

**STATISTICAL QUALITY CONTROL OF
HIV-1 ELISA TEST PERFORMANCE**

R.C. Gentleman, M.S. Hamada
D.E. Matthews and A.R. Wilson

IIQP Research Report
RR-92-05

May 1992

STATISTICAL QUALITY CONTROL OF HIV-1 ELISA TEST PERFORMANCE

R.C. Gentleman, M.S. Hamada, D.E. Matthews
Department of Statistics and Actuarial Science
University of Waterloo

and

A.R. Wilson
Woodbridge Foam Corporation

May 1992

ABSTRACT

Ensuring the quality and performance of HIV-1 ELISA testing in routine laboratory situations has previously been identified as an important, unresolved problem. An appropriate solution requires that the statistical issues which underlie the testing situation be properly understood. The use of Shewhart control charts would seem to be a natural answer to the problem. They are shown to be unsatisfactory, however, because a key aspect of the test is nested within each microplate of blood samples. Further investigation reveals that the logarithm of the measured optical densities is a more appropriate scale in which to discriminate between positive and negative control samples, as well as reactive and nonreactive specimens. A new statistical control chart, called the separation chart, is defined, and an algorithm for the preparation of separation (T) and range (R) charts is described. Retrospective analysis of nearly 1300 microplates using T and R charts conclusively demonstrates that these statistical process control methods represent an important advance in the continuing effort to maintain and improve the quality and performance of routine HIV-1 ELISA testing.

Key Words: Control chart, HIV-1 ELISA, Log transformation, Statistical quality control, Unpaired t test.

1 Introduction

Millions of enzyme-linked immunosorbent assays (ELISAs) are performed each year to detect antibodies to the human immunodeficiency virus, type 1 (HIV-1). These tests are carried out by agencies involved in collecting blood for transfusion purposes, or by diagnostic laboratories such as those operated by various levels of government in the interest of public health. The Centers for Disease Control have recently suggested that U.S. hospitals should test all new patients for evidence of infection with HIV-1 (see Joyce and Brown, 1991). If this suggestion becomes a routine component of hospital admission procedures, the number of HIV-1 ELISAs which are carried out annually will increase substantially.

Schwartz, Dans and Kinosian (1988) have distinguished between test efficacy (performance under ideal conditions) and test effectiveness (performance under average conditions), and have highlighted the important practical concern which test effectiveness represents. The HIV-1 ELISA is a complicated assay, involving a variety of reagents and laboratory techniques. Since there are many factors which can affect the performance of the test, it seems prudent to ensure that the testing procedure is performing as it should. The tools of *statistical quality control* (SQC) are commonly used for precisely this purpose in industrial applications. While the manufacturer of a serologic test ensures that its testing procedure can detect seropositive individuals with adequate precision under idealized conditions, the SQC procedures which we describe are designed to monitor the performance of the testing procedure under day-to-day conditions. For example, SQC procedures can detect problems which may be caused by the physical apparatus, laboratory procedures or laboratory personnel. This report focuses on the quality control procedures recommended by the manufacturer. Laboratories typically use additional in-house quality control procedures, but these are not considered here.

This report is divided into seven sections. We begin with a brief description of the ELISA test, and identify the need for ongoing quality control to ensure the satisfactory performance of these kits in day-to-day screening or diagnosis. The fundamental statistical issues in HIV testing are discussed in section 3 and the implications of these considerations with respect to the monitoring of test performance are summarized. Section 4 describes standard Shewhart control charts – the most obvious statistical tool for ongoing quality control – and outlines problems which arise when these charts are used in a routine way for monitoring ELISA test performance. In section 5 we propose a new type of control chart, the separation (T) chart, and suggest that the T chart, combined with the range (R) chart, directly addresses the key issues in the quality control of ELISA test performance. Concrete evidence of the value of these new statistical tools is presented in section 6, where we use T and R charts to retrospectively analyze data collected by the Virology Laboratory of the Ontario Ministry of Health (OMH). The data provided by the OMH contained no confidential information

and consisted of the measured optical densities (ODs) of negative and positive control specimens. The results of this retrospective study provide corroborative evidence that regular, systematic use of T and R charts could provide early warning signals that test performance has deteriorated. In the final section of the report, we summarize the insights which these investigations into statistical tools for quality control of ELISA test performance have provided.

2 Description of the HIV-1 ELISA

An individual who has been infected with HIV-1 but whose immune system has not been overwhelmed by the virus will have a high concentration of antibodies to the virus in his or her blood. It is these individuals that one hopes to identify using an ELISA. Routine screening of all blood donations for the presence of antibodies to HIV-1 was initiated in the United States as soon as the first HIV-1 ELISA was licensed by the Food and Drug Administration on March 2, 1985 (see Schorr *et al* 1985). Similar testing in Canada was implemented, nationally, by the Canadian Red Cross Society in November, 1985. It is important to realize that the ELISA is a screening assay and not a diagnostic test; we shall elaborate on the differences in the sequel.

In order to discuss methods of ensuring satisfactory assay performance on a routine basis, we need to understand the basic steps involved in obtaining an ELISA result. To produce an ELISA kit, a quantity of the HIV-1 virus is first manufactured under laboratory conditions and then deactivated by a chemical and ultraviolet light treatment. The deactivated virus is bonded to polystyrene microplates which consist of test wells arranged in an 8x12 matrix. Thus, any desired combination of at most ninety-six samples and/or standards may be assayed at one time. Samples or standards are added to the wells and diluted to an appropriate volume. After a suitable reaction time, any antibodies in the samples bind with the antigen. The wells are washed to remove any unbound antibody and then a conjugate solution (anti-human antibody attached to an enzyme) is added. The conjugate combines with the antibody/antigen complex if the sample contains any HIV-1-specific antibody. The microplate is washed a second time to remove any unbound conjugate and then a substrate is added. The substrate reacts with the enzyme to cause a specific color change. The reaction is stopped after a specified period of time and the OD of each well is read on a spectrophotometer which is designed specifically for evaluating microplates. The degree of color (OD) which is read by the spectrophotometer is proportional to the concentration of HIV-1 antibodies in the sample. A high OD indicates the presumptive presence of HIV-1 antibody, whereas a low absorbance reading for a test well identifies an HIV-1 antibody-negative specimen.

The ELISA is a relatively complex assay, involving a sequence of four chemical reactions and fifteen steps, each of which plays a critical role in the outcome of the assay. There are many

potential problem areas which could give rise to unacceptable variation in the assay result. Some of these problem areas are apparent from the test procedures, whereas other potential problems might only be apparent to experts in the field. A preliminary list of important factors may be found in Table 1. The same information is also presented in the form of an Ishikawa Diagram in Figure 1. The Ishikawa Diagram (see Ishikawa, 1982) is a valuable quality improvement tool which allows the factors contributing to some result, which is usually known as a quality characteristic, to be displayed in a concise, organized fashion. The objective of the HIV-1 ELISA is to produce an accurate evaluation of the specimens assayed; the factors listed in Table 1 contribute to the attainment of this objective.

To guard against such potential problems, ELISA kit manufacturers generally recommend running various control wells, such as two blanks, two negative controls and three or four positive controls as well as eighty-eight or eighty-nine specimens on a plate. For the plate results to be acceptable, specific criteria, referred to as kit rules, must be met. Table 2 summarizes the type and number of replicated standards used in seven of the ELISAs for HIV-1 licensed by the Food and Drug Administration in the United States.

3 Fundamental Statistical Issues

We begin with a general, informal introduction to the statistical issues which are associated with screening and diagnostic tests. Subsequent discussion will focus on the practical issues which are specific to the use of ELISA for HIV-1 screening.

The purpose of HIV-1 antibody testing (and of virtually every other similar laboratory test) is to determine the true status of the sample being tested. This process may be thought of as classification; specimens (and, therefore, individuals) are identified as either HIV-1 seropositive or seronegative depending on the test result. There are two types of errors that can be made in such a classification situation. Either seronegative samples may be classified as seropositive (Type I error), or seropositive samples may be identified as seronegative (Type II error). The impact, or cost, of each type of error depends on the purpose for which the test is being performed. In general, the probabilities of the two types of error cannot simultaneously be made arbitrarily small. In most cases, reducing the probability that one type of error will occur results in an increase in the probability of the other type of error. In blood screening situations, Type II errors are clearly more serious, and an appropriate test implementation would seek to eliminate this type of error, or to reduce its probability of occurrence to an absolute minimum. A Type I error in a screening context would result in the destruction of a unit of blood which could have been used for transfusion. Note that all positive screen results are confirmed by supplementary testing so that the probability of a

final Type I error is reduced. However, negative test results are usually accepted so that Type II errors are not detected. While Type I errors do have an impact on blood inventories and the costs of blood banking, these consequences are minor when compared to the social and medical impact of transfusing HIV-1-infected blood. On the other hand, if the purpose of the test is the diagnosis of infected individuals, a Type I error becomes relatively more important.

Associated with the probabilities of the two types of error are the concepts of the predictive value of a positive test outcome, and the predictive value of a negative test result. These relate the operating characteristics of the test to the prevalence of HIV-1 infection in the population. For example, the predictive value of a positive test result is the probability that an individual who is classified as infected on the basis of the test result is really infected with HIV-1. These predictive values are mainly of interest in the diagnostic setting. For a screening test, it is the potential infectivity of the unit of blood associated with the sample being tested that is important, not the status of the donor.

The identification of blood units as HIV-1 seronegative or seropositive using an ELISA outcome is accomplished by classifying the OD (on some appropriate scale of measurement) of a test well at the end of a fairly complicated laboratory procedure. As we shall shortly argue, the natural logarithm of the ODs appears to be the appropriate scale in which to classify specimens and monitor the performance of ELISA testing for HIV-1. A natural statistical assumption when attempting to classify individuals into one of two groups is that there are two populations, in our case HIV-1 seropositive and seronegative blood donors, and that the measurement of interest has some parametric distribution in each of the populations. Given estimates of the distributional parameters involved, and assuming that no measurement error occurs in determining the OD of a specimen, the probability that either error (Type I or II) will occur when a sample is classified on the basis of the test result can be computed. To the best of our knowledge, no work has been done to obtain accurate estimates of the distributional parameters for either the seronegative or the seropositive blood donor populations. Because the characteristics of the seropositive blood donor population might be expected to change during the course of the AIDS epidemic, identifying these distributions is a particularly difficult problem.

A further problem that arises is that the OD for a sample is measured with error. This measurement error contributes to the probability of misclassification, and can substantially alter the operating characteristics of the test. This variation is induced by the experimental procedure, and it is this variation that our SQC procedures will assess. If this induced variability becomes unacceptably large so that the probability of a Type II error increases unduly, then steps must be taken to identify and then eliminate the causes of the excessive variation. Conceptually, we can further divide this induced variation into two components, a between-plate component and a within-plate

component. Thus, we hypothesize the following model for the sample measurement (MOD) of the logarithmic OD (LOD), namely

$$MOD = LOD + \gamma_p + \eta_{(p)} ,$$

where γ_p is a random effect that is the same for all units on a given plate and $\eta_{(p)}$ is a random effect for a particular unit within a plate. Further, we assume that γ_p has mean zero and variance σ_B^2 , that $\eta_{(p)}$ has mean zero and variance σ_W^2 , and that they are independent. Some experimentation designed to quantify the magnitudes of these sources of variation and the validity of the assumptions would be very useful. Well-designed experiments would also indicate which of the factors in Table 1 seem to have the largest effects on MOD . Controlling these factors could result in an improved measurement process which exhibits substantially less variation than is being observed in the processes which are used to screen blood for antibodies to HIV-1 at present.

The classification procedures recommended by the manufacturers of all the screening kits with which we are acquainted are nested within a plate. That is, all of the quantities needed to classify individuals are determined from the same plate on which the individual observations are measured. Thus, for classification purposes, under the assumption that the within-plate variation is independent of the level or mean value, the between-plate variation is unimportant. However, prudent QC practice dictates that between-plate variation should be monitored since excessive variation here can identify problems before they adversely affect the classification algorithm.

The more important within-plate variation can be measured from the negative and positive controls that are run on each plate. These quantities are in essence replicated samples from a single source; hence, the variation they exhibit should be a reasonable estimate of σ_W^2 . While the manufacturers of ELISA kits advocate a number of checks within a plate which are, in some sense, designed to monitor the magnitude of σ_W^2 , we shall indicate areas in which these procedures do not detect substantial changes in the within-plate variation.

Next, we discuss reasons for preferring the logarithmic measurement scale. Figure 2a displays the ODs of the positive and negative controls, adjusted by their microplate means, for Du Pont HIV-1 ELISA microplates assayed between January 18 and 26, 1988 at the Virology Laboratory of the Ontario Ministry of Health. The variance in the negative controls is much smaller than that of the positive controls. This result should not be surprising; the phenomenon of non-constant variance for different absorbance levels is well known in the literature of analytical chemistry. Skoog and West (1982) suggest that the absorbance standard error varies proportionally with absorbance, that is, with the mean level. This suggests using a variance-stabilizing transformation which in this case would be a logarithmic transformation. Figure 2b presents the log-transformed version of the data displayed in Figure 2a. On the log scale, the positive and negative controls appear

to exhibit approximately the same amount of variation. Using the 240 positive and 120 negative samples that were available for the period from January 18 to 26, histograms and normal quantile-quantile (Q-Q) plots (see Chambers *et al*, 1983) of the adjusted observed *LODs* of the positive and negative control samples, respectively, were constructed (Figure 3). The plots suggest that the negative controls are slightly heavy-tailed and the positive controls are slightly skewed to the right with respect to a normal distribution. For practical purposes, however, these deviations are not substantial. Q-Q plots (not shown) were also computed for the example discussed in Section 6 which cover an extended period of time and show even stronger evidence in support of the normality assumption. This provides a justification for the assumption that the observed *LOD* for a positive control sample has a $N(\mu_P, \sigma^2)$ distribution while for a negative control sample it has a $N(\mu_N, \sigma^2)$ distribution.

Let us consider the kit rules for the Du Pont HIV-1 ELISA which are detailed in Du Pont (1988). Although this testing kit is no longer available, it was licensed by the Food and Drug Administration of the United States, and was the HIV-1 ELISA which provided the data used herein. The kit rules recommend running two blank wells, two negative controls and four positive controls on a plate; the plate is acceptable if it satisfies the following criteria.

1. At least one of the blanks must have an absorbance less than 0.2.
2. After subtracting the lower blank, at least one negative control must have an absorbance between 0.0 and 0.25.
3. After subtracting the lower blank, at least three of the positive controls must have absorbances between 0.5 and 2.0 and must be within thirty percent of the average absorbance of the usable positive controls.
4. The ratio of the average absorbance of the usable negative controls to that of the usable positive controls must be less than 0.2.

If the plate is acceptable, the average absorbance of the usable positive controls is multiplied by 0.5. The result is called the “Reaction Threshold Value”, which we denote by RTV. Other manufacturers refer to this threshold as the “Cut Off Value” or COV. Specimens with absorbances which exceed the RTV are declared reactive; otherwise, a sample is labeled non-reactive.

In general terms, the rules outlined in the Du Pont HIV-1 ELISA product information insert seem to be concerned with the location (i.e., the mean value) and the variability of the positive and negative controls and with the separation of the means for the two types of controls. Rule 2 concentrates attention on the location of the negative controls. It is not concerned with variability; in fact, it even disregards evidence of such within-plate variation by allowing one “outlier” to be

ignored. Rule 3 appears to address both the location and variation in the absorbances of the positive controls. It also allows the user to ignore an outlier, i.e., evidence of within-plate variation. Notice that the magnitude of the variation allowed depends on the mean. Rule 4 uses a ratio to define the necessary separation between the mean values for the positive and negative control samples. Apart from the minimum separation required, it does not appear to take account of variation in the estimates of the means for the two types of controls. The evaluation procedures of many of the current HIV-1 ELISA kits presented in Table 2 also allow some outliers to be discarded. Because of these shortcomings, it is natural to consider control charts, graphical SQC methods, which do take into account within-plate variability.

4 Shewhart Control Charts

Control charts (Grant and Leavenworth, 1988) are statistical tools which are widely used in industrial situations to monitor production and service processes. For example, a test instrument can be monitored by frequently running a test standard and charting the results. In the case of an HIV-1 ELISA, each plate incorporates two standards, i.e., the positive and negative controls. In this section, we consider the applicability of control charts for monitoring the HIV-1 ELISA screening process.

Regardless of the physical context in which a control chart is used, the particular type of chart employed depends on the characteristics of the data which are being collected. The simplest type of chart for continuous measurements such as ELISA absorbances is the \bar{X} and R chart, also known as Shewhart control charts (Grant and Leavenworth, 1988). Data for these charts are usually collected in samples (“subgroups”) of size four or five. For the HIV-1 ELISA, we take the subgroup to be the positive or negative controls on a given microplate. The \bar{X} chart is used to plot the within-subgroup means, while the R chart plots the within-subgroup ranges, i.e., the difference between the largest and the smallest values in the subgroup. Each chart has a center line, as well as lower and upper control limits. The center line identifies the mean value of the measurement which the chart monitors. The lower and upper control limits identify natural limits of variation for the measurement when the process is in control, i.e., in the absence of unusual sources of variation between subgroups. Typically, the probability that a plotted point falls within the control limits is 0.997 for a process which is in control. Consequently, a point which is plotted outside the control limits suggests that the process was out of control when the corresponding subgroup was obtained. Out-of-control points on the R chart usually exceed the upper control limit and suggest that variation in the measurement being monitored has increased. Out-of-control points on the \bar{X} chart suggest that the mean of the measurement has shifted; the direction of the shift is indicated

by the particular control limit, i.e., lower or upper, which has been exceeded.

Using data recorded for Du Pont HIV-1 ELISA microplates processed during early 1988 at the Virology Laboratory of the Ontario Ministry of Health, several \bar{X} and R charts for each type of control were plotted. Figure 4 shows separate \bar{X} and R charts for the absorbances of the negative and positive controls prepared from data collected between January 18 and 26, 1988; the same testing kit was used throughout the eight-day period represented on the chart. Figure 5 presents the corresponding charts based on data collected between January 26 and February 23, 1988. It is important to note that the testing kit was changed during this period; the time of the change is identified by a vertical dotted line on the charts. The center lines and lower and upper control limits in Figure 5 are based on the historical information provided by the data collected between January 18 and 26. (R charts based on subgroup sizes of six or less have a lower control limit of zero.) As the following discussion indicates, even though these charts reveal important aspects of the assay as plates are being processed, the suitability of these charts as a tool for monitoring test performance is questionable.

The R charts in Figure 4 for both types of control samples suggest that the within-plate variation was stable for the plates processed during the period concerned. However, the \bar{X} charts show numerous out-of-control points, which suggests that the means of both controls change significantly from microplate to microplate. This instability in the means from plate to plate might have been expected, since the Du Pont quality control procedures attempt to adjust for plate-to-plate differences by specifying acceptable ranges for the measured absorbances of the controls. This example suggests that \bar{X} charts based on the natural subgroups defined by the plate-specific positive and negative controls would not help in monitoring the accuracy of the HIV-1 ELISA. The frequency with which out-of-control points occur makes the \bar{X} chart virtually useless as a tool for monitoring the ELISA screening process. Moreover, provided the plate-specific changes in the means of the positive and negative controls are not too large, the ability of the assay to accurately discriminate between reactive and non-reactive samples may not be compromised. The \bar{X} chart fails to satisfactorily address this important question.

Figure 5 displays the corresponding \bar{X} and R charts during a period when the testing kit changed. While the R charts show few out-of-control points, suggesting that the variability of the positive and negative controls is relatively stable during the period shown, there is a noticeable decrease in the \bar{X} charts which coincides with the kit change. Thus, in order to use the \bar{X} chart to monitor HIV-1 ELISA performance, a new center line and control limits would need to be calculated each time a new testing kit was used. Since \bar{X} charts do not provide satisfactory information concerning the important question of accurate discrimination between reactive and non-reactive samples, this additional complication convinces us that the \bar{X} chart should not be used in this

particular application. On the other hand, these examples do demonstrate that R charts should be a useful tool for ensuring assay performance, because they monitor the within-microplate variation. If this variation increases substantially, the ability of the assay to accurately discriminate between reactive and non-reactive specimens may be compromised. As we have previously indicated, this is one of the key requirements which a suitable SQC procedure for the HIV-1 ELISA must satisfactorily address.

5 The T Chart

We now introduce a new control chart, the separation (T) chart, and argue that T and R charts together represent an effective combination for monitoring the continuing performance of an ELISA for HIV-1. The T chart is based on the assumption that the MOD for a positive control sample has a $N(\mu_P, \sigma^2)$ distribution while for a negative control sample it has a $N(\mu_N, \sigma^2)$ distribution (see Section 3). Then, if μ_P and μ_N are sufficiently well-separated relative to σ^2 , the ELISA should be capable of accurately differentiating between reactive and non-reactive samples. As we have previously indicated, although the manufacturer's evaluation procedures generally use the idea of maintaining an adequate separation between the negative and positive controls on a microplate, they fail to account fully for the variability observed in the control samples on a plate.

In general, most practitioners would resort to the familiar t test to determine whether the data from the two samples could be regarded as evidence contradicting the null hypothesis that the means of the corresponding populations were equal. The T chart which we propose simply displays a modified Student's t statistic derived from the two types of control samples. The T chart should be used in combination with separate R charts for the positive control samples and the negative control samples from each plate. Collectively, these charts should enable a user of the HIV-1 ELISA to monitor the stability of the screening test performance from plate to plate. Ideally, if the separate R charts are in control and the values on the T chart do not fall below a suitably chosen lower control limit, then the T and R charts may be regarded as indicating that the corresponding microplate specimens have been correctly differentiated into a reactive group, clustered around the mean of the positive controls, and a non-reactive group clustered around the mean of the negative controls.

If \bar{P} and \bar{N} denote the microplate average values of the log absorbances of the positive and negative control samples, respectively, and $\hat{\sigma}$ denotes the pooled estimate of standard deviation of the control samples' log absorbances, then the modified Student's t statistic which is plotted on

the T chart is

$$\frac{\bar{P} - \bar{N} - c}{\hat{\sigma}\sqrt{n_P^{-1} + n_N^{-1}}},$$

where n_N and n_P denote the number of negative and positive controls, respectively, used on the microplate. The extra term in the numerator of the t statistic, i.e., c , represents the smallest acceptable difference between the means of the logged positive and logged negative control populations. According to Rule 4 of the Du Pont test evaluation procedures (see Section 3), a suitable value for c would be $\log 5$. Thus, the T chart should detect a separation between the means of the logged positive and logged negative controls of $\log 5$ or smaller. The choice of an appropriate value for c can be suitably modified for ELISA kits from other manufacturers or for users with more or less stringent requirements.

In an appendix to this report we outline the steps involved in preparing and using T and R charts to monitor the HIV-1 ELISA. A T chart was plotted using the data from January 18 through 26, 1988 (see Figure 6a). The dashed line is 4.604, which is the 0.995 percentile for a t distribution with four degrees of freedom. This ensures, approximately, that a difference between the means of the logged positive and negative controls of $\log 5$ or smaller is detected 99.5% of the time. Note that the T chart requires only a lower control limit because only microplates for which the capability of the ELISA test to accurately discriminate is questionable need to be identified.

The observed t statistics were quite large during this period, ranging from 10 to 80. In view of the wide range of values observed, we revised the vertical scale of the T chart; Figure 6b displays the same t statistics on the log scale. Thus, the T chart can be used to monitor assay performance by investigating plates when small t statistics occur. A small value of the t statistic would constitute evidence of a serious deterioration in the ability of the assay to discriminate between non-reactive and reactive samples. Note that for the period from January 18 to 26, 1988 (see Figure 6b), the T chart provides no grounds for questioning the ability of any plate processed to correctly classify reactive and non-reactive samples. On the other hand, the T charts displayed in Figures 7 and 8 (see Section 6), do suggest that the reliability of several of the plates processed during the period of time represented on those charts is open to question.

The T chart should be used in tandem with separate R charts for the positive and negative control samples. That is, the R charts need to be checked first to ensure that the within-microplate variation is stable and is similar for both types of control samples on a plate. Out-of-control points on either R chart, i.e., points above the corresponding upper control limit, indicate that the within-plate variation for that particular type of control sample has increased substantially. Out-of-control points on the T chart, i.e., points below the lower control limit, indicate that the separation between the positive and negative controls is not adequate to ensure that the test is capable of accurately

discriminating between reactive and non-reactive specimens on the microplate.

Before presenting an application, we consider the sampling distribution of the t statistic and its ramifications. Strictly speaking, although the positive and negative controls are normally distributed, the statistic which we are plotting on the T chart does not have a Student's t distribution with $n_P + n_N - 2$ degrees of freedom if the variances are unequal. Such a situation is likely to occur if one or both of the R charts are out of control. Therefore, it is important to ascertain how often a difference of log 5 or smaller in the means of the logarithms of the positive and negative control ODs would then be detected. By matching moments, we can show that the sampling distribution of the statistic is approximately that of a constant k times a t distribution with f degrees of freedom, where k and f depend on the ratio σ_P^2/σ_N^2 . We have carried out some probability calculations which use this approximation for $n_P = 4$ and $n_N = 2$; for values of the ratio σ_P^2/σ_N^2 between 0.1 and 10.0, a difference, on the logarithmic scale, in the control means of log 5 or smaller would be detected at least 92.8% of the time. (A Monte Carlo study based on 100,000 simulations gave a slightly lower number of 91.3%.) Consequently, even for unequal positive and negative control variances, the T chart is still able to identify most occasions when the HIV-1 ELISA test cannot correctly discriminate between reactive and non-reactive specimens. On the other hand, based on the observed separation between logged positive and negative controls exhibited in the January 18 to 26 data relative to the observed variation, fewer than one in 10000 plates would be expected to generate a value on the T chart below the lower control limit.

6 Example

To demonstrate the benefits which could be derived from regular use of T and R charts to monitor the performance of routine ELISA testing for antibodies to HIV-1, we next obtained all the ODs for the negative and positive control samples used on microplates processed by the Virology Laboratory between January 18, 1988 and November 16, 1988. In order to facilitate the visual presentation of the corresponding T and R charts, we divided this period into three intervals of approximately equal length: January 18 – April 13, April 14 – July 20 and July 21 – November 16. During the first of these intervals, 385 consecutive microplates containing 1540 positive control samples and 770 negative control samples were processed. Since these T and R charts represented a continuation of the charts covering the initial period from January 18 to January 26, the values for the centre lines and lower and upper control limits for the positive and negative controls which were derived from the initial 60 microplates (January 18 to January 26) were used to prepare the series of three T and R charts covering all 1286 consecutive microplates processed by the laboratory during this ten-month period. All occasions when the lot number on the testing kit changed were identified

from the laboratory records and marked on the appropriate charts. All microplates for which the value plotted on the T chart was below the lower control limit (LCL), or which were rejected as invalid according to the assay evaluation requirements were identified. The outcome of the assay, and the corresponding evaluation based on the T and R charts, were compared; Table 3 provides a detailed summary of the results of this comparison.

Figure 7 presents the two R charts and corresponding T chart for the 385 consecutive microplates processed between January 18 and April 13, 1988. Although there are occasional range values which exceed the UCL on each of the R charts shown in Figure 7, these control charts suggest that during much of the period represented, the variation observed in the negative and positive controls on the microplates which were processed was relatively stable. Likewise, the T chart for the same period indicates that, with perhaps seven exceptions, the microplates which the Virology Laboratory processed between January 18 and April 13 were capable of accurately discriminating between the reactive and non-reactive specimens which were tested.

The T and R charts which we prepared corresponding to the 388 consecutive microplates processed between April 14 and July 20 were very similar to those shown in Figure 7, and therefore have been omitted. A total of ten out-of-control points were identified on the T chart during this period. All of these out-of-control points coincide with out-of-control values on one or both of the R charts. Additional details concerning the corresponding microplates are provided in the comparison described at the conclusion of this section. On the basis of these charts, we concluded that virtually all of the ELISA tests which were carried out in the Virology Laboratory between April 14 and July 20, 1988 were capable of accurately discriminating between the reactive and non-reactive specimens on each microplate.

Figure 8 shows the T and R charts for the 513 consecutive microplates processed between July 21 and November 16, 1988. In contrast to the charts in Figure 7 which summarize the favourable experience of the laboratory with respect to the performance of the Du Pont HIV-1 ELISA test during the first half of 1988, the appearance of the T and R charts in the second half of the year deteriorated noticeably. In particular, the R charts for the logarithmic ODs of the negative and positive control samples evaluated on the microplates were frequently out-of-control, i.e., statistically unstable. During the period from September 15 to November 16, and coinciding with the use of Lots 130T, 144E and 151E of the Du Pont HIV-1 ELISA test, 26% (79/307) of the microplates showed evidence of excessive variation in the range value calculated from the four positive controls on each plate. At approximately the same time, as the R chart for the negative controls in Figure 8 indicates, a further 17% (53/307) of these same microplates exhibited excessive variation in the range value of the corresponding negative control samples, and 26 (8%) microplates exhibited excessive variation in both types of controls. As we have previously indicated, increased

variation in the control samples on a microplate can compromise the capability of the plate to accurately discriminate between reactive and non-reactive specimens. In fact, as the T chart shows, a number of the microplates processed towards the end of this period (particularly those belonging to Lots 144E and 151E, which were used to test specimens between October 6 and November 16) may have been unable to accurately discriminate between reactive and non-reactive specimens. It is important to note that the Virology Laboratory had previously instituted additional quality control procedures, including duplicate testing of every specimen, which enabled lab personnel to identify the problems highlighted by the T and R charts; consequently, the reliability of the test results returned by the laboratory to physicians in the province of Ontario was ensured.

During the period from January 18 to July 21, only one microplate processed in the Virology Laboratory was rejected as invalid by the assay evaluation rules (henceforth referred to as the kit rules). In fact, the positive controls on that particular microplate were too variable, violating one of the kit rules. The T and R charts also identified this particular microplate as questionable; the points plotted on both R charts as well as on the T chart were outside the control limits. However, as Table 3a indicates, the T chart for this period contains an additional 16 out-of-control points corresponding to microplates which the kit rules did not reject as invalid. All of these microplates exhibited excessive variation in either the positive or the negative controls, variation which the kit rules ignore when outliers are discarded. In fact, a careful examination of the separate R chart for the positive controls during this six-month period reveals that excessive variability in the positive controls was observed on 43 (6%) microplates. An additional 103 (13%) out-of-control points were identified on the R chart for the negative controls. Only seven microplates generated out-of-control points on both R charts.

The results of a similar comparison for the 513 microplates processed between July 21 and November 16 are summarized in Table 3b. A total of 29 microplates were identified on the basis of the kit rules or because the corresponding points on the T chart were below the LCL. Five of these microplates were rejected by both the kit rules and the T chart. However, an additional 14 plates were identified as questionable by the T chart, but valid according to the kit rules. As Table 3b indicates, all 14 microplates exhibited excessive variation in one or both of the corresponding R charts, i.e., at least one of the points plotted on the separate R charts is out of control. During this same period, a total of 90 (18%) points plotted on the R chart for the positive controls were above the UCL; the corresponding total for the negative controls was 124 (24%).

Finally, some properties of the controls identified and discarded as outliers by the Du Pont rules are revealing. In the following, only microplates which were not rejected by the kit rules but which had a single outlier will be discussed. Between January 14 and July 20, 1988, 12 microplates contained a single positive control which was identified as an outlier because it was either out of

range or too variable. Each of these discarded positive controls was the smallest positive control OD on its respective microplate. When we compared these discarded positive controls with their microplate RTVs, we discovered that four of them were below their corresponding microplate RTVs. Between July 21 and November 16, 1988, a control was identified as an outlier on 47 different microplates. (Two negative controls were out of range; 45 positive controls were either out of range or too variable.) Again, the discarded positive controls were the smallest positive control ODs on their respective microplates. Moreover, 17 of these 45 outliers were below their corresponding RTVs. The two negative control outliers were slightly out of range, but fell well below their respective microplate RTVs. The fact that known positive controls which were discarded as outliers would have been classified as seronegative if they had been specimens is particularly disturbing. Note that the appropriate R chart was out of control for each of the 59 microplates mentioned, thereby highlighting the importance of using the R charts.

7 Conclusions

In the preceding sections of this report, we have described the HIV-1 ELISA test and have highlighted the important role which appropriate statistical methods can play in monitoring the ongoing performance of such tests in day-to-day operations. Although Shewhart control charts, i.e., \bar{X} and R charts, would seem to be ideally suited to such an application, a careful examination of the statistical nature of HIV-1 ELISA testing revealed that \bar{X} charts are virtually useless, because the key characteristics of ELISA testing, namely the positive and negative control samples, are nested within the individual microplates. R charts, however, should be useful in monitoring the stability of the within-microplate variation of each type of control sample.

Further analysis of the data from 1286 microplates revealed that a logarithmic transformation of the ODs results in separate distributions for the negative and positive control samples which can be reasonably approximated by the normal distribution. This transformation reduced the dependence between the mean level of the ODs and the variation, thus indicating that $\log(OD)$ is an appropriate scale on which to classify reactive and non-reactive specimens. This examination also facilitated the development of a new statistical control chart, the T chart, to monitor the capability of an individual microplate to correctly classify samples. A point which is below the lower control limit on the T chart indicates that the separation, on the logarithmic scale, between the two types of controls is inadequate to ensure HIV-1 ELISA performance. In such a situation, laboratory staff and management should take immediate action to ascertain the reason or reasons for the apparent deterioration in ELISA testing performance. Other considerations persuade us to suggest that the number of negative controls should be increased so that the number of control samples of each

type on a microplate would be the same. This change would increase the precision with which the variability of the negative control samples can be estimated, and enhance the performance of both the R chart for the negative controls, and the corresponding T chart.

The retrospective use of T and R charts on nearly 1300 microplates processed at the Virology Laboratory of the Ontario Ministry of Health in 1988 was employed to validate this new method of statistical process control. The results of that study demonstrated that T and R charts are clearly capable of identifying problematic microplates which the manufacturer's recommended operating procedures also identify. The same charts also revealed additional microplates whose ability to correctly discriminate may have been seriously compromised, microplates which the manufacturer's recommended operating procedures failed to identify.

In conclusion, we believe that these methods have the potential to become an important tool in the essential, continuing effort to maintain and improve the quality and performance of routine HIV-1 ELISA testing and therefore ought to be adopted in laboratories where such testing is done for diagnostic and screening purposes. Their adoption could be widespread if HIV-1 ELISA kit manufacturers would develop and recommend product-specific SQC procedures which users of their kits could then implement.

8 Acknowledgements

We are grateful to the Virology Laboratory, Laboratory Services Branch of the Ontario Ministry of Health for providing the data used in this research, and for permission to publish the results. In particular, we want to acknowledge the contribution of Carol Major, whose comments and discussion helped to improve an earlier version of this work, and also that of Robert James, whose interest and enthusiasm for this research we have greatly appreciated. Finally, we wish to thank Duncan Murdoch; his wizardry in recovering the data has earned our continuing admiration.

This research was supported by the Institute for Improvement in Quality and Productivity and the Natural Sciences and Engineering Research Council of Canada.

9 References

Barré-Sinoussi, F., Chermann J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Brun, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868–71.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for Experimenters*. New York, NY:

- John Wiley and Sons (232–238).
- Chambers, J.M., Cleveland, W.S., Kleiner, B. and Tukey, P.A. (1983). *Graphical Methods for Data Analysis*. Belmont, California: Wadsworth International Group (191–238).
- Du Pont (1988). Du Pont HIV ELISA Test Instructions. PN 770557.002.
- Grant, E.L. and Leavenworth, R.S. (1988). *Statistical Quality Control, 5th Edition*. New York, NY: McGraw-Hill Book Company (33–169).
- Ishikawa, K. (1982). *Guide to Quality Control*. Tokyo, Japan: Asian Productivity Organization (18–26).
- Joyce, C. and Brown, P. (1991). Blanket HIV tests urged for US hospitals. *New Scientist*, **131**, 1788, 5.
- Popovic, M., Sarngadharan, M.G., Read, E. and Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, **224**, 497–500.
- Schwartz, J.S., Dans, P.E. and Kinosian, B.P. (1988). Human immunodeficiency virus test evaluation, performance, and use. Proposals to make good tests better. *Journal of the American Medical Association*, **259**, 2574–2579.
- Skoog, D.A. and West, D.M. (1982). *Fundamentals of Analytical Chemistry, 4th Edition*. Toronto, Canada: Saunders College Publishing (539–546).

Appendix

Constructing T and R Charts

In the algorithm which follows, we assume that the natural logarithms of the ODs of the positive and negative controls are being used in the various calculations described. To implement these charts, data from at least 20 plates are required in order to derive the various control limits.

(i) The R Chart

For each of the two types of controls on a microplate

1. Compute a range (maximum minus minimum) value for each microplate and prepare a run chart of the range values, i.e., a plot of range versus the order in which the plates were processed.
2. Compute the mean of these range values (\bar{R}); plot \bar{R} as a solid line on the run chart.
3. Calculate the UCL for the R chart by multiplying \bar{R} and the appropriate value of the constant D_4 (see Grant and Leavenworth, 1988, p. 670, Table D). The choice of D_4 is determined by the number of positive (n_P) or negative (n_N) control samples on each microplate. Plot the UCL as a dashed line on the run chart.

Repeat steps 2 and 3 after removing any range values that exceed the current UCL, until there are no additional plotted points which exceed the current UCL. The final value of the UCL provides a baseline against which to monitor the within-plate variation of the corresponding type of control samples in the ELISA testing process.

As each subsequent microplate is processed, compute the appropriate range values and plot them on their respective control charts. If either plotted value exceeds its particular UCL, the variability of the corresponding control samples on the microplate has increased noticeably, and the cause of this increased variation which has been detected should be identified and eliminated.

(ii) The T Chart

1. Compute separate sample means and sample variances for the positive and negative controls on each microplate.
2. Compute the t statistic specified in formula (1) (see p. 12), using a suitable value for c , the smallest acceptable separation, on the logarithmic scale, between the means of the negative

and positive controls. Plot these t values on a run chart, using a logarithmic scale to accommodate the wide range of values observed. (Values of the t statistic which are less than one can be plotted as a zero.)

3. Determine the 0.995 quantile for a t distribution with $n_P + n_N - 2$ degrees of freedom (Box, Hunter and Hunter, 1978, p. 631, Table B1). Plot this value as an LCL (dashed line) on the run chart.

As each subsequent microplate is processed, compute its t statistic and add it to the T chart. If the microplate is capable of accurately discriminating between reactive and non-reactive specimens, the corresponding point on the T chart should lie above the LCL. A point below the LCL indicates that the separation between the mean of the logarithmic ODs of the positive controls and the corresponding mean of the negative controls is not adequate, when compared to the estimated standard error of the separation. The cause of this serious deterioration in assay performance should be sought, and eliminated as soon as possible.

Table 1: Factors Contributing to the Accuracy of an ELISA Assay Result

Equipment	Temperature control of laboratory and incubators. Calibration and maintenance of spectrophotometer.
Plates	Consistency of the manufactured antigen. Amount of antigen bonded to each well and consistency of coverage. Polystyrene characteristics.
Reagents	Purity of reagents (initially and lack of contamination). Concentration of reagents. Reagent shelf life.
Specimens	Blood or plasma. Age of specimens and storage conditions. Temperature of specimens at time of testing. Blood factors.
Procedures	Measurement of specimens and reagents. Plate washing procedures. Reaction and incubation times.

Table 2. Summary of seven HIV-1 ELISA test evaluation rules

Test	Number of		Microplate Evaluation Rules				RTV	Other Rules
	Neg	Pos	Wavelength (nm)	Negative Controls, N_i	Positive Controls, P_i			
Abbott Laboratories HIVAB HIV-1 EIA	2	3 ^b	492	$0.01 \leq N_i \leq 0.1$ $0.5\bar{N} < N_i < 1.5\bar{N}$	$0.4 \leq P_i \leq 1.999$ $0.5\bar{P} < P_i < 1.5\bar{P}$	$\bar{N}^a + 0.1\bar{P}$	$\bar{P} - \bar{N} \geq 0.4$	
Cambridge Biotech Recombigen (<i>env</i> & <i>gag</i>) HIV-1 EIA	2	3 ^b	450	$0 < N_i$ $0 < \bar{N} \leq 0.15$	$0.6 \leq P_i \leq 1.7$	$0.4\bar{P}$		
Cellular Products Retro-Tek HIV-1 ELISA	2	4 ^b	405	$0 < \bar{N} < 0.2$	$0.35 \leq P_i \leq 1.3$ $0.7\bar{P} < P_i < 1.3\bar{P}$	$0.5\bar{P}$		

Table 2. Continued

Test	Number of Controls		Wavelength (nm)	Microplate Evaluation Rules			
	Neg	Pos		Negative Controls, N_i	Positive Controls, P_i	RTV	Other Rules
Genetic Systems LAV EIA	3 ^b	2	450	$0 < N_i \leq 0.14$	$\bar{P} \geq 0.7$ $0.65\bar{P} < P_i < 1.35\bar{P}$	$\bar{N} + 0.225$	
Organon Teknika Voronostika HIV-1,2 ^d Microelisa System	3 ^b	1	405 ^c	$0.1 \leq N_i \leq 0.4$ $0.5\bar{N} < N_i < 1.5\bar{N}$	$P_i \geq 0.7$	$\bar{N} + 0.27$	$P_i - \bar{N} \geq 0.5$
United Medical UBI-Olympus HIV-1 EIA	2	4 ^e	492	$N_i \leq 0.1$ $0.5\bar{N} < N_i < 1.5\bar{N}$	$0.4 < P_i^+ < 2.0$ $0.5\bar{P}^+ < P_i^+ < 1.5\bar{P}^+$	$\bar{N} + 0.1\bar{P}^+$	$\bar{P}^+ - \bar{N} \geq 0.4$ $P_i^- > RTV$

^a An over-bar indicates the average of the usable controls

^b May discard at most one control outside the indicated range

^c Corresponding evaluation rules apply if the microplate is read at 690 nm rather than 405 nm

^d Organon Teknika Corporation also manufactures a Voronostika HIV-1 Microelisa System; the numbers of controls and microplate evaluation rules are identical to those shown in the table for the HIV-1,2 system

^e Two strongly reactive controls, P_i^+ , and two weakly reactive controls, P_i^-

Table 3. Comparison of the problematic microplates identified either by the Du Pont HIV-1 ELISA test evaluation rules or by the corresponding T chart. **a.** January 18–July 20, 1988 **b.** July 21–November 16, 1988.

	T chart in control		T chart out-of-control				
	Number of R charts in control		Number of R charts in control		Total		
Du Pont rules plate status	0	1	2	0	1	2	
a.							
Valid	-	-	-	2	14	0	16
Invalid	-	-	-	1	0	0	1
b.							
Valid	-	-	-	5	9	0	14
Invalid	3	5	2	1	2	2	15

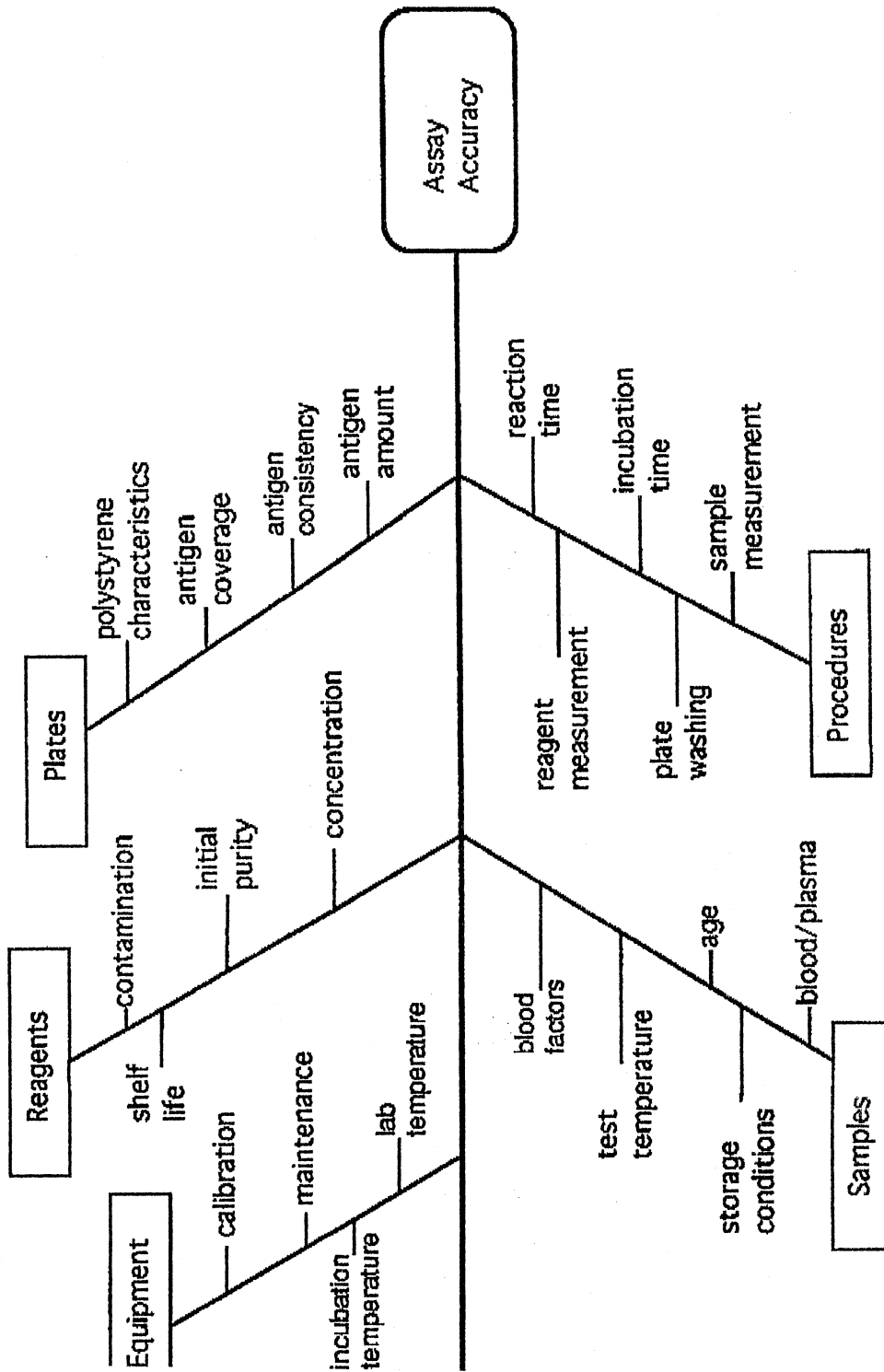
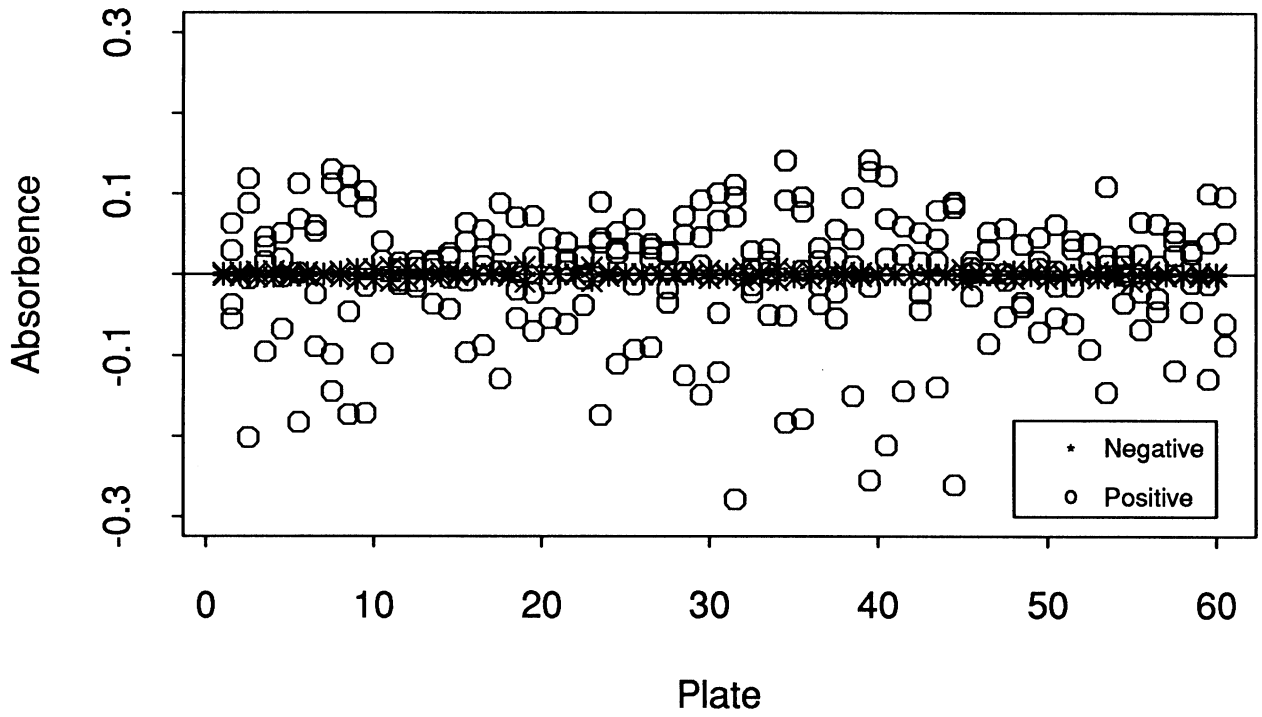


Figure 1: Ishikawa Diagram of ELISA Assay

Figure 2: Centered Controls and Log Controls
1-18-88 to 1-26-88

(a) Centered Controls



(b) Centered Log Controls

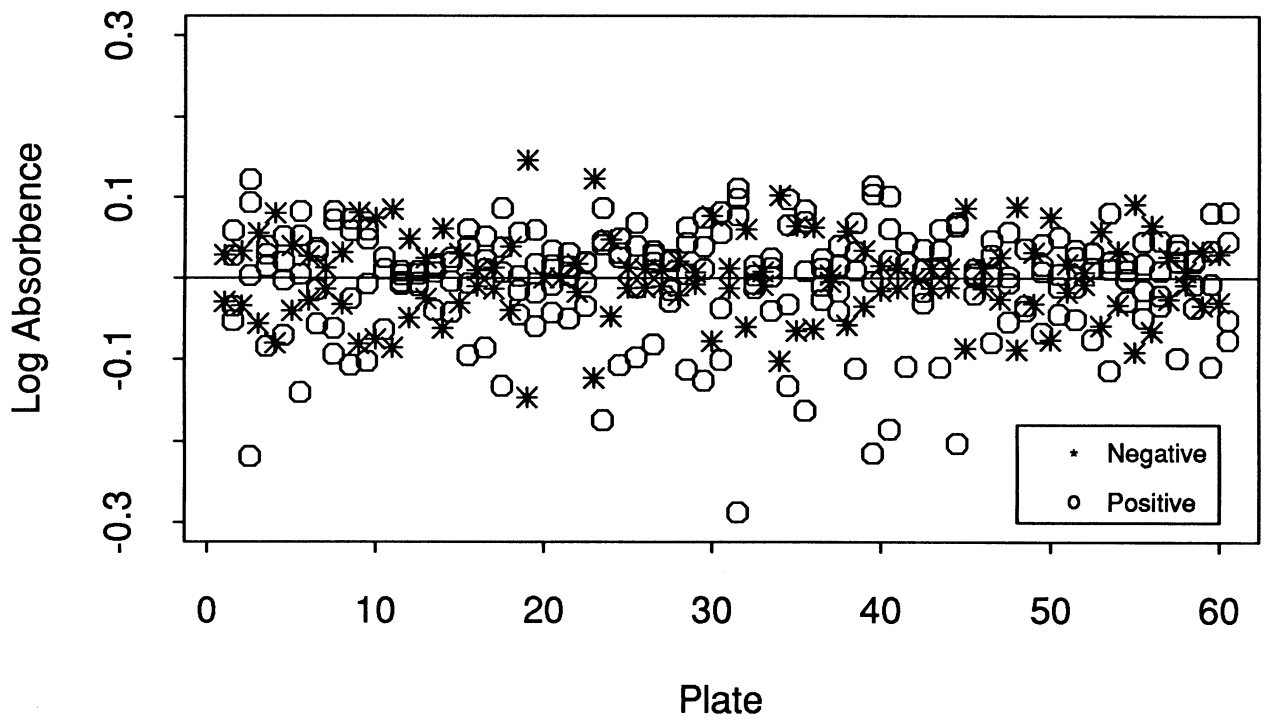
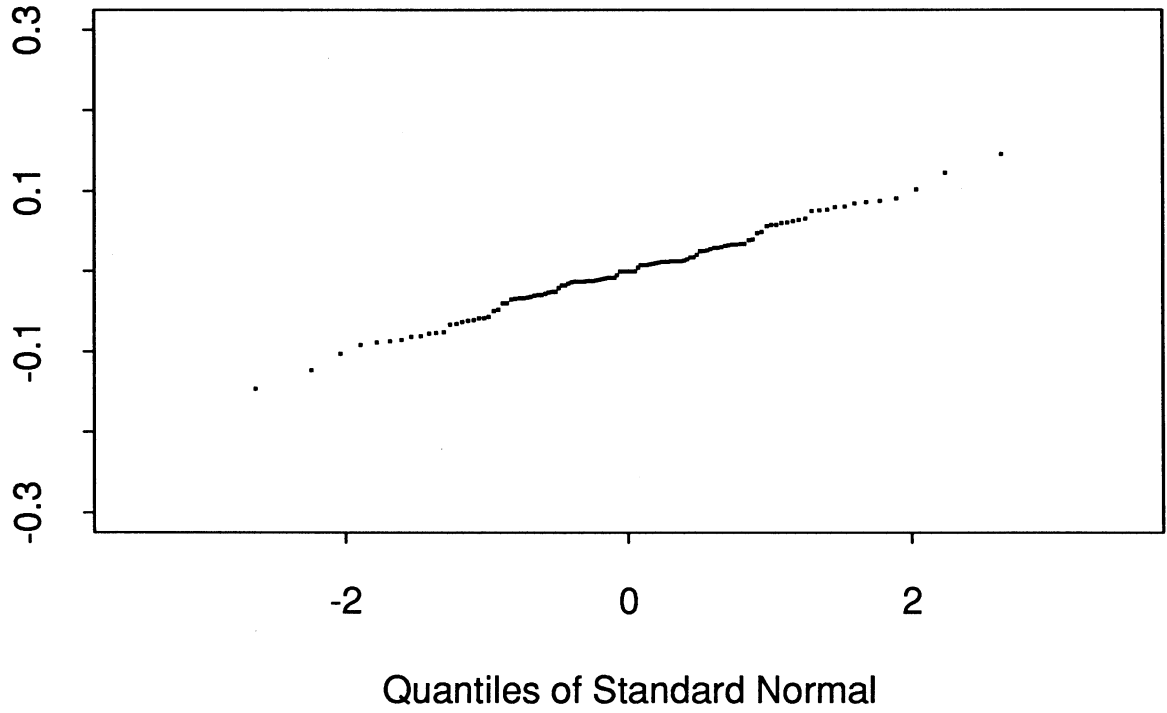


Figure 3: Q-Q Plots of Centered Log Controls
1-18-88 to 1-26-88

Q-Q Plot of Centered Log Negative Controls



Q-Q Plot of Centered Log Positive Controls

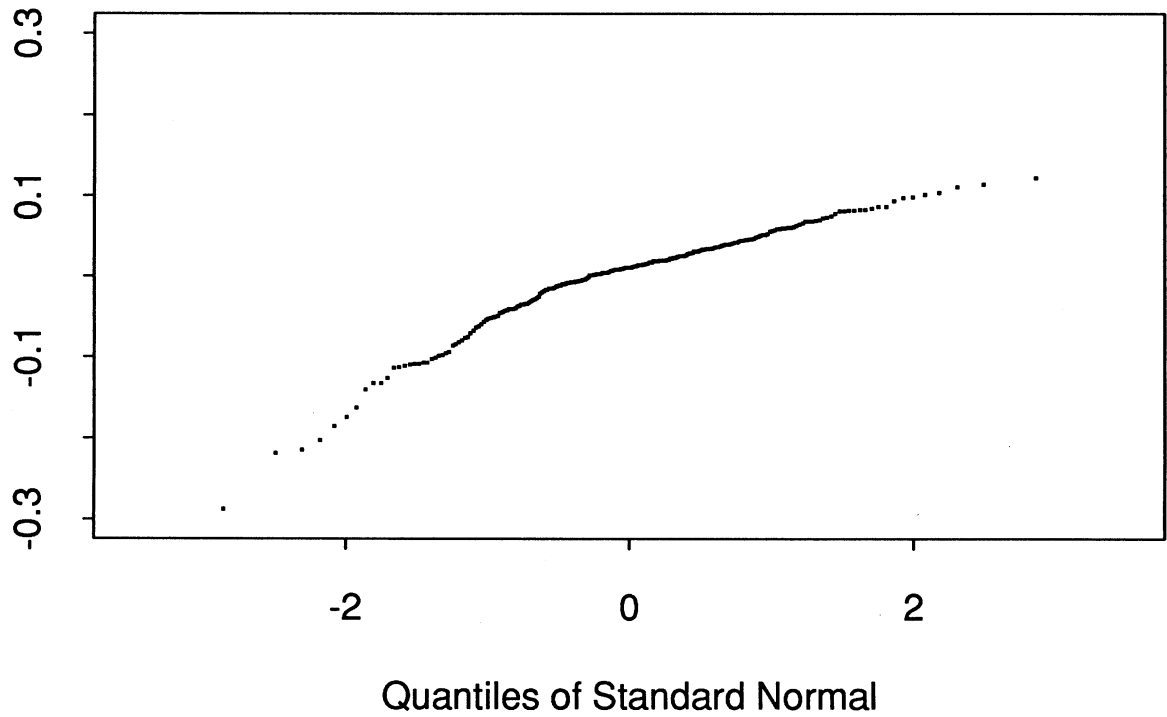
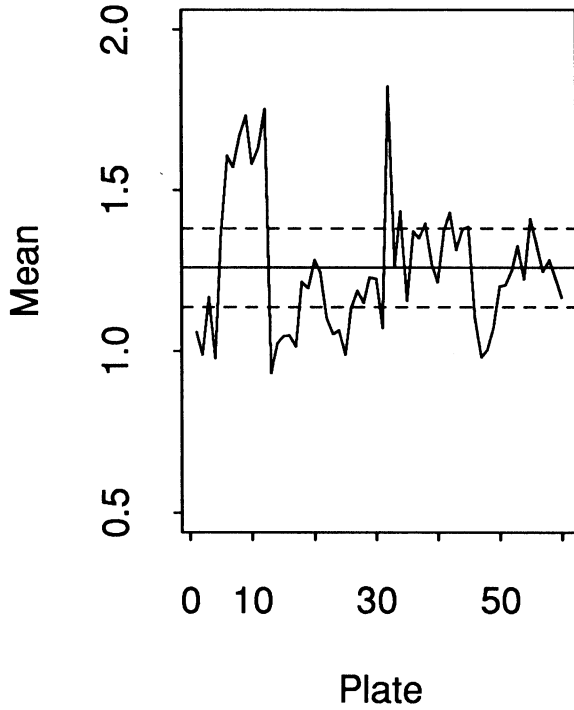
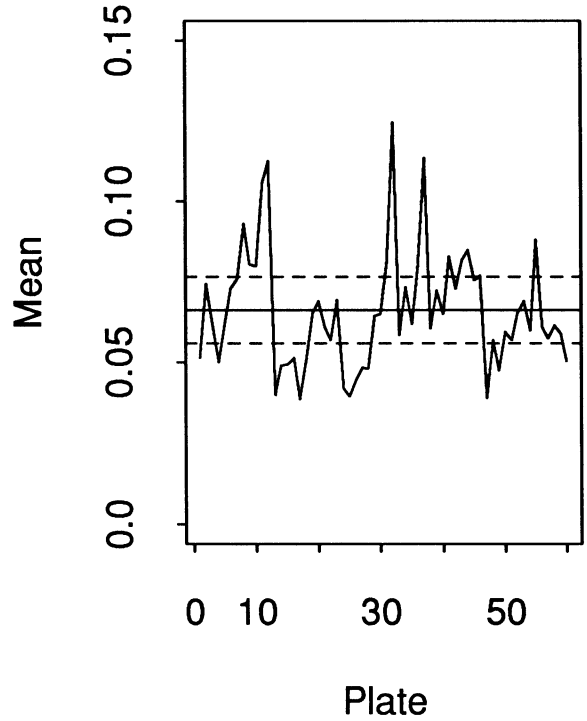


Figure 4: XBar-R Charts for Positive and Negative Controls
1-18-88 to 1-26-88

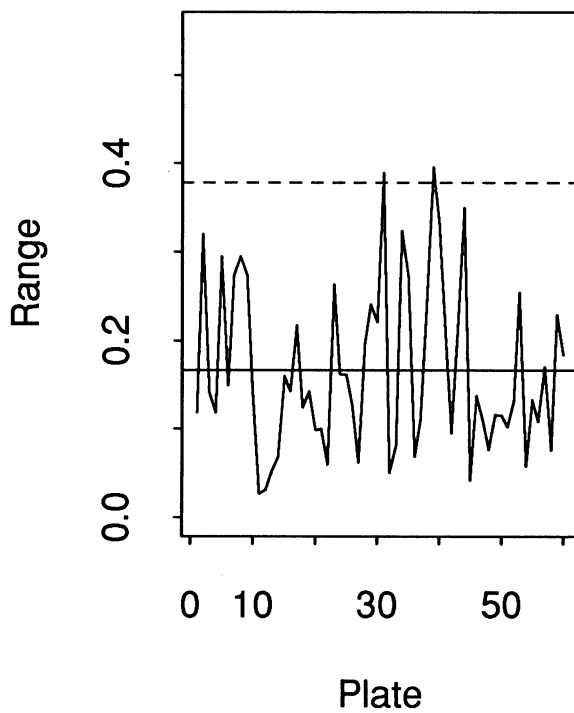
Xbar Chart for Positive Controls



Xbar Chart for Negative Controls



R Chart for Positive Controls



R Chart for Negative Controls

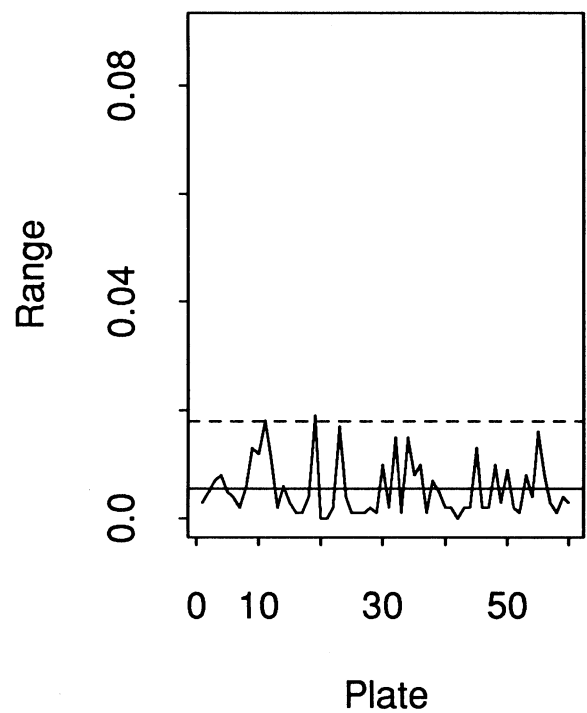
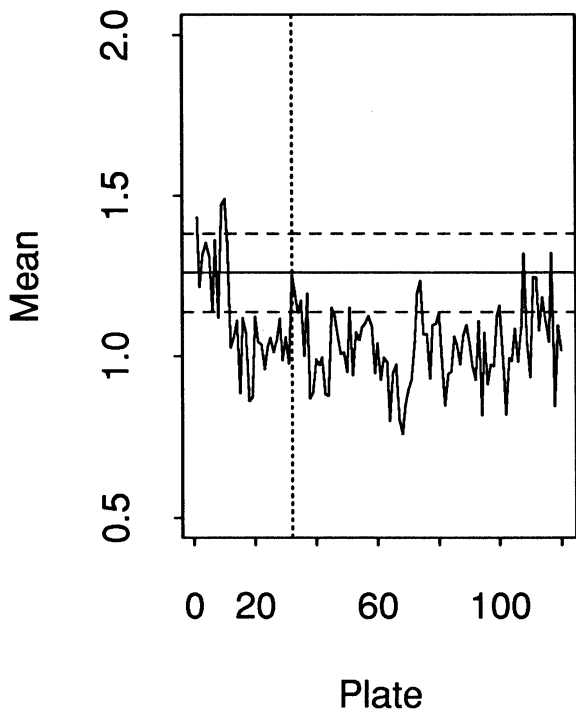
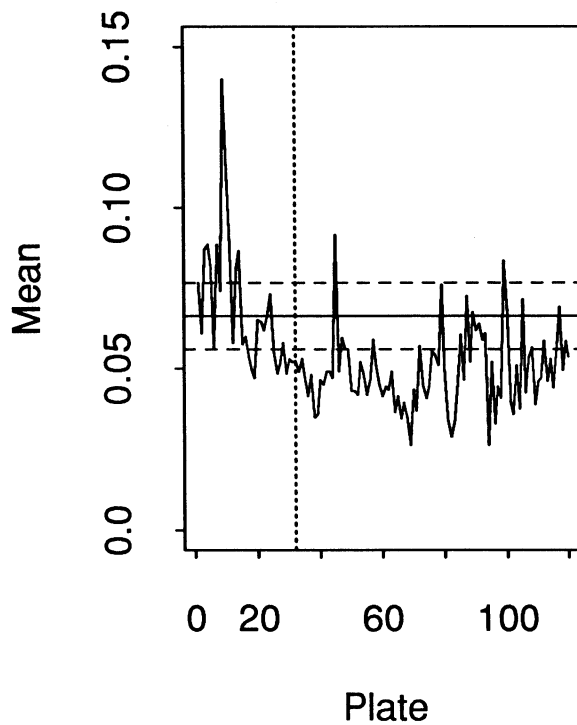


Figure 5: XBar-R Charts for Positive and Negative Controls
1-26-88 to 2-23-88

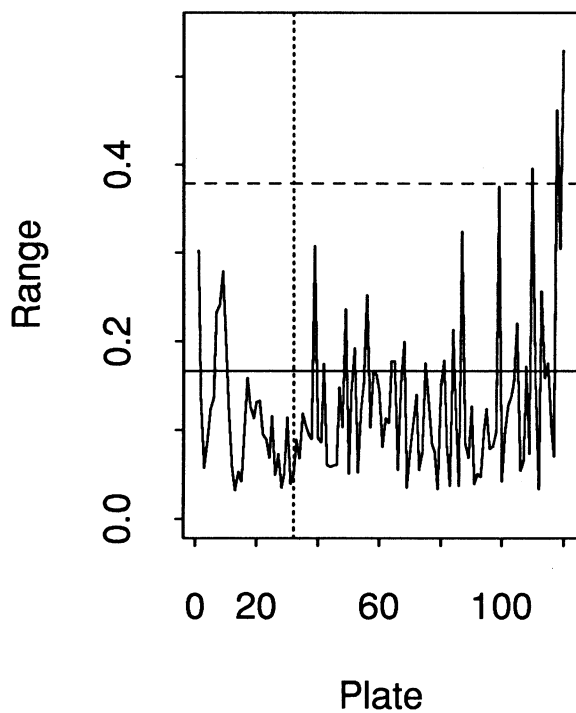
Xbar Chart for Positive Controls



Xbar Chart for Negative Controls



R Chart for Positive Controls



R Chart for Negative Controls

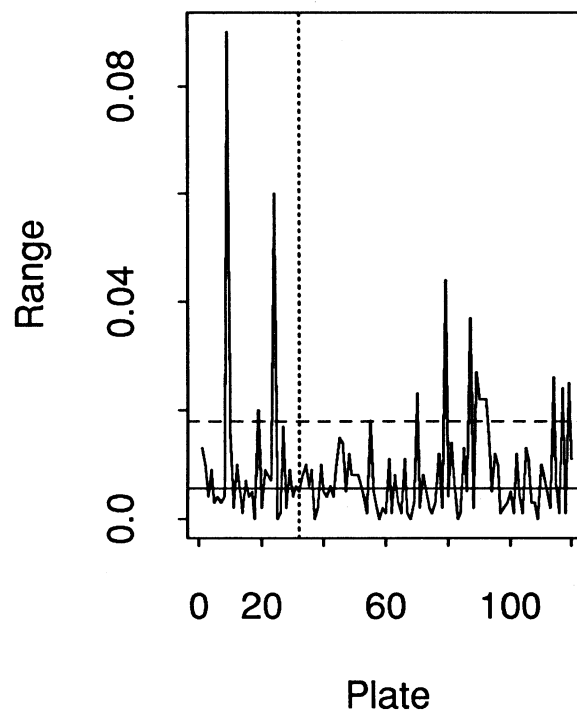
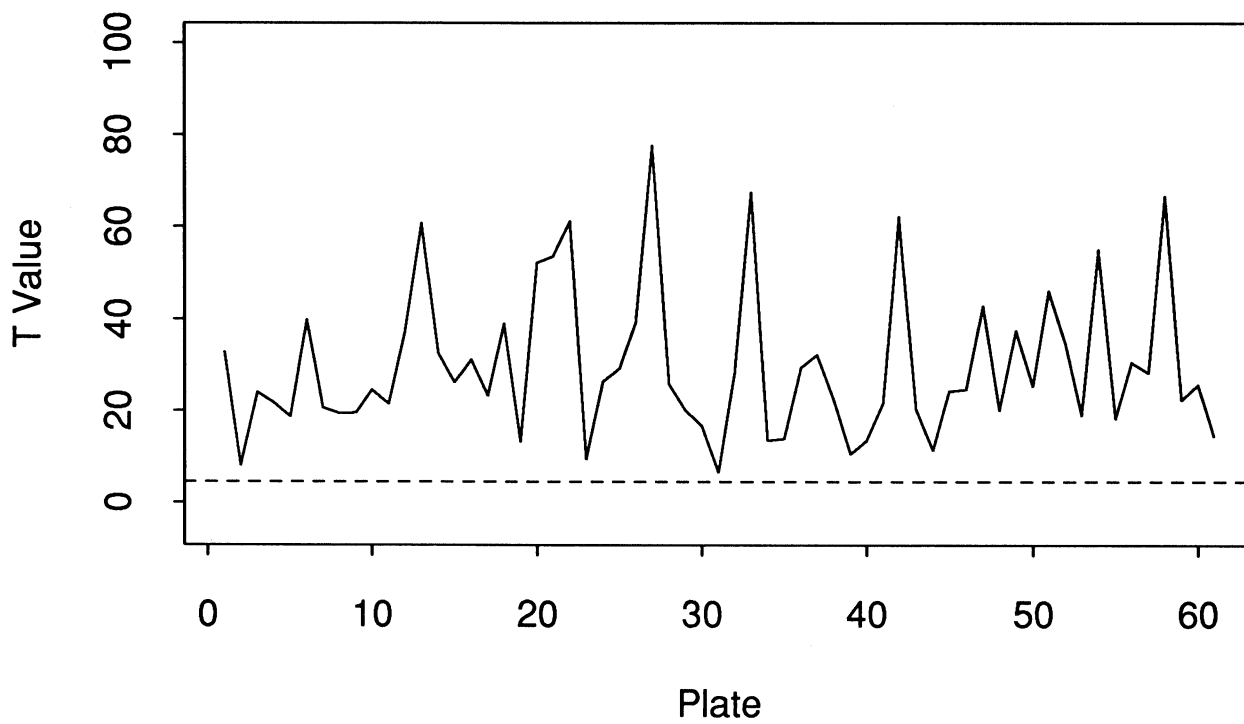


Figure 6: T Chart
1-18-88 to 1-26-88

(a) T Chart of Logged Controls



(b) Log Scaled T Chart of Logged Controls

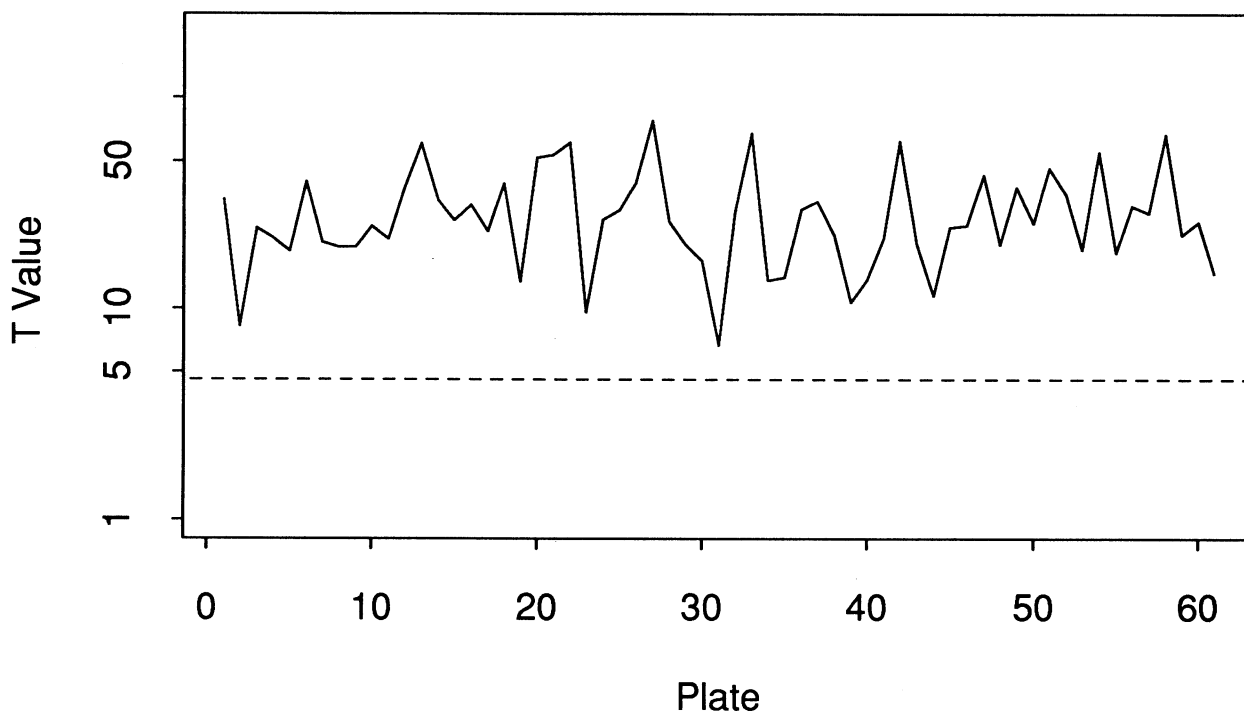
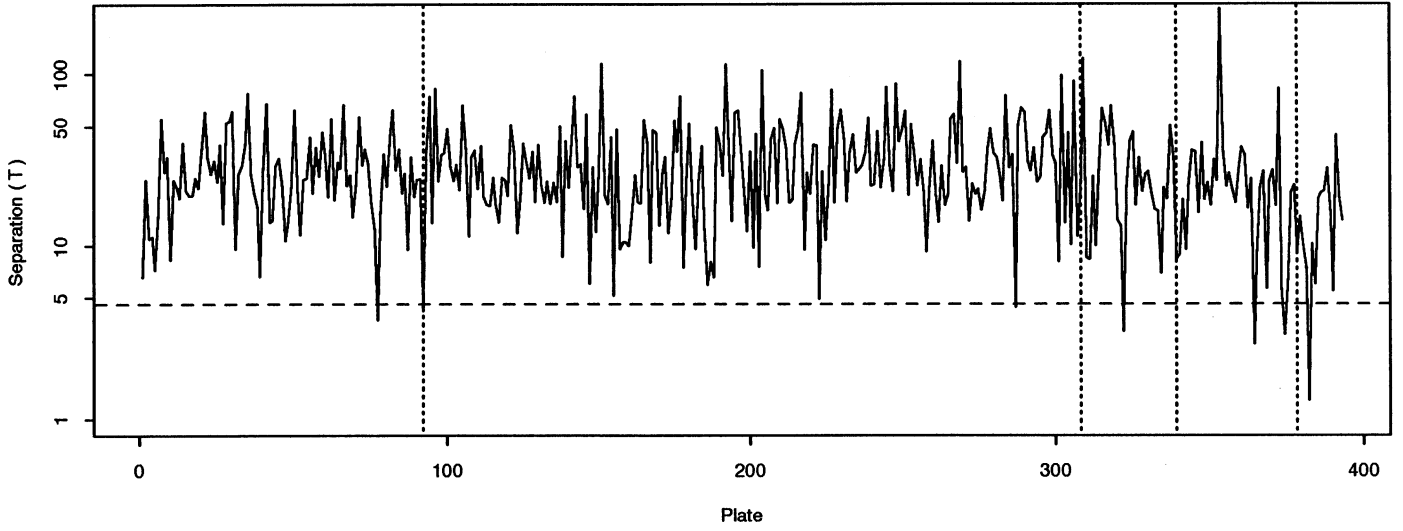
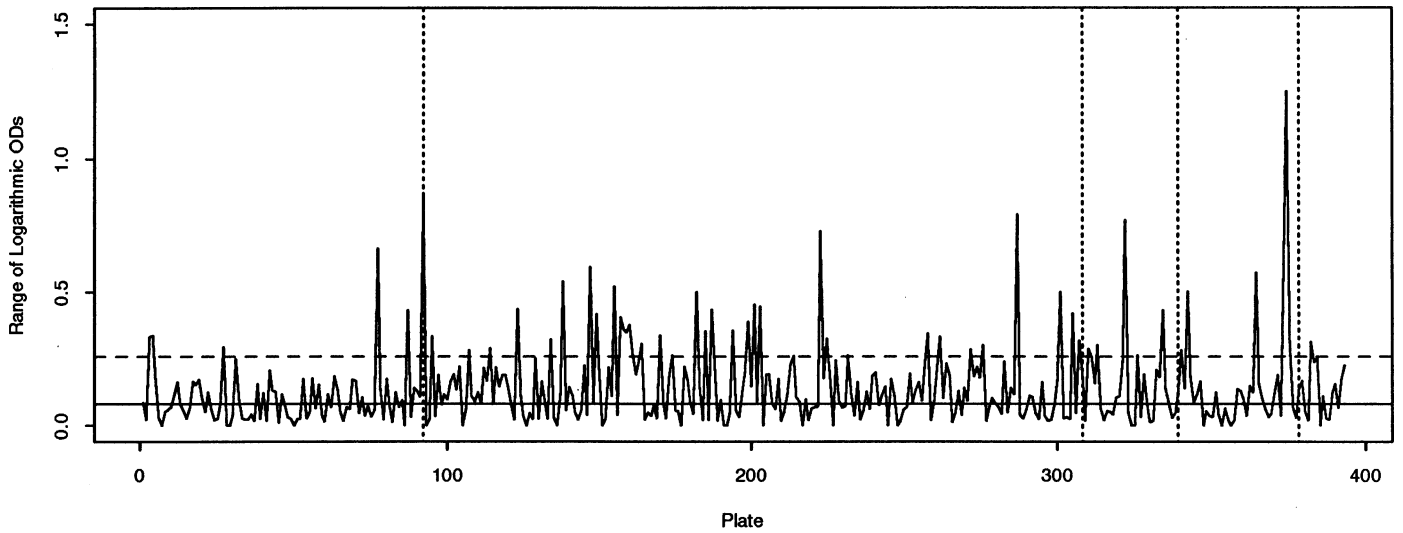


Figure 7: T-R Charts for 1-18-88 to 4-13-88

T Chart of Logged Controls



R Chart of Logged Negative Controls



R Chart of Logged Positive Controls

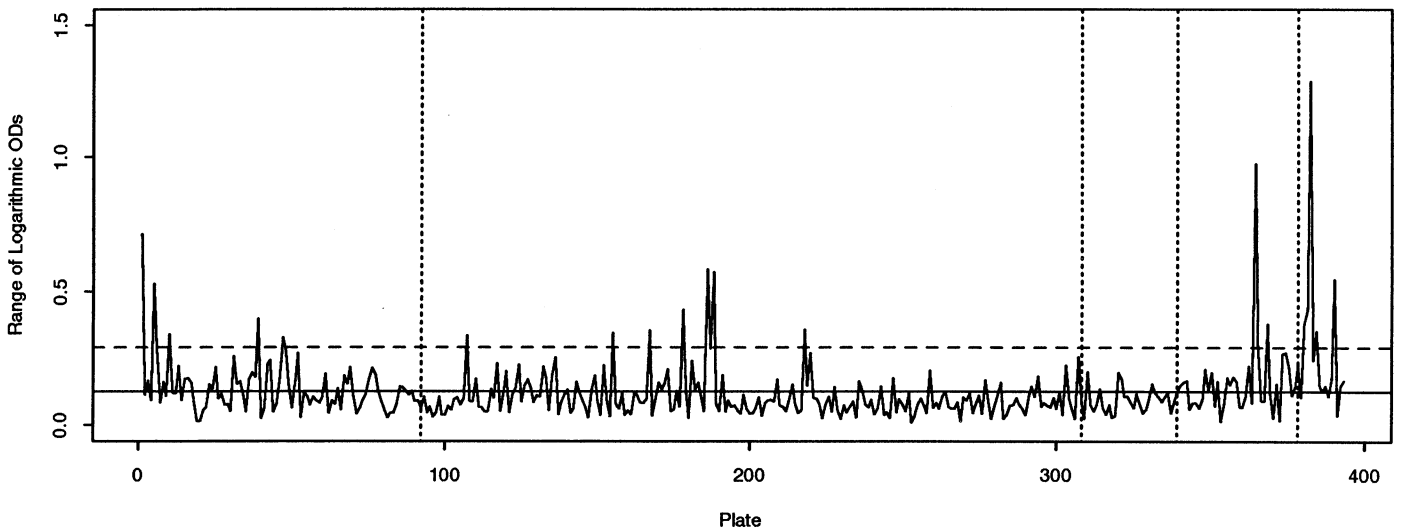
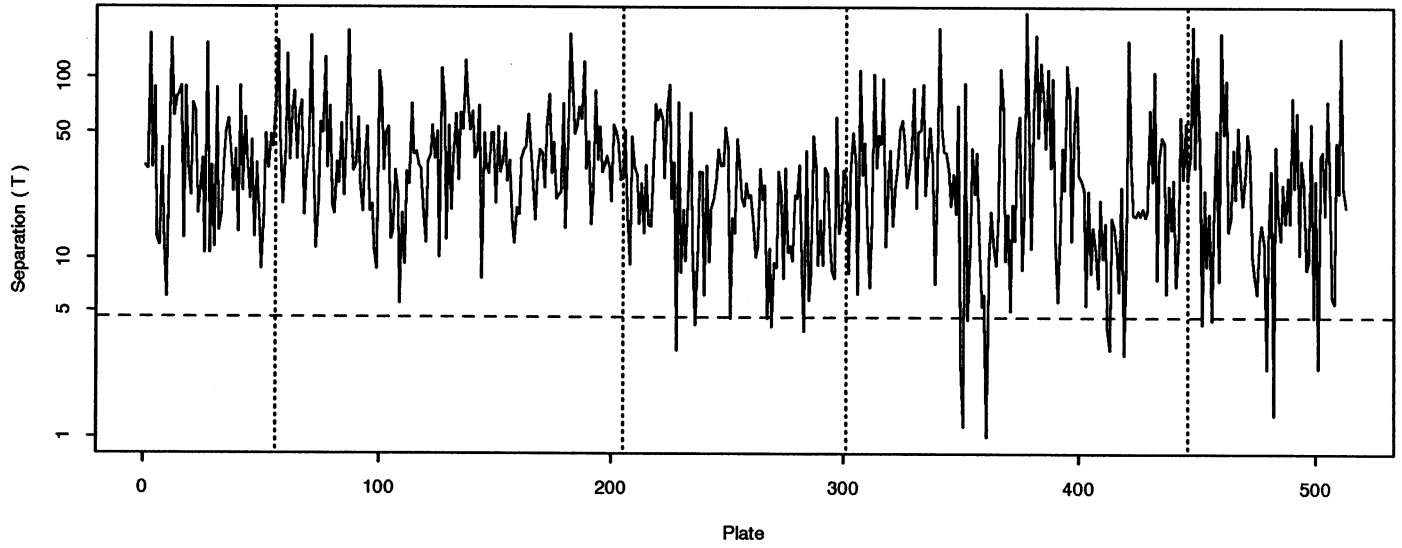
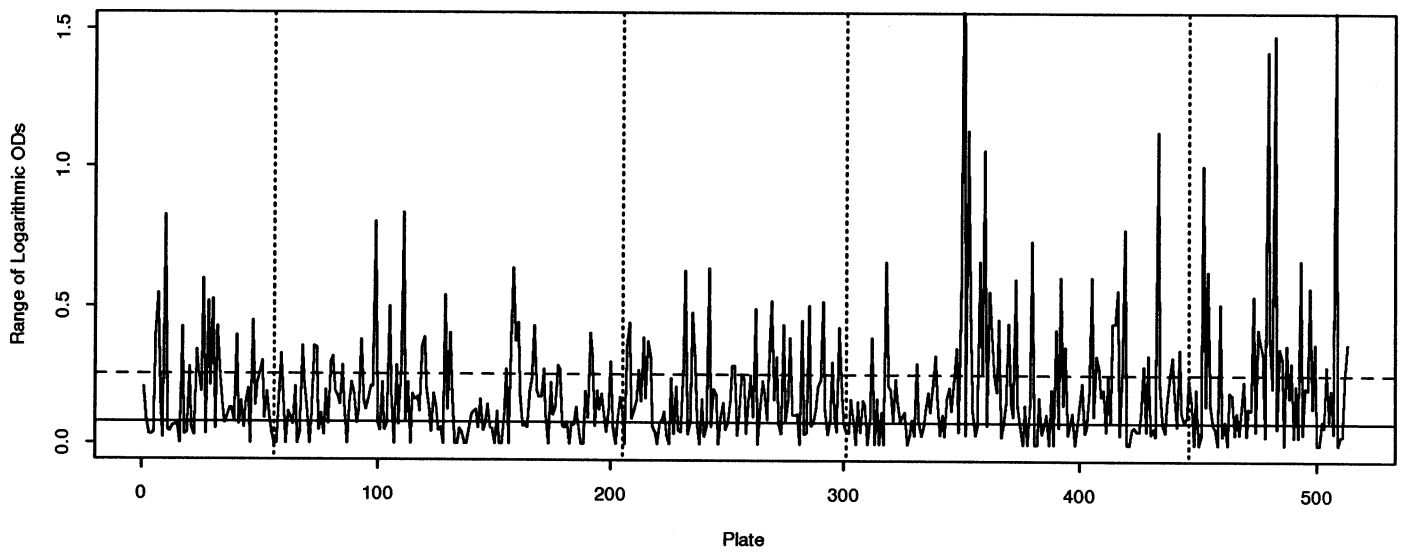


Figure 8: T-R Charts for 7-21-88 to 11-16-88

T Chart of Logged Controls



R Chart of Logged Negative Controls



R Chart of Logged Positive Controls

