Concurrent statistical evaluation during patient monitoring*

by S. T. SACKS, N. A. PALLEY and H. SHUBIN

University of Southern California School of Medicine
Los Angeles, California

and

A. A. AFIFI

University of California
Los Angeles, California

BACKGROUND

The Shock Research Unit, a specialized clinical research facility, has been developed by the University of Southern California's School of Medicine for the purpose of rendering intensive care to seriously ill patients. Included is a medium-sized digital computer with a real-time system which monitors the critically ill patient. In addition, the system is being used to study the underlying mechanisms of the disease process and for developing new techniques of evaluating and treating seriously ill patients.

The Shock Research Unit was started in 1962 and since that time approximately 1,000 patients have been treated at the unit. In mid-1963 a computer was obtained and a system developed for monitoring patients. In 1968 the Shock Research Unit obtained a third generation computer and applications programs for a much-enlarged system of real-time monitoring were written.

"The patient in circulatory shock is an example of the need for immediate response to clinical observations. Low blood pressure and reduced blood flow are characteristic of the patient in shock. He may become stuporous or comatose due to the inadequate circulation to the brain. His kidney may cease to function and his respiration may fail.

Assessment of the circulatory and respiratory status of such a critically ill patient requires measuring a number of variables: arterial and venous pressure; blood flow and volume; electrocardiogram; blood gases such as oxygen and carbon dioxide and blood constituents such as potassium. Repeated assessment of these variables is required since the critically ill patient is not in a steady state, but may undergo rapid and often unpredictable changes in status. A number of other patient monitoring facilities employing digital computers are currently in operation across the country. Two such units engaged in the simultaneous monitoring of multiple variables are at the Pacific Medical Center in San Francisco and the Latter Day Saints Hospital in Salt Lake City.

At the Shock Research Unit, a combination of sensors and transducers is used for measurement of vital signs and additional parameters of clinical importance, and the data is processed to derive numerical information helpful to the physician. Information which is supplied directly by sensors applied to the patient's body include eleven primary measurements and twenty-five parameters which are recorded and displayed with a frequency ranging from once a minute to once every twenty-four hours. All numeric and textual data are stored in an on-line patient file organized by type of data and by time. The data may be retrieved by the user through a bedside K/D and by current scheduled applications programs. Given this large amount of sequential data, the physician must have some means of combining it into a meaningful evaluation of changes in the patient's status. All present commercially available patient monitoring..."
systems depend on alarm limits which are manually set and may not be adaptable to the particular clinical situation. These alarms are based on absolute values of single variables and do not take into account relationships among variables. When these univariate alarm systems sound too frequently, the usual response of the person in charge is to set wider alarm limits. Such actions have no statistical basis and may be detrimental to patient care.

The availability of an on-line computer monitoring system allows the examination of many variables simultaneously as well as the observation of their interrelationships. The purpose of this paper is to describe the development and application of an automated screening procedure for evaluating physiological data obtained from continuous monitoring of critically ill patients. Three criteria, motivated by clinical experience, are used in the evaluation of the patient's data. The first is the absolute value of the monitored variables. These values contain the information necessary to determine the current status of the patient and to assure that the monitoring equipment is operating properly. The monitoring interval then may be modified appropriately. However, in a dynamic system the information of interest to the clinical staff is contained in the sequential changes in the measured variables. These may reflect sudden, unexpected changes in status or the expected responses to treatment and constitute the second criterion. Variations are sometimes more meaningfully interpreted relative to the absolute values of the variables. The third criterion, therefore, is the proportionate change in the monitored variables.

Many of the variables which are routinely monitored in a critically ill patient are highly correlated and the clinical significance of a given measurement becomes more apparent when examined in the context of the remaining measurements. For example, Figure 1 illustrates a scatter diagram of a typical random sample of simultaneously measured systolic (SP) and diastolic (DP) pressures. The data tend to accumulate in an elliptical region with the highest density near the center. While the vertical and horizontal dashed lines determine the 95 percent confidence intervals for SP and DP individually, the elliptical region includes the same region for the pair of measurements. Point A illustrates a measurement which lies within the individual normal limits but exceeds the bivariate limits, because low systolic pressure is combined with high diastolic pressure. Thus normalcy is determined by knowing both distance and direction from the mean. The region inscribed by the ellipse may be expressed by the inequality

$$D^2 \leq C$$

where $C$ is a constant and $D^2$ is the Mahalanobis distance for the two variables SP and DP. This Mahalanobis distance may be computed for any number of variables. If $Y$ is a vector of $r$ observations, $\bar{Y}$ is the mean vector, and $S$ is the covariance matrix, then

$$D^2 = (Y - \bar{Y})S^{-1}(Y - \bar{Y})$$

Because $D^2$ incorporates interdependencies among the variables, unusual combinations of variables or changes result in abnormally large $D^2$ values, even though the variables or changes when regarded individually may fall within normal limits.

Using data sampled from patients previously monitored in the Shock Research Unit, these ideas provided the basis for a system to respond in real-time to changes in patient status. The responses of the system include:

1. Informing medical personnel of unusual changes in the set of monitored variables.
2. Selecting data for display.
4. Running of special analysis programs.

These responses will be discussed in detail below.

**Development of method**

Forty-one patients were chosen randomly from the 750 patients monitored in the SRU up to August, 1967.
Only those patients who were observed for at least four hours were included.

Seven variables were selected from the many which are routinely monitored: systolic pressure (SP), diastolic pressure (DP), mean arterial pressure (MAP), mean venous pressure (MVP), rectal temperature (RT), heart rate (HR), and respiratory rate (RR). Each patient's record was examined and that four-hour period which had the least missing data was chosen. Measurements on sets of these variables were recorded on each patient at five-minute intervals over the selected four-hour period. Estimates for any missing data, which accounted for less than 5 percent of the total data, were made by interpolating between the nearest recorded values before and after the missing observation. From these absolute value sets the following were calculated: .5-, 10-, 15-, 30-, and 60-minute changes, proportionate .5-, 10-, 15-, 30-, and 60-minute changes.

Figure 2 illustrates the successive changes in a single variable over time. It should be noted that this choice of intervals was made in order to account for both long-term and short-term variations. An example of the calculation of 5-minute changes and proportional changes in a set of variables is illustrated in Table I. The value of the Mahalanobis $D^2$ was calculated using the appropriate sample mean vector and covariance matrix for each vector of absolute values, x-minute changes and proportionate x-minute changes ($x = 5, 10, 15, 30, 60$). In the remainder of this paper frequent reference will be made to these eleven types of measurements, namely, absolute values, the five x-minute changes, and the five x-minute proportionate changes.

This procedure yielded eleven distributions of $D^2$. Percentiles necessary for setting monitoring limits were then calculated for each of these $D^2$ distributions.

In monitoring a critically ill patient, it is inevitable that, on occasion, values for one or more variables are missing. This may occur if a measurement is taken while a catheter is being flushed or if the EKG leads are accidentally disconnected or simply because monitoring of all variables is not started simultaneously. A preliminary screening procedure will exclude such invalid measurements. Consequently, tables of $D^2$ for incomplete vectors of observations are necessary. It is possible to empirically derive and store the 1309 tables of $D^2$ percentiles necessary to handle all combinations of missing variables. However, in order to minimize storage requirements and allow the system to operate in real-time, an alternative procedure was developed as follows.

If the population mean vector and covariance matrix were used in computing $D^2$, and if the observation vector were normally distributed, the distribution of $D^2$ would be that of a Chi-square variable. Since the variables under consideration are not exactly normally distributed and sample estimates were used instead of population parameters, the empirical distribution of $D^2$ deviates from that of Chi-square. However, they do remain similar in shape. Thus it was hypothesized that the ratio of a percentile of the Chi-square distribution to the corresponding empirical percentile of $D^2$ for a given type of measurement is independent of the number of components in the vector of measurements. That is, for absolute values, x-minute changes and proportionate x-minute changes, we assume that:

\[
p\text{th percentile of } D^2 \text{ distribution based on } r \text{ components} = \frac{p\text{th percentile of Chi-square distribution with } r \text{ d.f.}}{}
\]

**TABLE I—An Example of 5 Minute Changes and Proportional Changes in the 7 Variables Set**

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Value at Time $t$</th>
<th>Value at Time $t - 5$</th>
<th>5-minute change at Time $t$</th>
<th>5-minute proportionate change at Time $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>122</td>
<td>122</td>
<td>0</td>
<td>0. = 0/122</td>
</tr>
<tr>
<td>DP</td>
<td>80</td>
<td>87</td>
<td>-7</td>
<td>-0.088 = -7/80</td>
</tr>
<tr>
<td>MAP</td>
<td>94</td>
<td>98</td>
<td>-4</td>
<td>-0.043 = -4/94</td>
</tr>
<tr>
<td>MVP</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0. = 0/1</td>
</tr>
<tr>
<td>RT</td>
<td>37.7</td>
<td>37.8</td>
<td>-0.1</td>
<td>0.003 = 0.1/37.7</td>
</tr>
<tr>
<td>HR</td>
<td>98</td>
<td>100</td>
<td>-2</td>
<td>-0.020 = -2/98</td>
</tr>
<tr>
<td>RR</td>
<td>21</td>
<td>22</td>
<td>-1</td>
<td>-0.048 = -1/21</td>
</tr>
<tr>
<td>$D^2$</td>
<td>12.9</td>
<td>8.5</td>
<td>11.8</td>
<td>43.1</td>
</tr>
</tbody>
</table>

Region: B A B C
Figure 3—Monitoring regions with alarm limits based on the empirical distribution of $D^2$

$p$th percentile of $D^2$ distribution based on 7 components

\[ = \frac{p \text{th percentile of Chi-square distribution with } 7 \text{ d.f.}}{\text{for } r = 2, \ldots, 6}. \]

From this assumption, approximate empirical percentiles for the $D^2$ distribution based on any number of components were generated from the $D^2$ distribution percentiles based on seven components.

Each $D^2$ distribution is divided into regions designating the degree of normalcy of a set of measurements as shown in Figure 3. These limits are based on the empirical distributions of each set of values in the base sample. In the figure the limits are taken as the 80th, 95th, and 99th percentiles. Region A may be classified as the normal region, containing values which are expected to occur 80 percent of the time. Region B, the moderately abnormal region, contains the values expected to occur between the next 20 percent to 5 percent of the time. Region C, the abnormal region, includes the 95th to 99th percentiles, and Region D may be interpreted as the highly abnormal region since it contains the extreme values which would be expected to occur only one percent of the time. Sets of three limits are computed for 0, 1, \ldots, 6 missing variables. The total number of tables to be stored is thus reduced from 1309 to 77, and with 3 limits per table, only 231 values are stored (11 change vectors \times 3 region limits \times 7 possible numbers of missing values). In addition eleven $7 \times 7$ covariance matrices based on the sample data are stored.

Note that the original choice of the three percentile limits is arbitrary and may be modified as experience is gained with the system.

System algorithm

The basic algorithm for this system comprises the following steps:

1. Each five minutes, the most recent set of observations is retrieved from the patient file.
2. The patient file is searched sequentially for the measurements taken 5, 10, 15, 30 and 60 minutes previously and also for alarm information recorded at those times. For those variables having a valid present and past value, changes and proportionate changes are computed. The appropriate stored covariance matrix, or a reduced matrix corresponding to the variables present in each change vector, is then used to calculate $D^2$. In some cases all components necessary for computing $D^2$ may be missing.
3. For each of the computed $D^2$ values, the appropriate set of three region limits is retrieved from the tabled values on the disc file. The index of the set of limits to be retrieved is determined by the change vector being considered (1, \ldots, 11) and its dimension (1, \ldots, 7).
4. Using the region limits, a category is assigned to each of the available $D^2$ values.
5. The identity of the most extreme $D^2$ and the category to which it is assigned are stored in the patient file.
6. Appropriate system action is taken as described below.

Uses of the system

The action taken by the patient monitoring system in response to the computed $D^2$ depends upon the most extreme category detected. Any $D^2$ value falling in the B, C or D regions will result in a signal to the ward staff. The signal is coded to indicate the degree of abnormality. A red and a yellow signal light and a chime, under computer control, are mounted near the status display. The chime signals a $D^2$ value detected in any of the three regions. In addition, the red light accompanies region D while the yellow light accompanies region C.

Some measurements such as heart rate and respiration rate, are usually read from the output of preprocessors which derive the information from the raw data. When an alarm occurs, the digital analysis program reads the original wave-form signal directly to verify the abnormal value encountered. The EKG analysis program, which may be run on demand, is also triggered in response to an alarm.
In many cases, the cause of the alarm is immediately obvious by noting the values on the status display, as shown in Figure 4. If the physician wishes to know in detail the values causing an alarm, he may call a program on the keyboard display which reports the current and previous values of the variables used in computing $D^2$. As illustrated in Figure 5, this program indicates the time over which the extreme changes have taken place and the magnitude of the changes. The physician may then indicate the probable cause of the alarm. The cause may be a change in the normal course of the patient’s progress or a specific treatment such as medication, fluid infusion, or adjustments made to the respirator. Changes from such causes, although resulting in an alarm, do not necessarily reflect emergent situations. In addition, the alarm may have been caused by an artifact in one of the monitored variables. This may occur if a catheter becomes clogged or is inadvertently flushed while the signal is being read or if EKG leads become dislodged. This information is stored in the patient file and at the same time the signal light is turned off.

Summaries of alarm information are stored in the patient file. These summaries contain the following information: the time of the alarm, its category and the specific change or proportionate change causing the alarm. In addition, if the ward staff responded to the alarm through the keyboard display, the summary will contain the time of the response and may thus contain a flag indicating that the cause of the alarm was an artifact and the particular variable or variables involved. If the information was supplied by the physician, the clinical event related to the alarm condition is also coded.

The alarm summaries are utilized by the patient monitoring system in a variety of ways. As indicated in Step 2 of the system algorithm, the patient file is searched so that changes may be computed. Summaries of alarms stored at the same time as any of the five previous sets of values are also retrieved. If the summary indicates that any value in the set used for calculating the $D^2$ was the result of an artifact, that value is deleted from computation. This prevents a single artifact from causing several alarms as that set of values is successively retrieved 5, 10, 15, 30 and 60 minutes later.

During the patient’s stay, the physician often reviews sets of data on the bedside keyboard display. Since on any display frame only six sets of simultaneous measurements may be viewed, as shown in Figure 6, it is desirable to eliminate redundant observations. This may be achieved by displaying only those sets of data falling in abnormal regions. In the same way the permanent hardcopy record, which is printed upon patient discharge, can be specified to contain only significant information. Under conditions of intensive monitoring where measurements are made every minute, this may represent a valuable reduction in the bulk of the hardcopy record to be reviewed. Those sets of variables containing values previously identified as artifacts may be deleted from both the keyboard

---

**Figure 4**—Bedside status display showing values of the monitored variables for two patients.
Figure 6—Hemodynamic data summary showing six sets of values of seven variables

display presentation and the hardcopy output. This procedure is also used with the storage oscilloscope at the bedside which displays a trend plot of monitored variables. By referring to the alarm summaries the program which produced this display deletes invalid data points from the plot.

In order to have the patient file accessible in case of system failure, teletypes, remote to the ward, produce hardcopy output containing the latest measurements. While the status display is automatically updated every 30 seconds, the teletype, because of its low speed, cannot be updated at the same rate. In order to make efficient use of the teletypes and fulfill their function of preserving data of interest, the teletype records are updated every half-hour or whenever abnormal values are encountered.

SUMMARY

The alarm system described in this paper both depends on and augments the capabilities of the digitally controlled patient monitoring system. It utilizes multivariate techniques to compute statistics which are sensitive to relationships among the monitored variables. The degree of abnormality of a computed statistic is evaluated in terms of empirical distributions. These distributions were derived from a population of critically ill patients similar to that being presently monitored by the system. The actions taken by the patient monitoring system in response to the alarm depend on the severity indicated by the alarm. The occurrence of the alarm and the cause, if known, become part of the patient's file and are accessible to applications programs involved in data display. The stored summary of the alarm information assists the clinical staff in case review, and provides a basis for evaluating and modifying the alarm system itself. Such modifications might include the redefinition of the percentiles which define the alarm categories or even the number of such categories. Extensions of this alarm system might enable a small centrally located digital processor to evaluate sets of data from a number of analog monitoring modules at various bedside.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Doctor Max Harry Weil, Director of the Shock Research Unit, whose able guidance and support made this work possible. We would also like to thank Doctor William Rand who, during his employment at the Shock Research Unit, contributed greatly to the collection of the data in the preliminary discussions leading to the formulation of the system. Our thanks also go to Miss Cecilia Pasos for typing the manuscript.

REFERENCES

1 M A ROCKWELL H SHUBIN M H WEIL
   Shock III: A computer system as an aid in the management of critically ill patients
   Communications of the ACM Vol 9 No 5 May 1966
2 D H STEWART D E ERBECK H SHUBIN
   A computer system for real-time monitoring and management of the critically ill
   AFIPS-Conference Proceedings Vol 33 December 1968
3 Ibid
4 J J OSBORN J O BEAUMONT J C RAISON
   Measurement and monitoring of acutely ill patients by digital computer
   Surgery Vol 64 pp 1057-1070 December 1968
5 T A PRYOR H R WARNER
   Time-sharing in biomedical research
   Datamation Vol 12 pp 54-63 April 1966
6 A A AFIFI W RAND H SHUBIN M H WEIL
   A method for evaluating changes in sets of computer monitored physiological variables
   Submitted for publication
7 T W ANDERSON
   Introduction to multivariate statistical analysis
   John Wiley and Sons New York 1958