A Study of Various Estimators for the Derivation of Quality Control Procedures from Patient Erythrocyte Indices

Article *in* American Journal of Clinical Pathology · May 1974

DOI: 10.1093/ajcp/61.4.473 · Source: PubMed

CITATIONS
90

READS
572

4 authors, including:

Brian S. Bull
Loma Linda University
162 PUBLICATIONS **4,743** CITATIONS

SEE PROFILE

A Study of Various Estimators for the Derivation of Quality Control Procedures from Patient Erythrocyte Indices

B. S. BULL, M.D., R. M. ELASHOFF, PH.D., D. C. HEILBRON, PH.D., AND J. COUPERUS, M.D.

Departments of Pathology and Laboratory Medicine, Loma Linda University, Loma Linda, California 92354, and Department of Clinical Pathology and Laboratory Medicine, San Francisco Medical Center, UC San Francisco, San Francisco, California

ABSTRACT

Bull, B. S., Elashoff, R. M., Heilbron, D. C., and Couperus, J.: A study of various estimators for the derivation of quality control procedures from patient erythrocyte indices. Am. J. Clin. Pathol. 61: 473–481, 1974. The derivation of quality control data from patient laboratory values presupposes, first, that the distribution of those values is stable over long periods of time; second, that it is possible to estimate accurately the mean of the entire population from a small sample of that population. That the first presupposition is true for erythrocyte indices has been shown elsewhere. The second presupposition is examined in this paper. Six estimators were tried in turn on two data batches, each consisting of 480 patients from two university hospitals. An estimator which smoothed the data (decreased the effect of outlying values) and at the same time showed the characteristics of a moving average was found sufficiently effective to serve as the major basis for quality control of an automated whole blood analyzer. (Key words: Quality control; RBC indices; Automated analytical devices; Calibration control; Patient values.)

WHEN CALIBRATED CORRECTLY and operating properly, automated analytical devices are a boon to the hematology laboratory. The tremendous capacities of such devices as the Model S Coulter Counter and the Technicon Hemalog permit the production of several thousand test results each working day. When calibration is incorrect from the start or drifts during the day, disaster results. The tremendous test throughput of these machines thus renders adequate and frequent calibration control procedures an absolute necessity.

Received October 15, 1973; received revised manuscript November 23, 1973; accepted for publication December 11, 1973.

The ideal calibration control method should give independent, confirmatory evidence that the machine was correctly calibrated initially, and it should be able to identify all types of calibration loss as soon as they occur. Calibration control should therefore be continuous. Not unexpectedly, the ideal is elusive and unobtainable by any past or present method. Calibration control of the Model S Coulter Counter is most frequently accomplished by analyzing a commercial whole blood standard for which the values have been provided and setting the machine accordingly. Errors arise in this method when (1) the blood standard was inaccurately analyzed at its origin, (2) changes have

occurred in the blood standard in transit or at the laboratory (stability problems), or (3) technicians handle data from analysis of the blood standards with bias. Thus, whole blood standards do not give error-free results and are expensive to use. Certified standards exist only for the hemoglobin determination.

Loss of calibration may occur in two ways. The first is an abrupt change in a measured parameter, usually as a result of mechanical or electronic breakdown. This may be either transient as a result of some temporary blockage of an aperture or a reagent line, or it may be permanent. The second type of calibration loss occurs more insidiously over a period of days or weeks and is presumably caused by electronic drift.

To detect these two types of calibration loss, a variety of quality control methods have been proposed. Primary or secondary standards can be interspersed between patient samples and the results analyzed statistically for calibration loss. This approach is expensive to implement and maintain, and might be at least partially vitiated as a control procedure by the technician's knowledge of what result is expected. For details of this method, called the "reference sample" method, see, for example, Britton, Brecher and Johnson.[5]

Another method derived by Britton and associates[4] uses specimens stored from one day to the next. Five blood samples which have been refrigerated overnight are analyzed. If the mean value of each parameter agrees with the corresponding measurement of the previous day, it is assumed that the calibration of the analyzer has not changed. Use of this method might result in a failure to detect calibration loss for 12 or more hours. It may also result in a considerable amount of wasted effort since recalibration of the machine is most often unnecessary.

In this paper is explored the usefulness of another group of methods of calibration control for the hematology parameters based upon the internal consistency of patient values. This approach has been criticized by Amador, Hsi and Massod[1]; however, Begtrup and co-workers[3] and Dixon and Northam[6] find it useful for certain tests.

The Basis of Using Patient Values for Quality Control

Considerable experience, including our own data, has shown the approximate constancy of the distribution of mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and mean corpuscular volume (MCV) values from day to day and week to week in medium- to large-sized hospitals. The reason for this finding is that the vast majority of patients in a general hospital, including anemic patients, have normal erythrocyte indices. In particular, such summary statistics as the mean, median, and percentage of patients lying within the normal range show slight daily, weekly, or monthly changes. Furthermore, these daily, weekly, and monthly distributions for the three measurements look approximately symmetrical about the respective means.

These facts suggest that periodic determination of the mean MCV, MCH, and MCHC of the patient population should make it possible not only to determine when calibration has been lost (the mean indices determined by the analytical device will no longer agree with the population means), but also, hopefully, to quantitate the departure and thus aid in recalibration. Implementation of such an approach to calibration control requires:

(1) accumulation of N patient values for MCV, MCH, and MCHC;

(2) estimation of a mean for each of these indices;

(3) analysis of the extent of the dif-

ference between the known population mean (μ) and the newly computed estimate to determine whether routine laboratory operation should be interrupted for machine inspection.

Value judgments are involved at each step of this process:

N should be as small as possible so that calibration rechecks will be available frequently. N should be as large as possible to avoid the problems inherent in estimating the mean of large populations from small samples. The latter problem is the more pressing because the distributions under study are not Gaussian, having increased numbers of both high and low values due to the presence in the population of hematologically sick patients. As N becomes smaller, some measure must be taken to decrease the contribution of the occasional outliers included in the sample.

The significance of any difference between the known population mean μ and that calculated from the sample N becomes greater as N becomes larger. Conversely, as N becomes smaller, significance diminishes and eventually disappears entirely.

What is desired, of course, is a calibration control procedure that minimizes false positives and yet is able to detect either a slow steady drift or an abrupt loss of calibration as rapidly as possible. The latter two requirements are in conflict. A detection procedure that is sensitive to transient changes in a measured parameter will have high variability and a large number of false positives—unnecessary machine inspection will be unavoidable. A detection procedure that emphasizes sensitivity to slow, steady drift and that is more highly damped than simple calculation of the sample mean $\overline{X}$ must give up some sensitivity in identifying transient changes and must delay the recognition of abrupt changes due to mechanical or electronic breakdown.

Table 1. Population Means

	Lab A	Lab B
Mean corpuscular volume	92 cu. μ	90 cu. μ
Mean corpuscular hemoglobin concentration	33%	33.5%

The means of batches of MCV and MCHC data from routine analysis of patient blood samples were determined by using each of six estimators in turn. The characteristics of each estimator and its likely utility for aiding quality control in a routine hematology laboratory form the substance of this report.

Materials and Methods

The data analyzed consisted of 480 consecutive values from each of two large university hospital laboratories, Lab A and Lab B. Both laboratories utilized Model S Coulter Counters; both machines were "in control" during the three days covered by the study as judged by independent reference whole blood standards.

The data were analyzed on the assumption that a batch mean which differed from the expected mean by ≥3% should lead to a recheck of machine calibration. The population means (μ) based on long-term experience at these laboratories are given in Table 1. (All three erythrocyte indices were studied, but only the MCV and MCHC are presented as the comparative performances of the estimators for MCV and MCH data were nearly identical.)

The following estimators of central tendency were investigated (estimators numbered 2, 3, and 6 proved significantly more useful than the rest).

(1) $\overline{X}$, the sample mean. This is the sample mean of the i^{th} batch, $\overline{X}_i$. This measure can be highly variable in non-Gaussian populations and makes no use of previous batch means.

(2) $\overline{X}_{MA}$*, a moving average mean.* This type of measure has been introduced by others working in quality control to make use of previous data on batch means. Let $\overline{X}_{MA,i}$ denote the moving average after i batches are observed. Then

$$\overline{X}_{MA,i} = r\cdot\overline{X}_i + (1 - r)\overline{X}_{MA,i-1}$$

where $\overline{X}_{MA,0} = \mu$ and $0 < r \leq 1$.

(3) $\overline{X}_B$*, an alternative moving average.* Suppose that an estimate of the true mean after i batches have been observed takes the form

$$\overline{X}_{B,i} = (2 - r)\overline{X}_{B,i-1} + r\cdot d$$

$\overline{X}_{B,i-1}$ is the estimator after $(i - 1)$ batches and d is some "signed function" of the patient values in the i^{th} batch. X_j is j^{th} value in batch i. For example,

$$d = \frac{1}{N}\sum_{j=1}^{N}(X_j - \overline{X}_{B,i-1})$$

$$= \sum_j \frac{\operatorname{sgn}(X_j - \overline{X}_{B,i-1})\cdot\left|X_j - \overline{X}_{B,i-1}\right|}{N}$$

which, if $\overline{X}_{i-1} = \overline{X}_{B,i-1}$ and $r = 1$, then

$$\overline{X}_{B,i} = \overline{X}_i$$

$$\operatorname{sgn}(X_j - \overline{X}_{B,i-1}) = +1 \text{ if } X_j > \overline{X}_{B,i-1}$$
$$= -1 \text{ if } X_j < \overline{X}_{B,i-1}$$
$$= 0 \text{ if } X_j = \overline{X}_{B,i-1}$$

A variety of modifiers was experimented with. The following expression performed best:

$$d = \operatorname{sgn}\left(\sum_{j=1}^{N}\operatorname{sgn}(X_j - \overline{X}_{B,i-1})\left|X_j - \overline{X}_{B,i-1}\right|^P\right) \times \left(\sum_{j=1}^{N}\frac{\operatorname{sgn}(X_j - \overline{X}_{B,i-1})\left|X_j - \overline{X}_{B,i-1}\right|^P}{N}\right)^{1/P} \quad \text{where } P = \tfrac{1}{2}.$$

Thus, $\overline{X}_{B,i}$ is used with d defined by the last formula and with $r = 1$. The resulting formula is not equal to $\overline{X}_i$, of course.

Our estimator $\overline{X}_B$ is, then, another type of moving average where the deviations, $\left|X_j - \overline{X}_{B,i-1}\right|$, have been "smoothed" to minimize both the effects of outliers and of sampling from non-Gaussian populations.

This quantity $\overline{X}_B$ is readily computed on a programmable calculator. If the positive results only of $X_j - \overline{X}_B$ are considered, the equation written above becomes simply

$$\overline{X}_{B,i} = \overline{X}_{B,i-1} + \left(\frac{\sum\sqrt{X_j - \overline{X}_{B,i-1}}}{N}\right)^2$$

The programming complexity is more accurately reflected in this way of writing $\overline{X}_B$. The algorithm for each erythrocyte index requires a storage register (for $\overline{X}_B$) and a memory of 30 program steps. Two illustrative examples demonstrate the decreased contribution of outliers to $\overline{X}_B$.

Example 1. Deviations of identical magnitude; no effect.

Mean of previous batch of MCV data: $\overline{X}_{B,i-1} = 90$.

Present Data Batch	$X_j - \overline{X}_{B,i-1}$	$\sqrt{X_j - \overline{X}_{B,i-1}} =$
Sample 1. MCV = 96	96 − 90 = 6	$\sqrt{6} = 2.45$
Sample 2. MCV = 96	96 − 90 = 6	$\sqrt{6} = 2.45$
Sample 3. MCV = 96	96 − 90 = 6	$\sqrt{6} = 2.45$
Sample 4. MCV = 96	96 − 90 = 6	$\sqrt{6} = 2.45$
Sample 5. MCV = 96	96 − 90 = 6	$\sqrt{6} = 2.45$
	$\sum = 30$	$= 12.25$
	$N = 5$	$= 5$

$\overline{X} = 90 + 6 = 96$ $\quad \overline{X}_B = 90 + (2.45)^2 = 96$

Example 2. Deviations of differing magnitude; decreased effect of outliers.

Mean of previous batch of MCV data:
$\overline{X}_{B,i-1} = 90.$

Present Data Batch	$X_j - \overline{X}_{B,i-1} =$	$\sqrt{X_j - \overline{X}_{B,i-1}}$
Sample 1. MCV = 92	$92 - 90 = 2$	$\sqrt{2} = 1.4$
Sample 2. MCV = 98	$98 - 90 = 8$	$\sqrt{8} = 2.8$
Sample 3. MCV = 96	$96 - 90 = 6$	$\sqrt{6} = 2.45$
Sample 4. MCV = 95	$95 - 90 = 5$	$\sqrt{5} = 2.23$
Sample 5. MCV = 99	$99 - 90 = 9$	$\sqrt{9} = 3.0$
	$\sum = 30$	$= 11.88$
	$N = 5$	$= 5$

$\overline{X} = 90 + 6 = 96 \quad \overline{X}_{B,i} = 90$
$+ (2.37)^2 = 95.64$

(4) $\overline{X}_{MED}$, the sample median. For the hematology parameters, the sample median will be too variable. A moving average median computed by analogy with the moving average mean was studied.

(5) $\overline{X}_T$, the trimmed mean. This measure is designed specifically to minimize the contribution of outliers. The simplest trimmed mean is defined in this way. The observations in a batch, ordered from lowest to highest, are designated

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(j)} \leq \cdots \leq X_{(N-1)} \leq X_{(N)}$$

where $X_{(j)}$ is the j^{th} smallest value. Then, $X_{(1)}$ and $X_{(N)}$ are discarded and the sample mean of the remaining observations is found. A moving average trimmed mean was also studied.

(6) $\overline{X}_J$, a measure due to Johns (unpublished report, see Andrews and associates[2]). This is an estimator which minimizes the effect of outliers. As for the trimmed mean, observations are first ordered, and the smallest and largest observations are discarded. Then the "middle" observations are combined into one sum and the others into a second sum. The Johns' mean is a weighted average of these two sums. A moving average Johns' mean was studied.

Different values of r and N were used to study the sensitivity of the several measures and to determine which values of r and N were preferable. Values of N studied were 7, 10, 14, 17, 20, 23; values for r were .30, .40, .50.

All measures were applied in turn to the data from Labs A and B. The data were first analyzed unmodified, and then the data were modified to simulate both abrupt shifts and prolonged drifts. The specific models used to simulate these two ways in which a machine can lose calibration are:

Model 1. Shift Model

The first 60 patient values have no error added. The next 60 patient values have an error added; the error is P% of the reference mean μ. The process is then repeated. This model simulates abrupt loss of machine calibration with stabilization at some different level. It is obvious that a highly damped estimator might aid in the quantitation of calibration loss, but it is certain to delay its recognition; thus examination of the response of each estimator with data perturbed by Model I will give some idea of the price paid for increased damping in decreased responsiveness. Since shifts of more than 5% are, in general, readily detected by the operator, shifts of 3% were examined. Machine shifts may occur either up or down; both are adequately simulated by the model employed since removal of a positive shift is, in terms of the calibration control procedures under study, indistinguishable from a negative shift.

Model 2. Cyclical Time Trend

An error linearly increasing with sequence number is added to the first 60 patient values, attaining a maximum of P%

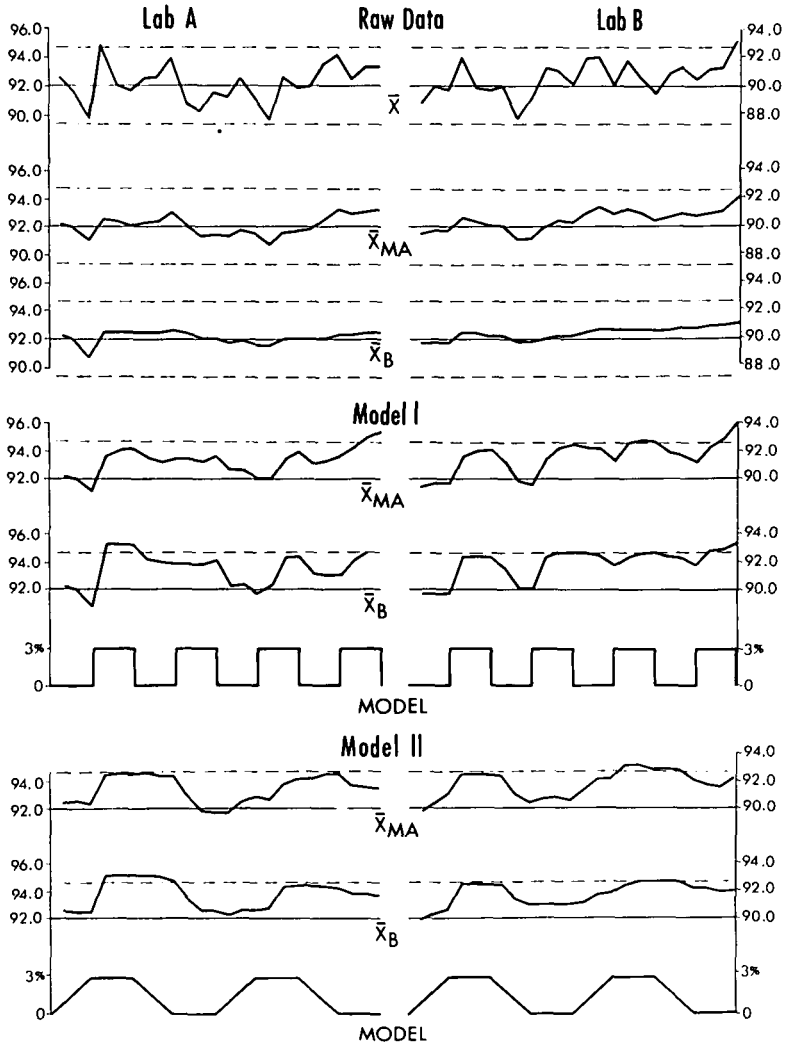

FIG. 1. Mean corpuscular volume. Four hundred eighty patients whose indices were determined in Laboratory A and a similar number of patients whose samples were analyzed in Laboratory B constitute the raw data plotted here. Twenty-four batches of 20 patients each are plotted on each line against the batch number. The top line $\bar{X}$ shows the results of plotting unmodified arithmetic means against the batch number. The second line $\bar{X}_{MA}$ demonstrates considerable smoothing as a result of incorporation of 60% of the previous mean into the following batch mean.

$\bar{X}_B$ demonstrates the additional smoothing that results from decreasing the contributions of outliers.

The ideal estimator should transform the raw data into a perfectly flat line (assuming that the blood analyzer does not drift or shift during the run). At the same time, an ideal estimator should be able to respond instantaneously to the data transformations depicted in Model 1 (shift model) and Model 2 (drift model).

A given estimator was judged on the basis of how well it was able simultaneously to meet both of →

of the reference μ. This error is held constant over the next 60 values. In the third successive set of 60 values, the error decreases linearly from P% to 0%. The fourth successive set of 60 patient values has no error added. The process then repeats itself. This model simulates the second way in which a machine might lose calibration—drift. As in the shift model, negative drift can be simulated by linear removal of a positive error, hence the cyclical nature of the model. Drifts totaling 3% were examined in this system.

Results

When the results of applying each estimator for a given r and N on each laboratory's MCV and MCHC values and the values as modified by models 1 and 2 were plotted on regular graph paper (see Figures 1 and 2), 962 graphs resulted. No overall analytic criterion to evaluate the graphs was computed—a possible subsequent paper would include such an analytic evaluation. Rather, they were evaluated visually relative to the following

the above criteria. Of the estimators studied, $\bar{X}_B$ most closely approached the ideal. It requires a programmable calculator for implementation. $\bar{X}_{MA}$, while not quite as successful, can be computed by hand.

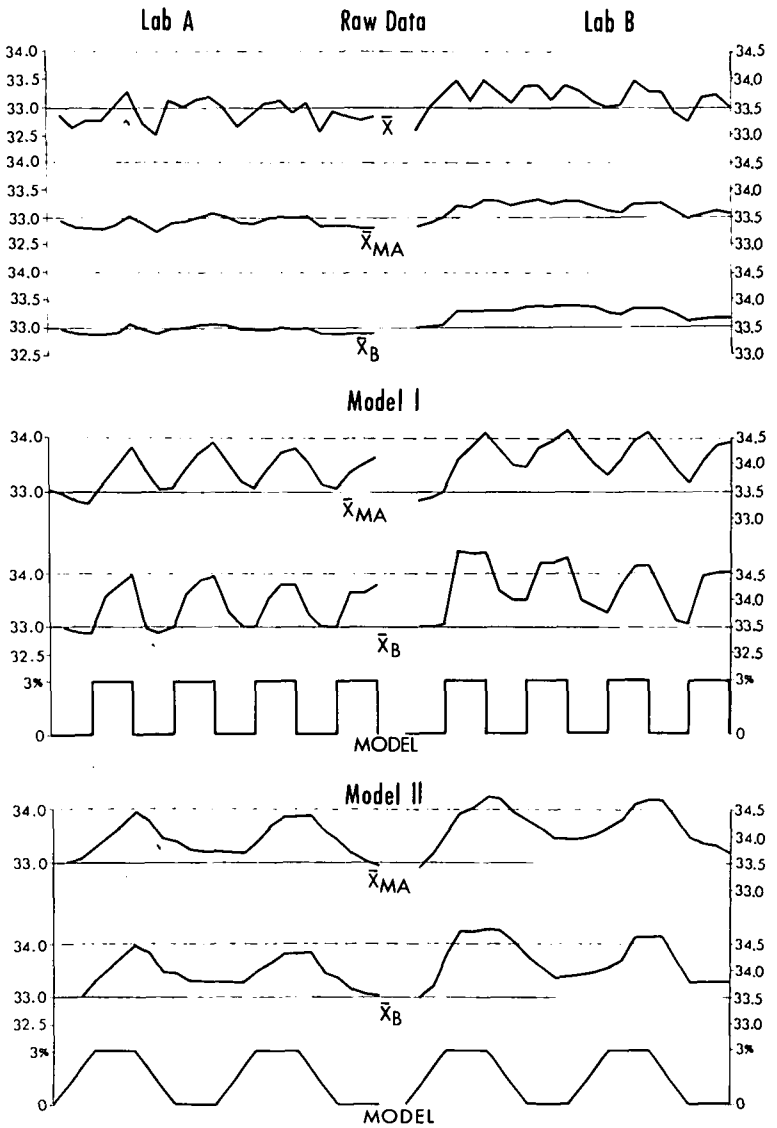

FIG. 2. Mean corpuscular hemoglobin concentration. See legend for Fig. 1. for explanation.

characteristics for an ideal result:

(1) A plot of the numerical value for the estimator against the serial batch number, based on the raw data, should lie close to the population mean value μ and the plot should be relatively smooth.

(2) The same plot, after the data have been modified by Model 1 or Model 2, should quickly demonstrate that calibration loss might have occurred. This plot should be smooth.

The more "smoothing" applied to the original data (*i.e.*, as N gets larger or r gets smaller), the less rapidly will the estimator respond to data drifts or shifts as simulated in Models 1 and 2. Alternatively, if there is too little "smoothing" of the orginal raw data, it looks progressively like $\bar{X}$; and due to the marked baseline instability, the changes introduced by the modified data (Models 1 and 2) cannot be easily identified visually. It is only as visual identification of machine shifts or drifts becomes simple and rapid that any such approach has maximum usefulness in the clinical laboratory situation.

After evaluating the several graphs for each estimator against the performance criteria, the following conclusion was reached. The estimator $\overline{X}_B$ with N = 20 and the moving average Johns' mean with r = .40 and N = 20 performed better than the other estimators; a slight preference for $\overline{X}_B$ could be supported, especially when calibration loss occurred by an abrupt change or shift rather than drift. The measure $\overline{X}_B$ is easily implemented on a programmable calculator, while the Johns' mean requires access to a computer. Both of these measures are moving averages, and both smooth data from the current batch, although the smoothing is carried out in quite different ways. Smaller values of N lead to highly variable values for each measure and produce graphs barely distinguishable from those based on $\overline{X}$.

All the non-moving average measures perform very badly (see, for example, the graph for $\overline{X}$ in Figures 1 and 2). All moving average measures show marked improvement over the corresponding non-moving average measures (see, for example, the graphs for $\overline{X}_B$ and $\overline{X}_{MA}$ in Figures 1 and 2). Thus, there is a strong indication that "moving" is more important as a variance reduction method than "smoothing" for the data types studied here. Of course, "smoothing" cannot be neglected, as a comparison of $\overline{X}_{MA}$ and $\overline{X}_B$ from Figures 1 and 2 indicates. The moving average versions of the sample mean and sample median are indistinguishable, while the moving average version of the trimmed mean is more nearly like the Johns' mean than the sample mean.

If neither a programmable calculator nor a larger computer is available, $\overline{X}_{MA}$ is the measure of choice. Hand calculation is not difficult, and $\overline{X}_{MA}$ does perform creditably.

The graphs presented in Figures 1 and 2 give the highlights of our findings. Graphs for other measures or for different values of r and N would not add significantly more information. There is, however, one observation suggested in the graphs reproduced as Figures 1 and 2 that was confirmed by the remaining graphs. The data from Lab B appear to be more "malleable," as judged by their response to the rapid changes of Model 1, than are the data from Lab A. This difference persisted even when the batches were varied so that while the batch size stayed constant, different patients were included in each batch. It is presumably due to some qualitative difference in the data that has been uncovered by the application of the estimator followed by the stress of rapid deformation induced by Model 1. This matter is under further investigation.

Discussion

While neither $\overline{X}_B$ nor $\overline{X}_{MA}$ may represent the optimal approach to the extraction of calibration information from patient data for the parameters studied, it seems clear that both procedures provide useful information. The information provided is the more useful because it is available at short intervals throughout the analytic run and thus gives very rapid indication of a calibration loss. Furthermore, both procedures stabilize the mean indices with sufficient precision so that it is possible not only to ascertain that calibration loss has occurred but also to quantitate the degree and the direction of the correction required. To illustrate this point, consider the application of estimator $\overline{X}_B$ to the raw data from Lab B. Note in Figure 1 that the Coulter S in Lab B held within ±.5% of the expected mean MCV of 90 for the first 240 patient samples (approximately two days' work in this laboratory). Starting at that point, a slow steady drift occurred, amounting at maximum to just over 1%. In the same group of patients, the MCHC, after starting at the expected mean value of 33.5, rose 1%, stabilized there for about 400 patient samples and then drifted downward again for the last 40–60 samples

analyzed. The MCV could have been recalibrated on the basis of this evidence by the 240th sample, and the MCHC could have been justifiably lowered at the 100th. However, it is in general impossible to move the potentiometers successfully in such small steps as 1%, and ordinarily recalibration is not attempted until the shift or drift approaches the 3% level. The clinical significance of a calibration which amounts to less than 3% is likewise questionable. Still, it is reassuring to be able to detect these tiny changes in machine calibration. The corrective procedure when the drift reaches the 3% mark is to recalibrate by adjusting the appropriate potentiometer 3%. The corrective action required in the case of a rapid shift is not so obvious since there are two possible causes. First, the machine may indeed have lost calibration, but it is also possible that an unusual group of patients (*i.e.*, from a cancer chemotherapy ward where macrocytosis is common) has produced the result. A rapid decision must be made as to which has occurred, for if the machine has drifted it must be recalibrated immediately. It is our policy under these circumstances to rerun five patient samples which were analyzed previously at a time when the estimator indicated that all parameters were "in control." Such samples form, in effect, a highly accurate secondary standard which is being continuously produced as long as the quality control charts show normal machine function. If the machine is no longer correctly calibrated, it will be evident from a comparison of the duplicate determinations on these five patient samples.

Of the three estimators $\overline{X}_{JMA}$, $\overline{X}_B$ and $\overline{X}_{MA}$, only $\overline{X}_B$ has had extensive trial in hospital laboratories. The combined experience with this estimator is now in excess of 15 laboratory years, with the longest experience in any single laboratory of three years. Furthermore, this can be done without the need to employ commercial whole blood standards. Since this will result in a considerable saving, the costs of a programmable calculator or the expense of programming a laboratory computer to accomplish this task will be offset in six months or less in most laboratories. If implemented on a laboratory computer, it requires no additional technologist time once programming and debugging are complete. If the calculation of $\overline{X}_B$ is performed on a programmable calculator, entering each batch of 20 patient samples and plotting the results requires less than 3 minutes.

Finally, it is likely that this general approach has applicability as a general quality control procedure to other clinical laboratory determinations. While this question was not examined in depth, it seems likely that the approach may have merit wherever the proportion of normal results is high in ratio to abnormals and the distribution is symmetrical. A further requirement seems to be that the population coefficient of variation for the parameter being studied be less than 10% if a calibration control at the 2–3% level is desired.

References

1. Amador E, Hsi BP, Massod MF: An evaluation of the "average of normals" and related methods of quality control. Am J Clin Pathol 50:369–378, 1968
2. Andrews DS, Bickel P, Hampel S, et al: Robust estimates of location; survey and advances. Princeton, N.J., Princeton University Press, 1972, 373 pp
3. Begtrup H, Leroy S, Thyregod P, et al: "Average of normals" used as control of accuracy, and a comparison with other controls. Scand J Clin Lab Invest 27:247–253, 1971
4. Britton GM, Brecher G, Johnson CA, et al: Stability of blood in commonly used anticoagulants: Use of refrigerated blood for quality control of the Coulter Counter Model S. Am J Clin Pathol 52:690–694, 1969
5. Britton GM, Brecher G, Johnson CA: Evaluation of the Coulter Counter Model S. Am J Clin Pathol 52:679–689, 1969
6. Dixon N, Northam BE: Quality control using the daily mean. Clin Chim Acta 30:453–461, 1970