

# A New Strategy for Phase I Analysis in SPC

Jyh-Jen Horng Shiau<sup>\*†</sup> and Jian-Huang Sun

The Phase I analysis in statistical process control usually includes a task of filtering out out-of-control data in the historical data set via control charting. The conventional procedure for this is an iterative procedure that first uses all the samples to set up initial trial control limits and discards *all* the 'out-of-control' samples accordingly, and then iteratively repeats the screening step on the remaining samples until no more 'out-of-control' samples are detected. For simplicity, the 'out-of-control' samples here refer to the samples with their monitoring statistics exceeding the trial control limits. It is found in this study that this procedure throws away too many useful in-control samples. To overcome this drawback, we propose and study a new iterative procedure that discards only *one* 'out-of-control' sample (i.e. the most extreme one) at each iteration. Our simulation study, using the Shewhart  $\bar{X}$  Chart for illustration, demonstrates that the new one-at-a-time procedure reduces dramatically the occurrences of false alarms. For cost-saving, we further suggest a new strategy on when to stop and inspect the process to look for assignable causes for samples signaling out-of-control alarms. To determine the control limits, both the traditional method that controls the individual false-alarm-rate and the Bonferroni method that controls the overall false-alarm-rate are considered. The performances of the proposed schemes are evaluated and compared in terms of the false-alarm rate and the detecting power via simulation studies. Copyright © 2009 John Wiley & Sons, Ltd.

**Keywords:** Bonferroni's adjustment; individual false-alarm rate; multiple tests; overall false-alarm rate; signal probability

## 1. Introduction

In statistical process control (SPC), control chart applications are often distinguished into Phase I and Phase II. In Phase I application, control charts are used off-line to determine retrospectively from a set of so-called historical data whether the process has been in control. In Phase II application, control charts are used online mainly for continual process monitoring. The purpose of this paper is to propose and study a new strategy for Phase I analysis.

In Phase I, process data are collected and analyzed with the goals to bring the process to a state of statistical control, and then to model the in-control process so that reliable control limits of the control chart can be established for online process monitoring later in Phase II. Woodall<sup>1</sup> pointed out that significant efforts for process understanding and process improvement are often required in the transition from Phase I to Phase II.

To construct a useful control chart for the online process monitoring, a suitable process/product quality-related monitoring statistic is chosen and control limits are set according to its in-control distribution. Phase I control charting is usually used to screen out out-of-control data from the data set so that a set of presumably in-control process data can be obtained to model the distribution of the monitoring statistic.

Some estimation problems in constructing control charts are explicitly mentioned by Woodall and Montgomery<sup>2</sup> and other researchers. In particular, Jones and Champ<sup>3</sup> pointed out that the decisions of being in- or out-of-control made by the  $\bar{X}$  control chart for the  $m$  samples in a Phase I data set are dependent because all the  $m$  monitoring statistics involve the same sample mean and standard deviation estimated from the Phase I samples. Thus, in order to have a fixed overall false-alarm rate, the control limits must be based on the joint distribution of the  $m$  charting statistics.

Assume that a set of samples, often called a historical data set, had been collected from the process. As some of these samples may not come from the in-control process, one would need a screening procedure. The conventional practice for Phase I analysis is an iterative procedure described below:

- (i) First, use the data to set up a set of initial *trial* control limits for the monitoring statistic, such as  $\bar{X}$ ,  $R$ , or  $S$ , to identify potential 'out-of-control' points. For simplicity, we only consider the charted points that exceed the control limits as the 'out-of-control' points in this study.

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan 30010, Taiwan

\*Correspondence to: Jyh-Jen Horng Shiau, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan 30010, Taiwan.

†E-mail: jyhjen@stat.nctu.edu.tw

Contract/grant sponsor: National Research Council of Taiwan; contract/grant numbers: NSC93-2118-M009-007, NSC95-2118-M-009-006-MY2

- (ii) If some samples signal 'out-of-control', then the operators or process engineers should investigate the process to see if there exist any assignable causes to explain why these points are out-of-control.
  - (a) If indeed some assignable causes are found, then appropriate corrective actions should be taken to eliminate the causes. If any of the corrective actions changes the nature of the process, one needs to recollect data from the new process and restart the screening process. If eliminating assignable causes does not affect the current process in nature, then we can simply discard the 'out-of-control' points and redo the control charting using the remaining data.
  - (b) If no assignable causes can be found, then one can choose either to keep or to discard these 'out-of-control' points. No one really knows which action is correct without further information as these points may exceed the limits simply by chance or by some uncovered assignable causes. For being conservative, many practitioners may choose to discard these 'out-of-control' points to avoid potential contamination in the data set.
- (iii) Repeat the above screening steps based on the remaining data set until no more 'out-of-control' points can be found.

Statistically, in any control charting, there are possibilities that some in-control samples may get wrongly discarded and some out-of-control samples may remain undetected, which are similar to committing Type I and Type II errors in hypothesis testing, respectively. A good control chart should be able to control these two types of error rates to some extent.

In this paper, we study this procedure and find surprisingly that the discard-all practice tends to mistakenly screen out too many (i.e. more than expected) in-control data points. To overcome this drawback, we propose a more effective iterative procedure for collecting in-control data by simply discarding, instead of *all*, but only *one* 'out-of-control' point (the most extreme one) and then updating the trial control limits at each iteration. This procedure will be referred to as the one-at-a-time (OAAT) procedure hereafter.

It is found from the simulation study that, with control limits constructed under the same *overall* false-alarm rate (i.e. the probability that at least one sample point in the data set signals 'out-of-control' when the process is in-control), the proposed OAAT procedure screens out much less in-control samples than the conventional discard-all procedure and still retains about the same power in detecting out-of-control samples.

With the proposed OAAT procedure, it is likely that there may be more iterations than the conventional procedure. To be more cost effective, instead of stopping the process at the end of each iteration to look for assignable causes for each of the 'out-of-control' points as the conventional practice does, we suggest a new strategy of performing the investigation for all of the 'out-of-control' points detected after the whole OAAT procedure ends. Note that the final number of 'out-of-control' points is most likely less than or equal to that of the conventional discard-all procedure. Thus, by reducing the frequency of investigations and possible process adjustments, the new practice would be more cost effective. The only disadvantage of the new procedure is very minor—it requires slightly more computing power, which is no longer an issue with the enhancing computer power nowadays.

We remark that, even with the conventional discard-all procedure, we would also recommend applying the new strategy of leaving all the investigation to the end when no more 'out-of-control' points can be detected. In this way, we only need to stop the process once and investigate all potential problems suggested by the current data set at that time.

Another important issue we would like to address in this paper is the criteria of performance evaluation of control charts used in Phase I analysis. Currently in the literature the evaluation is mostly based on the so-called 'signal probability' (e.g. Sullivan and Woodall<sup>4</sup>), which is defined as the family-wise signal rate, the probability that at least one sample point in the data set signals 'out-of-control'. The overall false-alarm rate mentioned above is the signal probability when the process is in control. Therefore, the signal-probability criterion can only evaluate the effectiveness of the control schemes on judging whether the whole data set comes from the in-control process or not.

As the signal probability does not differentiate true and false alarms for 'out-of-control' signals, it may not be a good measure to compare the performances of different screening methods. To fit the purpose of Phase I analysis better, we suggest using simultaneously the rate of correctly rejected samples and the rate of wrongly rejected samples as the comparison criteria. The former measures the detecting power (true-alarm rate) and the latter measures the false-alarm rate.

In early days, control charts were designed based on controlling the individual false-alarm rate for each sample no matter how many samples were tested (e.g. Hunter<sup>5</sup>, Vermaat *et al.*<sup>6</sup>). Recently many research works on Phase I analysis were based on controlling the overall false-alarm rate instead. With a fixed individual false-alarm rate, the overall false-alarm rate gets larger when the number of samples  $m$  gets larger. This is why many authors chose to adjust the individual false-alarm rate with Bonferroni-type procedures in order to control the overall false-alarm rate at a desirable level. For example, see Borror and Champ<sup>7</sup>, Champ and Chou<sup>8</sup>, Mahmoud and Woodall<sup>9</sup>, and Nedumaran and Pignatiello<sup>10</sup>. Nedumaran and Pignatiello<sup>11</sup> discussed many issues in constructing retrospective  $\bar{X}$  control chart limits when controlling the overall false-alarm rate at a desired level.

For its simplicity and popularity, we use Shewhart  $\bar{X}$  control chart to illustrate our new Phase I analysis strategy.

The remainder of this paper is organized as follows. Section 2 describes the conventional method in detail and explains why the discard-all practice has the problem of excessive false alarms. Section 3 presents the new strategy and the OAAT procedure. Section 4 compares the conventional discard-all method and the new OAAT method under controlling the individual false-alarm rate as well as under controlling the overall false-alarm rate by Bonferroni's adjustment. Section 5 concludes the paper with a brief summary and some remarks.

## 2. The conventional method

Let the in-control process be a normal process with mean  $\mu_0$  and variance  $\sigma_0^2$ . Assume that the set of process data in Phase I is in the form of  $m$  independent random samples of size  $n$ ,  $\{X_{i1}, \dots, X_{in}\}_{i=1}^m$ , taken in the order of process output. When the  $m$  samples are all from the in-control process,  $X_{ij}$ 's are independent and identically distributed (i.i.d.) as  $N(\mu_0, \sigma_0^2)$  for  $i=1, 2, \dots, m$  and  $j=1, 2, \dots, n$ . We first estimate  $\mu_0$  and  $\sigma_0^2$ , and then construct the Phase I control chart accordingly as follows.

### 2.1. Estimating process parameters

The most commonly used estimator of  $\mu_0$  is

$$\hat{\mu}_0 = \bar{\bar{X}} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n X_{ij} = \frac{1}{m} \sum_{i=1}^m \bar{X}_i$$

where  $\bar{X}_i$  is the sample mean of the  $i$ th sample. There are many estimators of  $\sigma_0$  given in the literature, see, e.g. Champ and Chou<sup>8</sup>, Nedumaran and Pignatiello<sup>11</sup>, and Champ and Jones<sup>12</sup>. The following are three variability-related statistics considered in Champ and Chou<sup>8</sup>:

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i, \quad \bar{S} = \frac{1}{m} \sum_{i=1}^m S_i \quad \text{and} \quad \bar{V}^{1/2} = \left( \frac{1}{m} \sum_{i=1}^m S_i^2 \right)^{1/2}$$

where  $R_i = \max(X_{i1}, \dots, X_{in}) - \min(X_{i1}, \dots, X_{in})$  is the sample range and  $S_i = (\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 / (n-1))^{1/2}$  is the sample standard deviation of the  $i$ th sample. Then by rescaling, the corresponding unbiased estimators of  $\sigma_0$  are

$$\frac{\bar{R}}{d_2}, \quad \frac{\bar{S}}{c_4} \quad \text{and} \quad \hat{\sigma}_0 = \frac{\bar{V}^{1/2}}{c_{4,m}}$$

where  $d_2 = E(R) / \sigma_0$ , which can be easily found in many quality control textbooks such as Montgomery<sup>13</sup>,

$$c_4 = \frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)} \quad \text{and} \quad c_{4,m} = \frac{\sqrt{2}\Gamma((m(n-1)+1)/2)}{\sqrt{m(n-1)}\Gamma(m(n-1)/2)}$$

Champ and Chou<sup>8</sup> chose  $\bar{V}^{1/2} / c_{4,m}$  as their preferred estimator of  $\sigma_0$  among the three by showing that

$$\text{Var} \left( \frac{\bar{V}^{1/2}}{c_{4,m}} \right) \leq \text{Var} \left( \frac{\bar{S}}{c_4} \right) \leq \text{Var} \left( \frac{\bar{R}}{d_2} \right)$$

Following their arguments, we adopt  $\hat{\sigma}_0 = \bar{V}^{1/2} / c_{4,m}$  in this paper.

### 2.2. Phase I Shewhart control chart

The Phase I Shewhart  $\bar{X}$  chart is constructed with the following lower control limit (LCL), center line (CL), and upper control limit (UCL):

$$LCL = \hat{\mu}_0 - k \frac{\hat{\sigma}_0}{\sqrt{n}}, \quad CL = \hat{\mu}_0 \quad \text{and} \quad UCL = \hat{\mu}_0 + k \frac{\hat{\sigma}_0}{\sqrt{n}}$$

where the multiple  $k$  is chosen to obtain the desirable false-alarm rate. The usual Shewhart chart takes  $k$  as 3. The statistic  $\bar{X}_i$  (the  $i$ th sample mean) is plotted against the sample number  $i$ . If any point falls below LCL or above UCL, it is taken as an evidence that the corresponding sample is out-of-control.

### 2.3. The individual and overall false-alarm rates

For the case when the  $m$  samples are all from the in-control process, it is easy to verify that  $\bar{X}_i - \bar{\bar{X}}$  is distributed as  $N(0, (m-1)\sigma_0^2 / (mn))$  for  $i=1, 2, \dots, m$ ,  $m(n-1)\bar{V} / \sigma_0^2$  is distributed as  $\chi_{m(n-1)}^2$ , the  $\chi^2$  distribution with degrees of freedom  $m(n-1)$ , and, furthermore,  $\bar{X}_i - \bar{\bar{X}}$  and  $m(n-1)\bar{V} / \sigma_0^2$  are independent. Therefore,

$$\frac{\sqrt{mn}(\bar{X}_i - \bar{\bar{X}})}{\sqrt{(m-1)}\sqrt{\bar{V}}} = \frac{\sqrt{mn}(\bar{X}_i - \bar{\bar{X}})}{\sqrt{(m-1)}\sigma_0^2} \bigg/ \sqrt{\frac{\bar{V}}{\sigma_0^2}}$$

**Table I.** The individual false-alarm rate,  $\alpha^*$ , and the overall false-alarm rate,  $\alpha$ , for various values of  $m$  and  $n$  when  $k=3$ , assuming  $m$  tests are independent

$m$	$n$	$c_{4,m}$	$\alpha^*$	$\alpha$
30	5	0.9979	0.0028	0.0793
30	10	0.9991	0.0025	0.0719
30	15	0.9994	0.0024	0.0698
50	5	0.9988	0.0027	0.1278
50	10	0.9994	0.0026	0.1207
50	15	0.9996	0.0025	0.1187
100	5	0.9994	0.0027	0.2381
100	10	0.9997	0.0026	0.2318
100	15	0.9998	0.0026	0.2300

is distributed as  $t_{m(n-1)}$ , the  $t$  distribution with degrees of freedom  $m(n-1)$ . As

$$\frac{\sqrt{mn}(UCL - \bar{X})}{\sqrt{(m-1)\sqrt{V}}} = \frac{\sqrt{mn}(k\hat{\sigma}_0 / \sqrt{n})}{\sqrt{(m-1)\sqrt{V}}} = \frac{k\sqrt{m}}{c_{4,m}\sqrt{m-1}}$$

the individual false-alarm rate for each sample is

$$\alpha^* = 2 \left( 1 - F_{t_{m(n-1)}} \left( \frac{k\sqrt{m}}{c_{4,m}\sqrt{m-1}} \right) \right)$$

where  $F_{t_{m(n-1)}}$  is the cumulative distribution function (c.d.f.) of the  $t_{m(n-1)}$  distribution.

If the  $m$  hypothesis tests are independent and the individual false-alarm rate of each test is  $\alpha^*$ , then the overall false-alarm rate  $\alpha = 1 - (1 - \alpha^*)^m$ . Table I lists the  $\alpha^*$  and the corresponding  $\alpha$  for various  $m$  and  $n$  when  $k=3$  by assuming that the  $m$  tests are independent. Note that these  $\alpha^*$ 's are not too much different. Also, the overall false-alarm rate  $\alpha$  increases as  $m$  increases or as  $n$  decreases as expected.

However, as described before, the independence assumption does not hold for control charts in Phase I application as all  $m$  tests use common parameter estimators,  $\bar{X}$  and  $\sqrt{V}$ . Fortunately, the assumption violation does not have significant impact on the false-alarm rate. Our simulation study indicates that the overall false-alarm rate for the case of  $m=30$ ,  $n=5$ , and  $k=3$  is about 0.078 (see Section 3.3), not too far from that of independent tests, 0.0793, given in Table I.

#### 2.4. Excessive false-alarm rate

The following argument explains why the discard-all approach tends to have more than expected false alarms. Assume that the  $n$  observations in a sample are i.i.d. Let observations from the in-control process follow  $N(\mu_0, \sigma_0^2)$ . Suppose that the historical data set is contaminated with  $100p$  percent of samples from an out-of-control process with process mean shifted from  $\mu_0$  to  $\mu_1$ . Then for any randomly selected sample,  $\bar{X}$  is distributed as a mixture distribution with mean  $\mu = (1-p)\mu_0 + p\mu_1$ . Thus, if we mistakenly treat all the data as from the in-control process and proceed control charting; we actually construct a control chart with the following *theoretical* center line and control limits:

$$CL = (1-p)\mu_0 + p\mu_1, \quad LCL = CL - k \frac{\sigma_0}{\sqrt{n}}, \quad UCL = CL + k \frac{\sigma_0}{\sqrt{n}}$$

Then, the individual false-alarm rate of a randomly selected sample is

$$\begin{aligned} \alpha^{**} &= P(\bar{X} > UCL \text{ or } \bar{X} < LCL | \mu = \mu_0) \\ &= 1 - \Phi(\sqrt{n}p\delta + k) + \Phi(\sqrt{n}p\delta - k) \end{aligned}$$

where  $\delta = (\mu_1 - \mu_0) / \sigma_0$  and  $\Phi$  is the c.d.f. of the standard normal distribution. Also, the detecting power of an individual test for  $\mu = \mu_1$  is

$$\begin{aligned} 1 - \beta^{**} &= P(\bar{X} > UCL_0 \text{ or } \bar{X} < LCL_0 | \mu = \mu_1) \\ &= 1 - \Phi(-\sqrt{n}(1-p)\delta + k) + \Phi(-\sqrt{n}(1-p)\delta - k) \end{aligned}$$

For the case of  $k=3$  and  $n=5$ , Tables II and III list the individual false-alarm rate  $\alpha^{**}$  and the detecting power  $1 - \beta^{**}$  for various values of the out-of-control proportion  $p$  and shift size  $\delta$ , respectively. When the process is in control (e.g.  $p=0$ ),  $\alpha^{**} = 0.0027$ , the nominal false-alarm level. Note that for a fixed shift size  $\delta$ , as  $p$  increases from 0, the false-alarm rate  $\alpha^{**}$  increases from 0.0027. This explains why there are more than expected false alarms when data are contaminated. Moreover, we also observe

**Table II.** The individual false-alarm rate,  $\alpha^{**}$ , for the case of  $k=3$  and  $n=5$ .  $p$  is the proportion of the out-of-control process with shift size  $\delta=(\mu_1-\mu_0)/\sigma_0$

$\delta$	$p=0.1$	$p=0.2$	$p=0.3$	$p=0.4$	$p=0.5$
0.4	0.002807	0.003132	0.003692	0.004511	0.005626
0.8	0.003132	0.004511	0.007085	0.011274	0.017670
1.2	0.003692	0.007085	0.014152	0.027032	0.048630
1.6	0.004511	0.011274	0.027032	0.058338	0.112921
2.0	0.005626	0.017670	0.048630	0.112921	0.222454
2.4	0.007085	0.027032	0.082262	0.196726	0.375729
2.8	0.008945	0.040260	0.130995	0.310087	0.551913
3.2	0.011274	0.058338	0.196726	0.445186	0.718270
3.6	0.014152	0.082262	0.279258	0.587040	0.847300
4.0	0.017670	0.112921	0.375729	0.718270	0.929508

**Table III.** The detecting power of an individual test,  $1-\beta^{**}$ , for the case of  $k=3$  and  $n=5$ .  $p$  is the proportion of the out-of-control process with shift size  $\delta=(\mu_1-\mu_0)/\sigma_0$

$\delta$	$p=0.1$	$p=0.2$	$p=0.3$	$p=0.4$	$p=0.5$
0.4	0.014152	0.011274	0.008945	0.007085	0.005626
0.8	0.082262	0.058338	0.040260	0.027032	0.017670
1.2	0.279258	0.196726	0.130995	0.082262	0.048630
1.6	0.587040	0.445186	0.310087	0.196726	0.112921
2.0	0.847300	0.718270	0.551913	0.375729	0.222454
2.4	0.966368	0.902038	0.775353	0.587040	0.375729
2.8	0.995792	0.977720	0.916621	0.775353	0.551913
3.2	0.999709	0.996778	0.977720	0.902038	0.718270
3.6	0.999989	0.999709	0.995792	0.966368	0.847300
4.0	1.000000	0.999984	0.999445	0.991023	0.929508

that the detecting power  $1-\beta^{**}$  decreases as  $p$  increases. This indicates that the conventional discard-all approach not only has problems with the false-alarm rate but also loses its detecting power quickly as the proportion of contamination  $p$  gets larger.

### 3. New strategy and the OAAT method for Phase I analysis

#### 3.1. Criteria for performance evaluation

Consider the problem of testing each of the  $m$  Phase I samples for being in control or not. Let  $m_0$  and  $m_1$  denote the numbers of in-control and out-of-control samples, respectively. Assume that  $R$  samples are rejected; and among them,  $R_0$  in-control samples are falsely rejected and  $R_1$  out-of-control samples are correctly rejected.

In Phase I, one of the goals is to determine whether the process is in control. So many research works use the signal probability described above as the performance measure to compare different control charts. However, practically, for the purposes of bringing the process to a state of statistical control and collecting in-control data, we not only want to know if the process is in control, but also need to know which samples are out-of-control when the answer is negative. The signal probability can only provide a measure for the first question. Note that when the historical data set contains a mixture of in-control and out-of-control data (i.e. when  $m > m_1 > 0$ ), then the signal probability (denoted by  $P$ ) is not the overall false-alarm rate (unless  $m_1=0$ ) nor the detecting power (unless  $m_1=m$ ), and out-of-control signals can be triggered by either true or false alarms. In other words, a high signal probability can be caused by high detecting power or by high false-alarm rate or both, without differentiation. For this reason, evaluating the performance by the signal probability is somewhat questionable.

It seems more realistic to evaluate a control scheme in Phase I analysis based on its ability of making correct decisions on samples—either in-control or out-of-control. Thus we propose using measures of false rejections and correct rejections simultaneously as evaluation criteria. The former has the flavor of the Type I error in hypothesis testing and the latter measures more or less the detecting power of the scheme.

### 3.2. A new strategy on when to inspect

As mentioned before, there may be some out-of-control samples undetected at each iteration of the control chart construction, especially when the trial control limits are calculated with a set of data contaminated by some out-of-control samples. If we inspect the process to look for assignable causes for every alarm whenever it signals, as the current practice suggests, the frequency of 'stop-and-inspect' actions may be unnecessarily high. Here we suggest a new strategy: run through the whole iterative procedure and then perform the inspection for assignable causes for all of the 'out-of-control' points at the end. This new strategy should be able to reduce the frequency of stop-and-inspect actions.

### 3.3. An illustrative simulation study of discard-all practice

Many SPC books have recommended that 20–30 subgroups of size 4 or 5 be used for estimating the process parameters (see, e.g. Montgomery<sup>13</sup>). Thus, we choose the setting of  $m=30$ ,  $n=5$ , and  $k=3$  for illustrating the behaviors of the above criteria in our simulation study. Without loss of generality, let  $\mu_0=0$  and  $\sigma_0=1$ . Denote the shift size of the process mean by  $\delta$ .

Consider  $m_1=0, 3, 6, 9, 12$  and  $\delta=0 (0.4)$ . For each combination of  $m_1$  and  $\delta$ , we simulate 1 000 000 data sets, each containing  $m_0=m-m_1$  random samples generated from  $N(0, 1)$  and  $m_1$  random samples from  $N(\delta, 1)$ . Each data set produces its own  $R_0$  and  $R_1$  under the conventional discard-all practice. When  $R_0+R_1>0$ , the alarm signals. We estimate the signal probability  $P$  by  $\hat{P}$ , the sample proportion of such signals among the 1 000 000 data sets. Let  $\bar{R}_0$ , the average of the 1 000 000  $R_0$ 's, be the estimate of  $E(R_0)$ , and  $\bar{R}_1$ , the average of the 1 000 000  $R_1$ 's, be the estimate of  $E(R_1)$ .

For the 41 combinations of  $m_1$  and  $\delta$  considered, the standard errors of  $\hat{P}$  are about 0–0.00049, and the standard errors of  $\bar{R}_0$  and  $\bar{R}_1$  are about 0–0.00227. Table IV shows the values of  $\hat{P}$ ,  $\bar{R}_0$ , and  $\bar{R}_1$ . We observe the following from the table:

- For  $m_1=0$  (i.e. the process is in control),  $\hat{P}$ , now estimating the overall false-alarm rate, is about 0.078. On the other hand,  $\bar{R}_0$ , the estimated false-alarm rate, is 0.0828, which is greater than  $\hat{P}=0.078$ . The reason is that, when  $R_0 \geq 1$ , no matter how many (false) alarms exist in the data set, it only counts as 1 toward estimating  $P$ .
- As expected,  $\bar{R}_1$  increases as the size of the shift  $\delta$  increases. But  $\bar{R}_0$  increases as well—this is the cost of instability.
- As the percentage of the out-of-control samples increases, the false-alarm rate ( $\bar{R}_0/m_0$ ) increases.
- $\bar{R}_1$  moves in the same direction as  $\hat{P}$ , and yet it contains more information than  $\hat{P}$ . It indicates, on average, how many true out-of-control samples the method can detect. On the other hand,  $\bar{R}_0$  measures how many false alarms can occur on average. As to  $\hat{P}$ , recall that when  $0 < m_1 < m$ ,  $\hat{P}$  is just the alarm-signaling rate, and alarms can be triggered by samples from either state. Also note that  $\hat{P}$  reaches 1 when the shift size  $\delta$  gets to 2.4 while  $\bar{R}_0$  and  $\bar{R}_1$  are still growing as  $\delta$  gets larger.

By examining the simulation results of the discard-all practice, it is noted that the number of false alarms is higher than we would expect.

### 3.4. The OAAT method

To reduce the wasteful excessive false alarms as shown by the above simulation study, we propose and study a new practice—discard only one sample at a time instead of discarding all as in the conventional method. Our simulation study shows that this OAAT procedure can reduce the number of false alarms dramatically, which in turn would reduce the amount of time in conducting unnecessary investigations for non-existing assignable causes and reserve more in-control data so that process modeling would be more efficient. The main reason why this procedure works is that the most extreme point is more likely to be an out-of-control sample than others. In contrast, the conventional method discards all the 'out-of-control' samples at each iteration, thus being more vulnerable for some of the discarded samples to be in-control samples.

We describe the OAAT procedure along with the new strategy below:

*Step 1.* Construct the trial control limits with all collected data.

*Step 2.* If no 'out-of-control' samples are identified with the control limits, stop iterating and go to Step 4; otherwise, discard the most extreme sample.

*Step 3.* Construct the trial control limits with the remaining samples; go to Step 2.

*Step 4.* If there is no sample discarded, claim the process is in control; otherwise collect all the samples discarded in the above iterations and inspect the process for assignable causes.

## 4. The performance of the OAAT procedure

We conduct a simulation study to evaluate the performance of the new OAAT procedure. Four factors that would influence the performance are studied, including the number of samples ( $m$ ), the subgroup size ( $n$ ), the proportion ( $p$ ) of the samples that are generated from the out-of-control process, and of course the size of the process shift ( $\delta$ ). Both the individual and overall false-alarm rates are studied.

**Table IV.**  $\hat{P}$ ,  $\bar{R}_0$ , and  $\bar{R}_1$  of the discard-all procedure based on 1 000 000 replications for various values of  $m_1$  and  $\delta$  when  $m=30$ ,  $n=5$ , and  $k=3$

$m_1$	$\delta$	$\hat{P}$	$\bar{R}_0/m_0$	$\bar{R}_1/m_1$
0	0	0.0780	0.0828/30	0/0
3 (10%)	0.4	0.1104	0.0769/27	0.0426/3
	0.8	0.2853	0.0867/27	0.2478/3
	1.2	0.6530	0.1011/27	0.8394/3
	1.6	0.9336	0.1241/27	1.7609/3
	2	0.9965	0.1539/27	2.5399/3
	2.4	1	0.1940/27	2.8986/3
	2.8	1	0.2445/27	2.9872/3
	3.2	1	0.3085/27	2.9991/3
	3.6	1	0.3864/27	3.0000/3
6 (20%)	0.4	0.1322	0.0767/24	0.0684/6
	0.8	0.3651	0.1101/24	0.3512/6
	1.2	0.7590	0.1726/24	1.1836/6
	1.6	0.9729	0.2742/24	2.6721/6
	2	0.9995	0.4276/24	4.3063/6
	2.4	1	0.6549/24	5.4090/6
	2.8	1	0.9736/24	5.8643/6
	3.2	1	1.4109/24	5.9802/6
	3.6	1	1.9864/24	5.9981/6
9 (30%)	0.4	0.1442	0.0784/21	0.0815/9
	0.8	0.3926	0.1514/21	0.3652/9
	1.2	0.7664	0.3005/21	1.1841/9
	1.6	0.9715	0.5729/21	2.7942/9
	2	0.9994	1.0277/21	4.9658/9
	2.4	1	1.7382/21	6.9747/9
	2.8	1	2.7604/21	8.2450/9
	3.2	1	4.1422/21	8.7973/9
	3.6	1	5.8693/21	8.9612/9
12 (40%)	0.4	0.1521	0.0829/18	0.0870/12
	0.8	0.3979	0.2054/18	0.3276/12
	1.2	0.7475	0.4906/18	0.9919/12
	1.6	0.9582	1.0576/18	2.3681/12
	2	0.9983	2.0402/18	4.5105/12
	2.4	1	3.5516/18	7.0427/12
	2.8	1	5.5863/18	9.2968/12
	3.2	1	8.0168/18	10.8186/12
	3.6	1	10.5676/18	11.5927/12
4	1	12.9224/18	11.8906/12	

#### 4.1. Controlling the individual false-alarm rate

Set  $m=30, 50, 100$  and  $n=5, 10, 15$ . For each combination, consider the cases of  $m_1=0, 0.1m, 0.2m, 0.3m, 0.4m$  and  $\delta=0.4$  (0.4) 4. In this subsection, we study the case when the individual false-alarm rate is controlled. For simplicity, set  $k=3$ . For each scenario, simulate 1 000 000 data sets and calculate  $\bar{R}_0$  and  $\bar{R}_1$  as described in Section 3.3. The estimated standard errors of  $\bar{R}_0$  and  $\bar{R}_1$  are about 0–0.0069.

Table V gives the  $\bar{R}_0$  of the two procedures (discard-all versus OAAT) when all the  $m$  samples are from the in-control process. Note that the  $\bar{R}_0$  of the OAAT procedure is less than that of the discard-all procedure in every case under study. This demonstrates that the OAAT procedure does signal fewer false alarms. However, the reduction is not extensive, less than 2.1%.



**Table V.** Average numbers of false alarms,  $\bar{R}_0$ , of the discard-all procedure and the OAAT procedure when all the samples are *in control* for various combinations of  $m$  and  $n$ . Here  $k=3$

$m$	$n$	Discard-all	OAAT
30	5	0.0834	0.0818
30	10	0.075	0.0738
30	15	0.0726	0.0716
50	5	0.1384	0.1362
50	10	0.1291	0.1273
50	15	0.1271	0.1254
100	5	0.2749	0.2711
100	10	0.2651	0.2617
100	15	0.2639	0.2606

Figures 1 and 2 plot, respectively, for  $n=5$  and 15 the simulated  $\gamma_1 = \bar{R}_1/m_1$  and  $\gamma_0 = \bar{R}_0/m_0$  versus  $\delta$  for various values of  $m$  and  $m_1$ . The case of  $n=10$  has the results in between and is omitted to save space. The following are observed:

- The detecting power ( $\bar{R}_1/m_1$ ) of the OAAT procedure is slightly better than that of the discard-all procedure in general and the advantage is more profound as the proportion of 'out-of-control' samples  $p$  increases.
- The false-alarm rate ( $\bar{R}_0/m_0$ ) of the OAAT procedure is uniformly smaller than that of the discard-all procedure and the advantage is more profound as the proportion of 'out-of-control' samples  $p$  increases. Note that all the  $\bar{R}_0/m_0$  curves of the OAAT procedure lie flat on  $x$ -axis, which indicates that the new OAAT procedure seldom signals false alarms.
- We note that the improvement of the OAAT procedure over the discard-all procedure in terms of the false-alarm rate is extremely large for many cases, especially when  $m$ ,  $n$ , and/or  $p$  are large.
- As  $n$  increases, the detecting power ( $\bar{R}_1/m_1$ ) of either procedure increases; the false-alarm rate ( $\bar{R}_0/m_0$ ) of the discard-all procedure increases, but that of the OAAT procedure decreases in most cases; and the advantage of the OAAT procedure in the false-alarm rate ( $\bar{R}_0/m_0$ ) is more profound.
- As to the effect of  $m$ , it is observed that the false-alarm rates ( $\bar{R}_0/m_0$ ) are fairly close for various values of  $m$  when  $n$  and  $p$  are fixed. This is because we control the individual false-alarm rate by setting  $k=3$  for each scenario; hence, the theoretical false-alarm rates are not much different for various  $m$  and  $n$  as seen in Table I.

In summary, the OAAT method offers a better alternative to the discard-all procedure for practical use. It is as powerful as the discard-all procedure, but can diminish the false-alarm rate dramatically.

#### 4.2. Controlling the overall false-alarm rate

More and more researchers/practitioners consider controlling the overall false-alarm rate in Phase I analysis. In this subsection, we use the Bonferroni method to control the overall false-alarm rate  $\alpha$ . Let  $k = \sqrt{(m-1)/m} c_{4,m} t_{m(n-1), \alpha/2m}$  where  $t_{\nu, \gamma}$  is the 100(1- $\gamma$ )th percentile of the  $t$  distribution with  $\nu$  degrees of freedom. Set  $\alpha=0.05$ . It is noted that  $k$  increases as  $m$  increases or as  $n$  decreases.

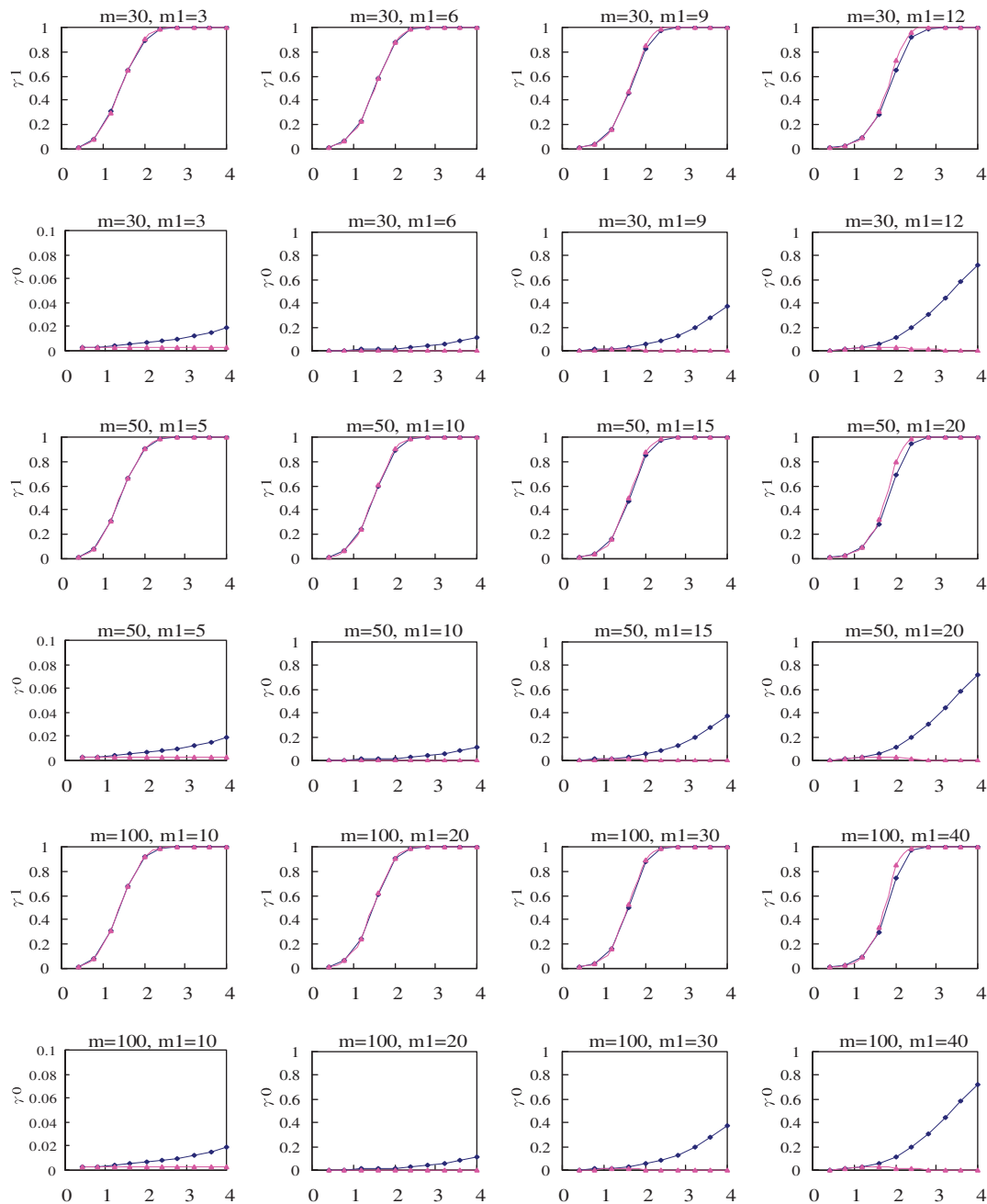
As above, consider  $m=30, 50, 100$  and  $n=5, 10, 15$ . For each combination of  $m$  and  $n$ , consider the cases of  $m_1=0, 0.1m, 0.2m, 0.3m, 0.4m$  and  $\delta=0.4$  (0.4) 4. Simulate 1000000 data sets for each scenario and calculate  $\bar{R}_0$  and  $\bar{R}_1$  for each scenario. The standard errors of  $\bar{R}_0$  and  $\bar{R}_1$  are about 0–0.00998.

When all the  $m$  samples are from the in-control process, similar to the results of controlling the individual false-alarm rate given in Section 4.1, the study indicates (not shown) that the  $\bar{R}_0$  of the OAAT procedure is smaller than that of the discard-all procedure in each scenario and the false-alarm reduction is small, less than 1.3%.

When some of the samples are from the out-of-control process, simulation results are similar to that in the last subsection. To save space, we only present the case of  $n=5$ . Figure 3 plots the simulated  $\bar{R}_1/m_1$  and  $\bar{R}_0/m_0$  values versus  $\delta$  for various values of  $m$  and  $m_1$ . For the effect of  $m$ , the following are observed:

- The detecting power ( $\bar{R}_1/m_1$ ) of both procedures decreases in general as  $m$  increases—this is the effect of multiplicity control, which means that when controlling a fixed overall false-alarm rate ( $\alpha$ ) for  $m$  samples, the individual false-alarm rate ( $\alpha^* = \alpha/m$ ) gets smaller as  $m$  increases. Thus the out-of-control region gets smaller, hence harder for charting points to fall into no matter they are from in-control or out-of-control processes. Therefore, both the detecting power and the false-alarm rate are reduced for either procedure.
- The detecting powers ( $\bar{R}_1/m_1$ ) of the two procedures are about the same. But, the OAAT procedure performs much better in terms of the false-alarm rate ( $\bar{R}_0/m_0$ ) than the discard-all procedure. However, it is noted that the advantage decreases as  $m$  increases. This is because, when  $m$  increases, fewer false alarms can occur for the Bonferroni method; but the OAAT procedure already has very few false alarms; thus, the difference between the two procedures becomes smaller. In other words, the effect of multiplicity control is more for the discard-all procedure than for the OAAT procedure.



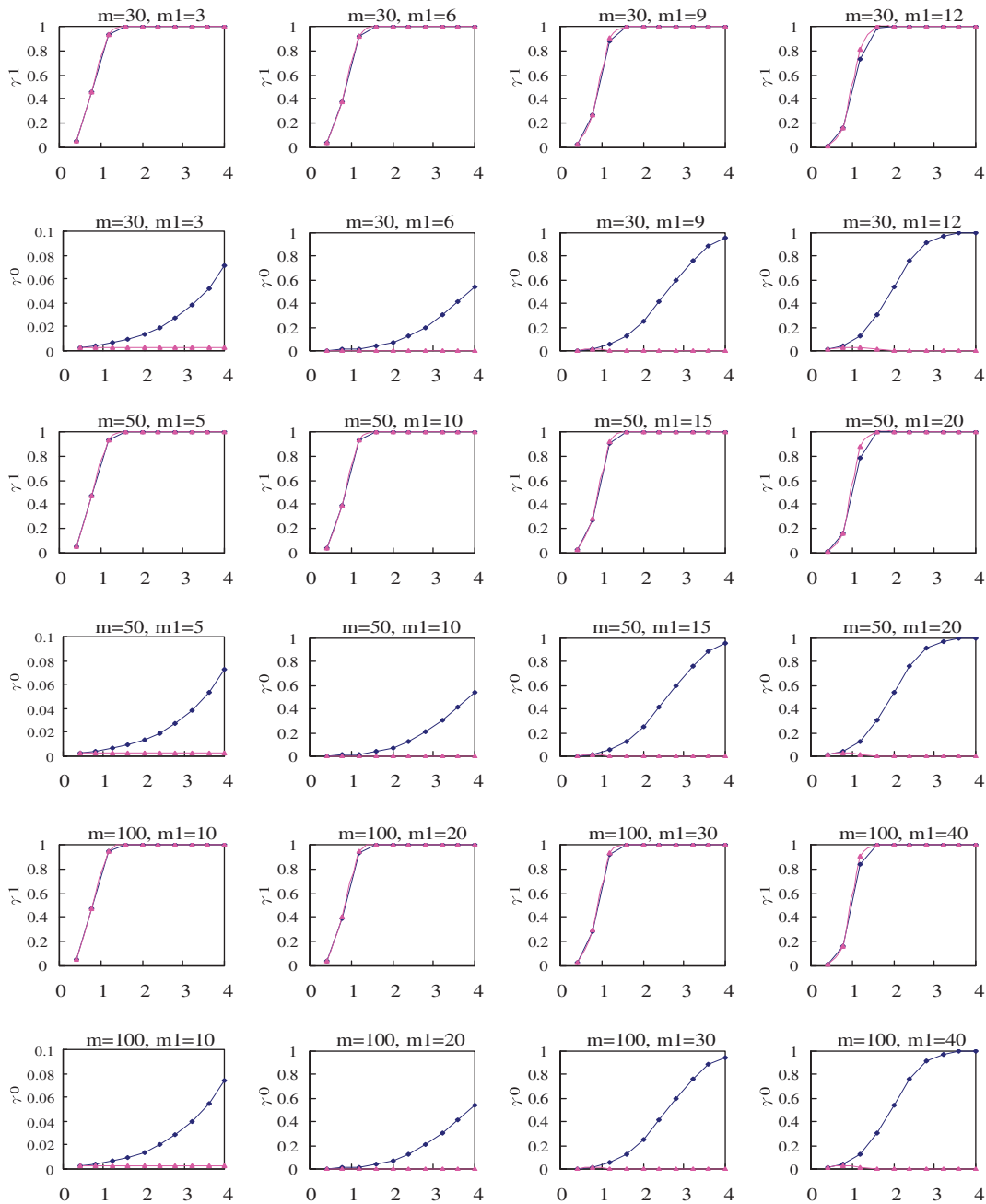


**Figure 1.**  $\bar{R}_1/m_1$  ( $\gamma_1$ ) and  $\bar{R}_0/m_0$  ( $\gamma_0$ ) for various combinations of  $m$  and  $m_1$  ( $n=5$ ) when  $k=3$ . The x-axis is the shift size  $\delta$ . The solid line with diamonds corresponds to the discard-all procedure and the dashed line with triangles corresponds to the OAAT procedure. This figure is available in colour online at [www.interscience.wiley.com/journal/qre](http://www.interscience.wiley.com/journal/qre)

## 5. Summary and concluding remarks

In this paper, we study some strategies of Phase I analysis in control charting. It is found that the conventional practice of discarding all 'out-of-control' samples at each iteration in constructing control charts has a major drawback of throwing away too many in-control samples. To overcome this drawback, we propose and study a new OAAT procedure that only discards the most extreme sample at a time. Our simulation study demonstrates that the OAAT procedure reduces dramatically the occurrences of false alarms and this advantage is more profound when the process is more unstable (i.e. more out-of-control samples or larger process deviation).

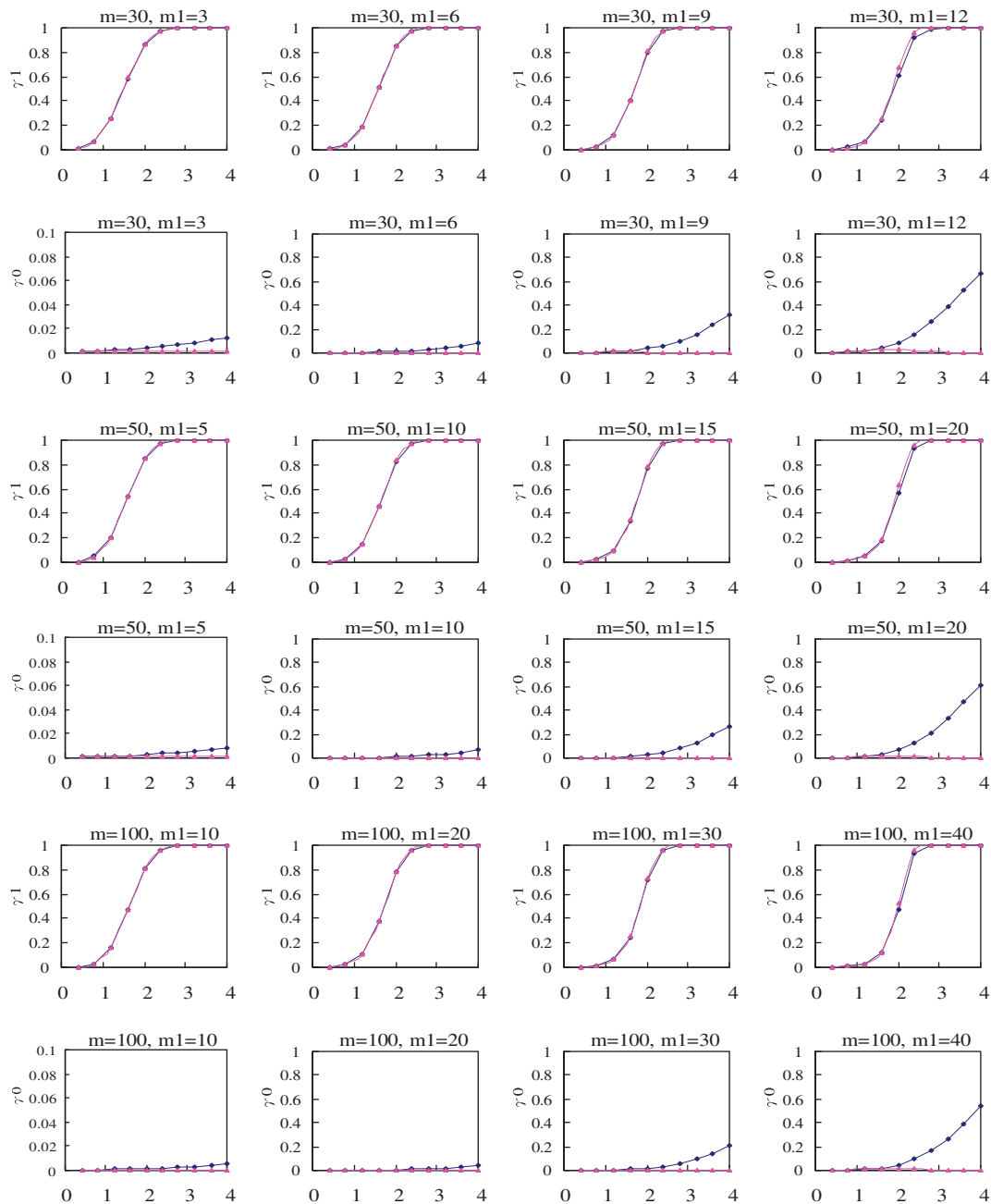
We also suggest a new strategy on when to inspect the process to look for assignable causes for samples signaling out-of-control alarms. Instead of carrying out investigations whenever an alarm signals, the new strategy is: run through the whole iterative procedure and then perform investigations for all 'out-of-control' samples after the remaining samples are all in control. This practice may save a tremendous amount of cost and time in bringing the process to the state of statistical control.



**Figure 2.**  $\bar{R}_1/m_1$  ( $\gamma_1$ ) and  $\bar{R}_0/m_0$  ( $\gamma_0$ ) for various combinations of  $m$  and  $m_1$  ( $n=15$ ). The x-axis is the shift size  $\tilde{\delta}$ . The solid line with diamonds corresponds to the discard-all procedure and the dashed line with triangles corresponds to the OAAAT procedure. This figure is available in colour online at [www.interscience.wiley.com/journal/qre](http://www.interscience.wiley.com/journal/qre)

We study two approaches of error control: controlling the individual false-alarm rate and the overall false-alarm rate. Note that in conventional practice, the individual false-alarm rate is kept fixed for each sample tested in all iterations; but, when using Bonferroni's adjustment to control the overall false-alarm rate, it becomes larger as more 'out-of-control' samples are removed as the screening process progresses.

We also study two criteria for evaluating the performances of different control schemes, including the signal probability ( $P(R \geq 1)$ ) considered in the literature and both numbers of false and true alarms,  $E(R_0)$  and  $E(R_1)$  (or  $E(R_0/m_0)$  and  $E(R_1/m_1)$ ), suggested in this paper. The signal probability considers 'the process' as a whole. It only assesses the ability of judging if a process is in control or not. It does not assess the scheme on its performance of screening out out-of-control samples and/or keeping in-control points. With a mixture of in-control and out-of-control samples in the data set, perhaps it is more appropriate to use simultaneously two indicators—one to assess the false-alarm rate and the other to assess the detection power, such as  $E(R_0)$  and  $E(R_1)$  (or  $E(R_0/m_0)$  and  $E(R_1/m_1)$ ).



**Figure 3.**  $\bar{R}_1/m_1$  ( $\gamma_1$ ) and  $\bar{R}_0/m_0$  ( $\gamma_0$ ) for various combinations of  $m$  and  $m_1$  ( $n=5$ ) when the overall false-alarm-rate  $\alpha$  is controlled at 0.05 by Bonferroni's adjustment. The x-axis is the shift size  $\delta$ . The solid line with diamonds corresponds to the discard-all procedure and the dashed line with triangles corresponds to the OoAT procedure. This figure is available in colour online at [www.interscience.wiley.com/journal/qre](http://www.interscience.wiley.com/journal/qre)

Finally, we remark that the  $\bar{X}$  chart used in this paper is just for demonstration of the above ideas. The new strategy for Phase I analysis, evaluation criteria for control schemes, and the new OoAT procedure can be applied to any other control charts.

## Acknowledgements

We would like to express our gratitude to the Editor and an anonymous referee for careful reading and insightful views and comments. This research is partially supported by the National Research Council of Taiwan, Grant Nos. NSC93-2118-M009-007 and NSC95-2118-M-009-006-MY2.

## References

1. Woodall WH. Controversies and contradictions in statistical process control (with Discussions). *Journal of Quality Technology* 2000; **32**(4):341–378.
2. Woodall WH, Montgomery DC. Research issues and ideas in statistical process control. *Journal of Quality Technology* 1999; **31**(4):377–386.
3. Jones LA, Champ CW. The design and performance of phase I control charts for times between events. *Quality and Reliability Engineering International* 2002; **18**:479–488.
4. Sullivan JH, Woodall WH. A control chart for preliminary analysis of individual observations. *Journal of Quality Technology* 1996; **28**:265–278.
5. Hunter JS. The Box–Jenkins manual adjustment chart. *Quality Progress* 1998; **31**:129–137.
6. Vermaat MB, Ion RA, Does RJMM, Klaassen CAJ. A comparison of Shewhart individuals control charts based on normal, nonparametric, and extreme-value theory. *Quality and Reliability Engineering International* 2003; **19**:337–353.
7. Borror CM, Champ CW. Phase I control charts for independent Bernoulli data. *Quality and Reliability Engineering International* 2001; **17**:391–396.
8. Champ CW, Chou S-P. Comparison of standard and individual limits phase I Shewhart  $\bar{X}$ ,  $R$ , and  $S$  charts. *Quality and Reliability Engineering International* 2003; **19**:161–170.
9. Mahmoud MA, Woodall WH. Phase I analysis of linear profiles with calibration applications. *Technometrics* 2004; **46**:380–391.
10. Nedumaran G, Pignatiello JJ. On constructing  $T^2$  control charts for retrospective examination. *Communications in Statistics—Simulation and Computations* 2000; **29**:621–632.
11. Nedumaran G, Pignatiello JJ. On constructing retrospective  $\bar{X}$  control chart limits. *Quality and Reliability Engineering International* 2005; **21**:81–89.
12. Champ CW, Jones LA. Designing phase I  $\bar{X}$  charts with small sample sizes. *Quality and Reliability Engineering International* 2004; **20**:497–510.
13. Montgomery DC. *Statistical Quality Control: A Modern Introduction* (6th edn). Wiley: New York, 2009.

### Authors' biographies

**Jyh-Jen Horng Shiau** is a Professor in the Institute of Statistics at the National Chiao Tung University, Taiwan, where she has been a faculty member since 1992. She holds a BS in Mathematics from the National Taiwan University, Taipei, Taiwan, an MS in Applied Mathematics from the University of Maryland Baltimore County, an MS in Computer Science and a PhD in Statistics from the University of Wisconsin-Madison. Formerly, she taught at Southern Methodist University, the University of Missouri at Columbia, and the National Tsing Hua University and worked for the Engineering Research Center of AT&T Bell Labs before moving to Hsinchu, Taiwan. She is a former managing editor of an *International Journal Quality Technology & Quantitative Management* (2004–2006). Her primary research interests include industrial statistics, nonparametric and semiparametric regression, and functional data analysis. She is a lifetime member of the International Chinese Statistical Association.

**Jian-Huang Sun** received his MS degree in Statistics from the National Chiao Tung University under the supervision of the first author. His primary research interests include statistical quality control and mathematics education.