these data, the system performance trends for the two "well-behaved" patient groups could be examined and compared with results obtained for the overall patient population. In addition, because the reduction of false-positive alarms was a continuing objective of the effort, special steps were undertaken to identify those patients who produced excessive numbers of false alarms and to collect sample tracings and tape recordings of the associated ECG sequences for subsequent off-line analysis.

Since the incorporation of software Revision 9 in May 1979, the combined data analysis and computer algorithm refinement effort was carried forward on a continuing basis. Software revisions were developed and introduced at intervals ranging from one to three months as important sources of false alarming were identified and isolated and the detection algorithms were appropriately refined. The most recent software

revision (Number 18) was installed in April 1980. Within this series, only four or five revisions actually involved a significant change to the ECG arrhythmia detection algorithm itself. The remainder were associated either with changes in the software operating system or with refinements to improve display formats and control functions.

In the course of the data collection and analysis effort, some unanticipated observations began to emerge as more data were analyzed. Because the average monitored period per patient was approximately three days and 70 to 90% of the available beds typically were occupied, it had been assumed that the effect of individual patient variations would be effectively suppressed by computing the average alarm rates for the units over an interval of 10 to 15 days. However, it was soon noted that variations in both the average true and average false alarm rates (alarms per patient day) were much more pronounced than was initially anticipated. To better quantify this effect, average true and false alarm rates per patient day were computed for each 15-day segment both for all CCU patients and for the two previously identified patient subgroups. Results of these computations are presented in Fig. 4 together with time periods and computer software revision numbers associated with each computed data point.

In the published literature, the ratio of true-to-total alarms obtained from one or another test set of sample ECG's by currently marketed arrhythmia detection systems is often cited as a measure of performance accuracy. Accordingly, the magnitude of this measure and its variations were also monitored throughout the performance assessment and system refinement process. The resulting statistics are illustrated graphically in Fig. 5, which indicates that, in the main, this measure of performance had indeed improved with successive refinements to the arrhythmia detection software algorithm. Thus, the expected result of a progressive reduction in false alarms due to software improvements would appear to have been achieved. No data from Revisions 10

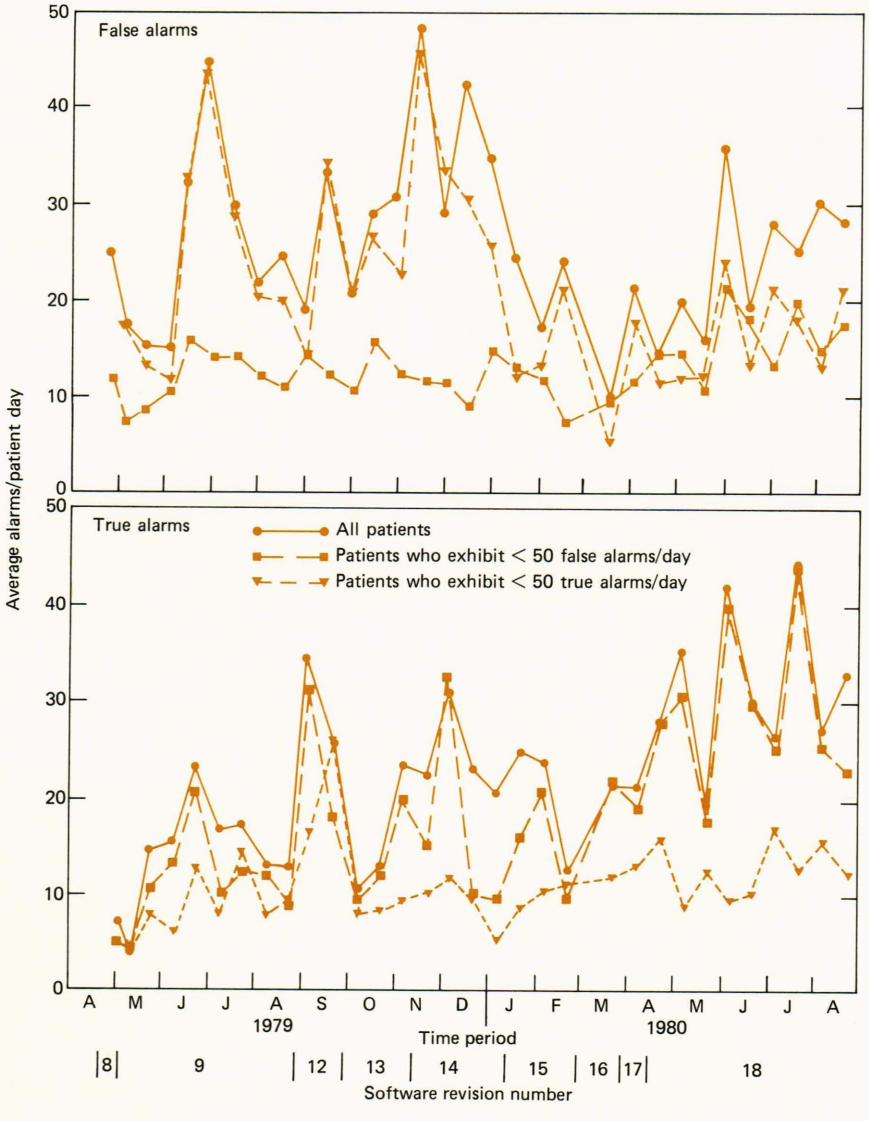

Fig. 4—Unexpectedly large variations were observed in computed biweekly averages for both false and true ECG arrhythmia alarm rates. The average true alarm rate for the total patient population increased over the evaluation period while the false alarm rate remained relatively fixed.

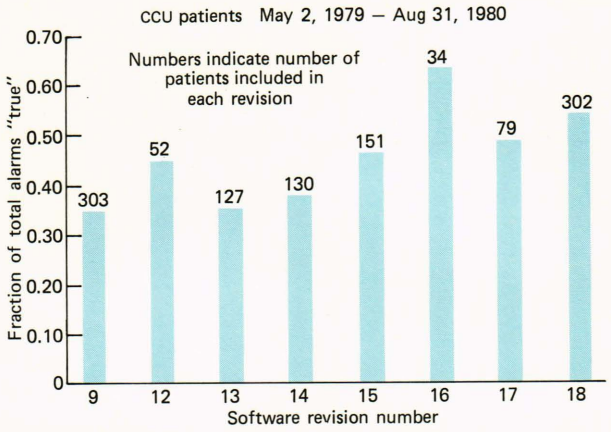

Fig. 5—The overall arrhythmia detection accuracy was improved by revisions that were introduced to eliminate specific causes of false alarms.

and 11 were considered in the evaluation because both revisions contained programming errors that led to their replacement in a matter of days.

However, as the data analyses progressed, separate assessments of false and true alarm rates over time began to cast doubt on this simple, albeit quite logical, conclusion. Finally, as a result, a series of formal statistical analyses was performed using as input the data presented in Fig. 4.

For this purpose, a linear regression was performed for each of the six data sets presented in Fig. 4. The results of this analysis are summarized in Table 1. In this formulation, "p" (probability) values approaching 1.0 indicate that the data points, with high probability, could have been generated by a purely random sequence of numbers; small values indicate that generation via this mechanism is highly unlikely, thereby confirming the statistical validity of the linear regression fit to the data. For example, $p = 0.001$ indicates the likelihood of such a random occurrence to be only 1 in 1000.

The principal conclusions to be drawn from Table 1 are the following:

1. When the total patient population is considered, no statistically significant reduction in the overall average false-alarm rate occurred over the data collection period. However, if only the "well-behaved" patient group that exhibits less than 50 true alarms per day is considered, a sizable decrease in false-alarm rate is observed. On the other hand, the false-alarm rate rises slightly over the data collection period for the restricted patient group that exhibits less than 50 false alarms per day.
2. A large and statistically significant increase in true alarm rates is evidenced both for the total patient population and for either of the two subgroups. As might be expected, this increase is less dramatic for the subgroup that includes only those patients who exhibit the limited number of true alarms.
3. For the restricted patient population subgroup that exhibits less than 50 false alarms per day, false alarm deviations about the linear regression line are sharply reduced relative to the total patient population; true alarm deviations are similarly reduced with the patient subgroup that exhibits less than 50 true alarms per day.
4. Exclusion of those patients who exhibit large numbers of false alarms does not materially affect the observed standard deviation of true alarm values about the regression line. Similarly, the false alarm deviation obtained for the reduced true alarm group is approximately equal to the deviation obtained for the full patient population.

It is seen from observations (1) and (2) above that, although revisions to the arrhythmia detection algorithm introduced during the evaluation effort were directed at eliminating sources of excessive false alarms, the average false-alarm rate for the total patient population actually changed little over the period. At the same time, the average true-alarm rate rose dramatically. Thus, the performance improvement reflected in Fig. 5 did not, as originally thought, result from a reduction in false-positive alarm activity.

Taken together, observations (3) and (4) and the data presented in Fig. 3 indicate that the large week-to-week changes noted in both true- and false-alarm rate averages computed for the total patient population derive from two relatively independent and numerically small, but highly variable, patient subpopulations that produce the majority of either false or true alarms. The number of patients in either of these segments is small although their alarms represent a large fraction of the total; thus a numerically small change in the size or a variation in the composition of either subgroup will appear as a large change in the observed number of alarms.

The unexpected trends in true- and false-alarm rates as summarized in (1) and (2) above appear reasonable only if some major change in operational environment or system application over the 16-month data collection period is assumed. And, in fact, it was subsequently determined that a significant change in the CCU patient profile did occur over the period. In particular, for reasons not directly related to this effort, the proportion of CCU patients exhibiting serious arrhythmia problems increased significantly during the study period while the fraction of patients with less complicated cardiac involvement correspondingly decreased.

The observed results are not inconsistent with the known change in patient population because complicated arrhythmia patients will, on the average, generate a considerably larger number of true alarms than a population composed of relatively uncomplicated cardiac patients. At the same time, this more complex patient group could be expected to produce a greater number of false alarms due to the variations in ECG waveform and timing commonly

Table 1

SUMMARY OF LINEAR LEAST SQUARES STATISTICAL ANALYSIS OF CCU PATIENT POPULATION

Software Revisions 9-18, 1178 patients, 3176 monitored patient-days, May 1979 – Aug 1980

		Average False (F) and True (T) Alarms per Day					
		Total Patient Population		*Patient Population A**		*Patient Population B**	
		F	*T*	*F*	*T*	*F*	*T*
Direction of linear regression fit over period		(not significant)	↑	→	↑	↓	↑
Range of linear regression over the period		26–25	12–33	11–15	9–30	26–16	8–14†
Standard deviation of data about linear regression line		8.8	6.9	2.8	7.4	8.7	2.8†
Test values (t) for linear regression fit (p)	(t)	0.275	5.03	2.93	4.70	1.84	3.37
	(p)	>0.5	<0.001	<0.01	<0.001	<0.10	<0.005†

**Population A*: patients who exhibit <50 false alarms per day; *Population B*: patients who exhibit <50 true alarms per day.

†With a single outlying data point deleted. The overall character of the linear regression fit was little affected by the inclusion of the isolated point (Sept 15-30, 1979) in the analysis. However, confidence in the validity of the fit was materially reduced, with the resulting value of p near 0.10. Accordingly, because the difference between that data point and the corresponding linear regression value was greater than 4 times the standard deviation (3.98), the point was omitted from the regression data summarized above. No explanation has thus far been put forward that would account for the anomalous results obtained during this particular data collection segment.

exhibited by them. Indeed, the average false-alarm rate is seen to improve somewhat (from 26 to 16 alarms per day) over the data collection period when only those patients who exhibit less than 50 true alarms per day are considered. Therefore, it is hypothesized that although a real improvement in the inherent false alarm capability of the system was achieved by the successive revisions to the arrhythmia algorithm, the improvement was accompanied by a corresponding increase in the number of complex and changing ECG patterns. Thus, in the aggregate, little if any net change in false-alarm rate was observed for the entire CCU patient population over the 16-month system refinement effort.

SYSTEM ACCEPTANCE AND APPLICATION BY CCU STAFF MEMBERS

As was noted earlier, false-alarm rates in excess of 80% were regularly produced by the arrhythmia detection algorithm as initially incorporated in the patient monitoring system. Under these conditions, the system was deemed by the staff to be more of a hindrance than a help in the detection of patient arrhythmias. The "harassment" produced by recurrent false alarms was severe enough to interfere with the intended function of checking and reporting valid alarms for clinical action.

The initial revision, which reduced the tendency to cause alarms as the result of muscle noise, poor ECG lead attachment, and similar artifacts, decreased the number of false alarms by a factor of over two so that the overall average false-alarm rate was lowered to 65% from about 75 to 80%. At that point, clinical personnel indicated that, although considerable improvement was still needed, the system was "becoming useful." Moreover, the users then first observed that, in practice, a sizable fraction of the observed false alarms was usually produced by only a very few patients.

As illustrated by Fig. 5, no really dramatic improvement in average system alarm performance was achieved by the early revisions (9 to 13) introduced during the first six months of the data collection effort. Nevertheless, during the same period, acceptance of the system increased significantly. Both increased confidence in the system operation and a greater understanding of particular performance

limitations were regularly expressed by users during that period.

By the end of the evaluation period (August 1980), user acceptance of the system as an effective clinical tool had become general, and the system was being routinely relied on by the staff both for the monitoring of ECG's in the CCU and for the retrospective review of the status of selected patients. Nevertheless, as is illustrated by Table 1, no significant reduction in false-alarm rates had actually been achieved.

Interestingly, throughout the entire data collection period, clinical users generally seemed to feel that the system was producing fewer and fewer "false-alarm problems" as successive revisions were introduced. It would appear that the subjectively perceived level of "false-alarm harassment" was being reduced even though, as demonstrated by the resulting alarm data, the false-alarm rate for the overall unit remained relatively constant. Moreover, clinical personnel who were closely involved with the system application and data collection process over the entire period were generally unaware that a sharp increase in the number of true alarms had, in fact, occurred.

On the other hand, a small but continuing reduction in the fraction of patients producing large numbers of false alarms was identified during the data analysis effort. These results, which are presented in Table 2, appear to correlate with the qualitative assessments of system utility provided by the users; they suggest that acceptance of the arrhythmia detection system as a regular adjunct of patient care became more general after performance was improved so that a large fraction of the false alarms was being produced by only a very few identifiable patients. Observations at the central monitor stations suggest that, once these conditions are achieved, recurrent false alarms generally are of a few types at most. Thus, observing and checking them for validity by the monitor watch does not materially interfere with the continued monitoring of other patients. On the other hand, when the false alarms are more generally distributed over all monitored patients, the same total number of alarms appears to be much less easily tolerated.

These observations also suggest strongly that the fraction of patients producing large numbers of false alarms, or some similar measure, may be a better predictor of system effectiveness in the clinical environment than the measures (e.g., fraction of beats correctly identified) currently in vogue with system designers and suppliers.

Table 2

PERCENTAGE OF PATIENTS WHO PRODUCE LARGE NUMBERS OF FALSE ALARMS VERSUS SOFTWARE REVISION NUMBER

Revision Number	*Percentage of Patients Who Exhibit an Average False-Alarm Rate of*	
	>50 per Day	*>100 per Day*
5 (prior to correction of noise artifact problem)	55	29
7 (after noise artifact revision)	15	8
9 (start of performance evaluation)	15	7
18 (end of performance evaluation)	12	5

REFERENCES

[1]D. A. Frost, F. G. Yanowitz, and T. A. Pryor, "Evaluation of a Computerized Arrhythmia Alarm System," *Am. Cardiol.* **39** (1977).

[2]American Society for Hospital Engineering, *Arrhythmia Monitoring Systems*, American Hospital Association AHA Catalog No. 112.

[3]S. Corday, "Welcome Trends in Cardiology," *Reports from the American College of Cardiology Meeting* (1980).

ACKNOWLEDGMENTS—Selection, refinement, and evaluation of the computerized patient monitoring system involved the efforts and dedication of individuals from the Johns Hopkins Hospital and the system vendor — the Hewlett Packard Medical Products Division — and APL. The continuing support of several members of the JHH clinical staff is especially noteworthy: Dr. E. L. Nagel was responsible for overall direction of the effort; Dr. L. C. Becker actively participated in the identification of initial system objectives and selection criteria; Dr. B. Bulkley made available the support needed to collect the requisite performance data in the CCU area; and Ms. D. Zwarra, RN, provided day-to-day coordination of the data collection task.

Checkout, installation, and servicing of the monitoring equipment were the responsibility of the JHH Clinical Engineering Services Group, which assisted in the selection of initial design concepts. The group's supervisor, Mr. D. H. Gordon, also served as a member of the initial system definition and selection team and actively assisted in the performance evaluations during the early stages of this effort.

Special thanks are due Dr. W. H. Guier of APL for his continuing participation in the effort, from the earliest identification of system objectives through the system selection and data collection process. His assistance in the review and critical analysis of the collected system performance data and in development of the performance measurement criteria that were ultimately employed were indispensable.