Research Article

Shewhart-Type Charts for Masked Data: A Strategy for Handling the Privacy Issue

Said Farooq Shah,1 Zawar Hussain,2 Muhammad Riaz,3 and Salman Arif Cheema4

1Department of Statistics, University of Peshawar, Peshawar, Pakistan
2Department of Social & Allied Sciences, CUVS, Bahawalpur, Pakistan
3Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
4Department of Applied Sciences, National Textile University, Faisalabad, Pakistan

Correspondence should be addressed to Said Farooq Shah; farooqshah@uop.edu.pk

Received 28 June 2020; Revised 20 August 2020; Accepted 14 September 2020; Published 26 September 2020

Data privacy is a serious issue and therefore needs our attention. In this study, we propose masking through randomized response techniques (RRTs) to ensure the privacy and thus to avoid falsification. We assume that the process characteristic is of sensitive nature, and due to privacy issue, the actual measurements cannot be shared with the monitoring team. In such situations, the producer is very likely to falsify the measurements. Consequently, the usual control charting techniques will mislead about the process status. We discuss different data masking strategies to be used with Shewhart-type control charts. The usual Shewhart-type control chart appears to be a subchart of the proposed charts. Average run length (ARL) is used as a performance measure of the study proposals. We have evaluated the performance of the proposed charts for different shift sizes and under different intensities of masking. We have also carried out a comparative analysis for various models under varying sensitivity parameters. We have also compared the performance of the proposals with the traditional Shewhart chart. It is observed that the B-L control chart under the RRT model performs better for smaller shifts and for larger shift sizes, the G-B chart under an unrelated question model performs better. A real-life application of the study proposal is also considered where monitoring of thickness of paint on refrigerators is of interest.

1. Introduction

Control charting technique deals with the early detection of special cause variation in the process. The idea was proposed by Shewhart for monitoring production process and is also widely applied in other fields such as health care, social sciences, and business decision-making. For example, Muhammad et al. [1] explored various phenomena, such as number of surgeries and stay time in hospital, and interpreted various control charts for continuous and attribute data. Sherlaw-Johnson and Beardsley [2] used k-sigma and variable control charts for detection of changes in quarterly observed emergency admission of individuals, of specific age group, into hospital for different diseases. In their work, they highlighted need of frequent and continuous data collection on different indicators via patient experience surveys. In other health settings, SPC has been used for monitoring mortality [3, 4], the frequency of health care acquired infection [5] and [6, 7], and a number of process measures including readmission rates, lengths of stay, or bed days [8, 9].

On the extension side of control charting techniques, motivated by the Shewhart (1939, p. 134) statement, “the only way one can experience any quality characteristic quantitatively is by means of an operation of measurement,” Linna and Woodall [10] addressed the issue of measurement error in a production process and quantified its impact on power of the Shewhart control chart for monitoring the characteristic under study. In social and health sciences, these measurements, for monitoring
purposes, may be collected directly from the persons in the population or extracted from the record. In both the situations, there are variables on which one may need data, but due to privacy issue, it becomes hard to get accurate data. For example, one may be interested in the monitoring of abortion rate, average amount of under reporting in income tax report, number of unsuccessful surgeries, and number of times illicit drugs were taken, but observing data directly on these issues will lead to misleading results, see, for example, Horvitz et al. [11], Eichhorn and Hayre [12], and Bar-Lev et al. [13]. Randomized response techniques (RRTs), pioneered by Warner [14], provide tools through which one can collect trustworthy data for drawing inferences about true quantities with high trust but low precision.

Nowadays, data privacy is the major issue for the manufacturing industries and business-related organizations. In both fields, control charting techniques are widely suggested for improving the quality and enhancing the business decision-making, respectively. For procuring the data, one possible way is to mask the actual measurement of a process before storing it or at least when releasing the information for monitoring purposes. The simple and probabilistic way of masking the characteristic is to do it through randomized response models. As in quality control charting techniques, the nature of the data is either qualitative or quantitative. The randomized response models provide masking mechanism both for qualitative and quantitative data. It is important to note that individual randomized observation cannot be used to know the true observation but the actual average behavior is estimable.

In the current electronically connected era, the quality control, in manufacturing process, through outside professionals is not uncommon; the exact information sharing with other organizations may pose a serious threat to data security, in situations where the involvement of outside inspectors is unavoidable and sharing the exact data is threat to the privacy. The RR model is highly capable of addressing this issue; that is, one has to just mask the data through the RR model before sharing it with the quality control department, and, in this paper, we show that the quality control department is still capable of monitoring the process parameters without knowing the actual measurements of the process characteristic.

The aim of this paper is to address the issue of false reporting and data privacy in control charting techniques with the help of incorporating randomized responses. Background of the study is given in Section 2, and hence, an introduction to RRT and statistical process control is provided. In Section 3, data masking strategies and resultant Shewhart-type control charts are proposed for monitoring the process mean. Section 4 gives performance evaluations for the proposed charts. Section 5 is dedicated to comparative analysis among the proposed charts along with the existing ones. Section 6 presents an application of the study proposals. The study is concluded, and some future research aspects are listed in Section 7.

2. Background

This study is based on the two prominent statistical techniques, namely, randomized response technique and statistical process control chart. So, a brief review of these two is given in the following sections.

2.1. Randomized Response Technique. About a half century ago, Warner [14] proposed the randomized response method as a survey technique to reduce potential bias due to nonresponse and social desirability when asking questions about sensitive behaviors and beliefs. The method asks respondents to use a randomization device, such as a coin flip, whose outcome is unobserved by the interviewer. Depending on the particular design, the randomization device determines which question the respondent has to answer [14–17], the type of expression the respondent uses to answer the sensitive question [18], or if the respondent should give a predetermined response [19, 20]. By introducing random noise, the randomized response method conceals individual responses and protects respondent privacy. As a result, respondents may be more inclined to answer truthfully.

Randomized response technique is basically a data-generating technique and is extended in various ways. Depending on the type of the parameter, we want to estimate through data; RRTs are divided into two main types, namely, qualitative RRTs and quantitative RRTs. In qualitative RRTs, estimation of proportions remained basic interest, and when it comes to estimation of population mean or similar, then quantitative RRTs are used. For recent advancements in RRTs, we refer to Shah et al. [21] and Blair et al. [22]. On the other hand, RRTs have been effectively merged with other tools and new research areas have been proposed. For example, Du and Zhan [23] merged RRT for privacy preserving in data mining, and Shah et al. [24] successfully merged RRT and paired comparison modeling to ensure privacy of the judges in paired comparison experiment and hence reduce the impact of social desirability bias.

There are various other fields with whom RRT may be combined to produce results less affected by false reporting, and we think that statistical control charting is one of them. Thus, in the coming subsection, we explain briefly this with a brief review.

2.2. Statistical Control Charts. In statistical process control (SPC) settings, the variations in measurements are categorized into two types: the special cause variations and common cause variations, the latter one can never be eliminated fully, and that is why these are acceptable. The first one is considered the bad one, and if present, the process is categorized as out-of-control. Furthermore, these variations are attributed to shift in the process parameters. Early detection of a shift is encouraged to insure the quality. The control chart (CC) is one of the statistical tools used for the early detection of the shift. The idea was developed by
Shewhart [25] about a century ago in America for monitoring manufacturing process. Currently, there are worthy advancements in these techniques. For example, exponentially moving average (EWMA) and cumulative sum (CUSUM) control charts are proposed while considering the importance of magnitude of the shift, and p-chart, u-chart, s-chart and many more are proposed by considering different quality characteristics of interest. Furthermore, a lot of the literature has been dedicated to adjust the limits of the Shewhart chart for different environments. For example, Raji et al. [26] evaluate the impact of outliers on the Shewhart chart and evaluate the applicability through data from the semiconductor-manufacturing industry; and Celano and Chakraborti [27] proposed a distribution-free Shewhart control chart for monitoring finite horizon productions.

Now, we turn to discuss the basic and simple Shewhart control chart for monitoring mean of the continuous random variable $X$, known as the $x$-bar control chart. The established ingredients of this chart are (1) an estimator of the process parameter, referred as plotting statistic, a function of the sample observation $(s)$, (2) a central line corresponding to the average value of the estimator, and (3) control limit, upper control limit (UCL), and lower control limit (LCL), usually set as $\pm 3$ standard error around the average value. For mathematical formulations, let us assume that a quality characteristic follows a distribution with mean $\mu_0$ and variance $\sigma^2$ and consider that $x$ is the plotting statistic, then the upper and lower $k$-sigma limits are, respectively, defined as

$$
UCL = \mu_0 + k\frac{\sigma}{\sqrt{n}} \\
LCL = \mu_0 - k\frac{\sigma}{\sqrt{n}}
$$

There are several rules for classifying the process out-of-control or in-control, called run rules, but simplest among them is if a value of the plotting statistic falls outside these control limits, then the process is said to be out-of-control; otherwise, there are no evidences of special cause variations. One of the performance measuring criteria of CC is ARL, defined as the average number of subsamples required for a CC to detect a shift in the process’s parameters. The common cause variations may lead CC to signal; in such situations, the calculated ARL is categorized as in-control ARL and is denoted by $ARL_0$. False alarm rate, the reciprocal of $ARL_0$, is the probability that the value of plotting statistic falls outside the control limits given that the process’ parameters are in-in-control state, that is, $(1/ARL_0) = \Pr[UCL < x < LCL | \mu_0 = \mu_0]$ where $\mu_0$ is said to be the value for which the underlying process is considered to be in-control state. In (1) and (2), $k$ is a positive constant, typically used equal to 3 for a process where the underlying quality characteristic’s distribution is assumed to follow normal distribution. In such cases, $k$ equal to 3 yields $ARL_0$ equals to 370. When the plotting statistic does not follow normal distribution and then for performance measure, one may fix the false alarm rate at 0.0027 (or any other value) and the value of $k$ should be found. For more detail, we refer to Mehmoond et al. [28] and Hardy et al. [29].

### 3. Different Types of Masking and Proposed Charts

This section is devoted to construction of generalized control charting technique which takes a broad view that of Shewhart and result in a flexible way of monitoring the sensitive character. This mechanism exploits both the quantitative randomized response models and the techniques used for monitoring the process mean in statistical process control. We start by introducing a masking mechanism to the producer which they may choose to mask the data. Then, we provide control limits of Shewhart-type charts, and those are capable of catering the intensity of masking used by the producers. Lastly, we provide a theoretical comparison with the traditional Shewhart chart.

In the coming sections, let us denote the sensitive character of interest by a random variable $X$, let $S$ be a random variable used in masking the response, generated from an arbitrary distribution, and let $Z$ be the observed randomized response. In the RR studies, $S$ is also called the scrambling variable.

#### 3.1. Masking through Greenberg et al. [30] Model and G-B Chart

In the study of Greenberg et al. [30], they coined the idea of quantitative data collection in survey sampling on sensitive issues. We exploit the study of Greenberg et al. [30] and state that the producer/respondent reports the actual measurement $X$, with probability $P$, or the value $S$ with probability $(1-\text{	extit{P}})$. Thus, the masked data may be modeled as

$$Z_G = \begin{cases} 
X, & \text{with probability } p, \\
S, & \text{with probability } 1-p,
\end{cases}$$

where $p$ is defined in the same way as defined by Warner [14], known as the masking parameter and is selected by the producer and is known to the monitoring team. Furthermore, let us assume that $X$ follows a distribution with mean $E(X) = \mu_X$ and variance $\text{Var}(X) = \sigma^2_X$, and $S$ is independent of $X$ with mean $E(S) = \mu_S$ and variance $\text{Var}(S) = \sigma^2_S$. Now, it is nontrivial to show that mean and variance of $Z$ are

$$
\mu_{Z_G} = p\mu_X + (1-p)\mu_S, \\
\sigma^2_{Z_G} = p^2\sigma^2_X + (1-p)\sigma^2_S + p(1-p)(\mu_X - \mu_S)^2.
$$

Then, an unbiased estimate of $\mu_X$ proposed by Greenberg et al. [30] is

$$
\hat{\mu}_X = \frac{\overline{Z}_G - (1-p)\mu_S}{p}.
$$

And its variance is given as follows:

$$
V(\hat{\mu}_X) = \frac{\text{V}(\overline{Z}_G)}{p^2} = \frac{\sigma^2_Z}{np^2}.
$$
Using the above facts, we have a Shewhart-type control chart, for monitoring the mean of \( X \), with lower confidence limit and upper confidence limit given, respectively, as

\[
\text{LCL}_1 = \mu_X - k \frac{p}{n} \left( \frac{2}{n} \sigma_X^2 + (1-p)\sigma_S^2 + p(1-p)(\mu_X - \mu_s)^2 \right), \tag{7}
\]

\[
\text{UCL}_1 = \mu_X + k \frac{p}{n} \left( \frac{2}{n} \sigma_X^2 + (1-p)\sigma_S^2 + p(1-p)(\mu_X - \mu_s)^2 \right), \tag{8}
\]

If we look at model (3) and limits (7) and (8), then it is nontrivial to notice that for \( p = 1 \), we have the unmasked data, and this chart reduces to the usual Shewhart-type chart limits are given in (1) and (2). In coming sections, we refer this chart as the G-B chart.

3.2. Masking through Eichhorn and Hayre [12] and Bar-Lev et al. [13] Models and the B-L Chart. Eichhorn and Hayre [12] proposed a new randomized response technique for estimating the population mean response of the quantitative sensitive character. Each measurement is recorded in masked form. That is, the producer is directed to give a coded or scrambled observations obtained by multiplying the true measurement value by some random numbers unknown by the monitoring authorities. The monitoring team does not know which random number the producer has used for coding, but the distribution by which the random number has been generated is known.

Let \( X \) is the variable representing the responses to the sensitive quantitative character, \( S \) is the variable representing the random number used for the purpose of coding the responses, and \( Z_E = XS \) represents the coded response, where \( S \) and \( X \) are being independent of each other. Then,

\[
\mu_{Z_E} = \mu_S \mu_X, \quad V(Z_E) = \sigma_X^2 \sigma_S^2 + \mu_X^2 \left( 1 + C_X^2 \right) \sigma_S^2, \tag{9}
\]

where \( C_X^2 \) is the coefficient of variation of \( X \).

Thus, the unbiased estimator of \( \mu_X \), proposed by Eichhorn and Hyre [12], is given as

\[
\bar{\mu}_X = \frac{z_E}{\mu_S}, \tag{10}
\]

and its variances is

\[
V(\bar{\mu}_X) = \frac{\sigma_X^2 + \mu_X^2 \left( 1 + C_X^2 \right)}{n}. \tag{11}
\]

Let us denote the sensitive character of interest by a random variable \( X \), let \( S \) is a random variable used in masking the response, generated from an arbitrary distribution, and \( Z \) is the observed randomized response. For simplicity and without any loss, we assume independence between \( X \) and \( S \). According to Bar-Lev et al. [13], \( Z \) is distributed as

\[
Z = \begin{cases} 
X, & p, \\
XS, & 1 - p,
\end{cases} \tag{12}
\]

where \( p \) is the same role defined by Warner [14], known as the masking parameter and is presented by the investigator. Furthermore, let us assume that \( X \) follows a distribution with mean \( E(X) = \mu_X \) and variance \( \text{Var}(X) = \sigma_X^2 \) and \( S \) is independent of \( X \) with mean \( E(S) = \mu_s \) and variance \( \text{Var}(S) = \sigma_S^2 \). Now, it is nontrivial to show that mean and variance of \( Z \) are

\[
E(Z) = \mu_X \left( p + \mu_s (1 - p) \right),
\]

\[
\text{Var}(Z) = \mu_X^2 \left( 1 + C_X^2 \right) \left( p + \mu_s^2 (1 + C_s^2) (1 - p) \right).
\]

Using the above information, Bar-Lev et al. [13] proposed an unbiased estimator of true mean \( \mu_X \) as

\[
\bar{\mu} = \frac{\bar{Z}}{p + \mu_s (1 - p)}, \tag{14}
\]

where \( \bar{Z} \) is the sample mean of randomized data. The standard error of the estimator given in (14) is

\[
\tau = \frac{1}{\sqrt{n}} \left[ \sigma_X^2 + \mu_X^2 \left( 1 + C_X^2 \right) \left( p + \mu_s^2 (1 + C_s^2) (1 - p) \right) \right]^{1/2}. \tag{15}
\]

Using equations (14) and (15), the \( k \)-sigma lower control limit (LCL), central line (CL), and upper control limit (UCL) of Shewhart control chart, for monitoring the mean of sensitive process \( X \), are defined, respectively, as

\[
\text{LCL}_2 = \mu_X - k \tau, \\
\text{CL}_2 = \mu_X, \\
\text{UCL}_2 = \mu_X + k \tau.
\]

This model is the general form of direct responses and the model proposed by Eichhorn and Hayre [12]; that is, for \( p = 1 \), this model reduces to the direct response mechanism, and for \( p = 0 \), this becomes the multiplicative randomized response model proposed by Eichhorn and Hayre. For simplicity, we denote this proposed chart by B-L (Bar-Lev), in the coming sections.

3.3. Masking through Additive Models and the G-A Chart. For gain in efficiency of the estimates, several techniques have been considered. Such improvements cost the simplicity of the model, but additive randomized response models are one of those which are simple in application and better in terms of the variances of the estimates. Gupta et al. [31] proposed an optional randomized additive response model, which, with gain in efficiency, is also capable of producing the information for estimation of the sensitivity level. Here, we simplify the Gupta et al. [31] model as

\[
Z = \begin{cases} 
X, & p, \\
X + S, & 1 - p.
\end{cases} \tag{17}
\]
In this simplified version of the Gupta et al. [31] model, if \( p = 0 \), then we get the model proposed by Pollock and Beck [32]. The unbiased estimator of the sensitive mean may be reported as
\[
\hat{\mu} = \bar{z} - p + \mu_x(1 - p).
\]
It is trivial to show that the variance of \( \hat{\mu} \) is
\[
V(\hat{\mu}_x) = \frac{\sigma^2 + (1 - p)(\sigma^2_s + \rho S^2)}{n}.
\]

By setting \( p = 0 \), in (17), (18), and (19), one may obtain the model, estimator, and its variance, proposed by Pollock and Beck [32], respectively.

The upper control limit (UCL\(_3\)) and lower control limit (LCL\(_3\)) for this model are, respectively, given as
\[
\begin{align*}
\text{UCL}_3 &= \mu_x + k\sqrt{\frac{1}{n}(\sigma^2_{\bar{x}} + (1 - p)(\sigma^2_s + \rho S^2))}, \\
\text{LCL}_3 &= \mu_x - k\sqrt{\frac{1}{n}(\sigma^2_{\bar{x}} + (1 - p)(\sigma^2_s + \rho S^2))}.
\end{align*}
\]

Furthermore, this chart is named as the G-A chart, which means the Gupta-additive chart, and by putting \( p = 1 \), the G-A chart reduces to the usual Shewhart chart.

4. Performance Evaluations

The performance of the proposed chart is evaluated through the traditional Monte Carlo simulations for various values of the masking parameter \( p \). We assumed that \( X \) follows a normal distribution with mean 3 and variance 1. The distribution of the masking variable \( S \) is assumed to be Poisson with mean \( \mu_s = 3 \). First, we searched for the values of \( k \) for different values of \( p \) by fixing the ARL\(_0\) approximately equal to 370, and then we introduced shift of various magnitude and calculated the out-of-control average run length (ARL) and standard deviation of run length (SDRL) for evaluation of power of detecting shift, in \( X \), of the proposed chart. The shift size in the mean \( \delta \) of \( X \) is defined as
\[
\delta = (\bar{X}_n - \bar{X}_1) / \bar{X}_1,
\]
where \( \bar{X}_1 \) is the shifted mean of \( X \). The results are reported in Table 1–3 for different settings. We compared the performance with the traditional Shewhart chart.

4.1. The G-B Chart

According to the results provided in Table 2, this chart suffers with biased ARL issue for smaller shift sizes. Noted from (5), the plotting statistic does not exist for \( p = 0 \), which implies that fully unrelated data (when \( Z = S \)) are not usable for monitoring the process mean. Furthermore, for larger shifts (greater than or equal to 1), the performance of the chart improves with increase in the \( p \) and converges to the usual Shewhart chart when \( p \) becomes 1.

4.2. The B-L Chart

As discussed in Section 3.2 that the B-L model is the generalized form of the Eichorn and Hayre [12] model, we consider the B-L model for evaluation. The ARL values reported in Table 2 show the usual behavior; that is, as the shift size increases, then we expect earlier detection. It is also observed that there is no linear relationship between \( p \) and the performance of the control chart but the chart performs better at the extreme values of \( p \); that is, when \( p = 0 \) or \( p = 1 \), then ARL decreases early as the shift size increases. It is worth mentioning that at \( p = 0 \), we have fully masked data; that is, privacy is fully ensured; when \( p = 1 \), then we have fully unmasked data, and the proposed control chart reduces to the usual Shewhart chart; generality of the proposed chart should be witnessed. Furthermore, when \( p \) decreases from 0.9 to 0.1 (with a step size equal to 0.1), then the detection power of the chart increases. Furthermore, in the given setting, it is observed that for some combinations of \( p \) and shift size, the proposed chart is ARL biased, and these situations are highlighted in Table 2.

4.3. The G-A Chart

The control limits are provided in Table 1 for different values of the masking parameter, \( p \), by fixing the ARL\(_0\) equal to 370. The ARL bias issue is once again encountered for small shift sizes. The relationship between extent of masking and average run length is not linear; that is, there is an upward trend in the ARL for smaller shift sizes but gradually a downward trend for larger shift sizes (greater than or equal to 1). Similarly, this chart is the generalization of the simple Shewhart and Pollock and Beck [32] charts.

5. Comparative Analysis

This section provides a comparison among the proposed charts of this study along with the traditional Shewhart chart.

5.1. Comparisons among the Proposed Charts

In this section, graphical comparison of the proposed G-B, G-A, and B-L charts is provided for different shift sizes. Figure 1 depicts the relationship between \( p \), \( \delta \), and ARL of the proposed charts. It is observed that the B-L chart outperformed G-B and G-A charts when \( p \) is less than or equal to 0.7 for different shifts. As the shift size increases, the performance of G-A chart increases, comparative to the G-B chart, for smaller values of \( p \). By fixing \( p \) greater than 0.7, the G-B chart performs better than the other two charts, and when \( p = 1 \), then all of the charts converge to a simple Shewhart chart.

5.2. Proposed versus Shewhart Chart

As we witnessed, the proposed charts are the generalized form of the traditional Shewhart chart. It is worth mentioning that masking has two facts (i) it insures privacy and (ii) it introduces extra noise in the variances of the estimates. Due to the last fact, one may say that the traditional chart, provided in (1) and (2), has narrower limits than the proposed one and hence has more power to detect shift. However, we do not buy this argument and argue that the traditional Shewhart chart is incapable to detect falsified responses which we witnessed in Table 2. To further elaborate the applicability of the proposed work, we
consider a situation where the process is assumed to be of sensitive nature and the monitoring team does not know that the measurements are falsified. Due to sensitivity, the producer falsified the responses as $Z = X\delta$, where $X$ is the actual value and $S$ is the scrambling variable defined earlier. Due to the lack of prior information, the monitoring team has to use the traditional Shewhart-type control chart (with limits provided in (1) and (2)) for monitoring purposes. If so, then one may witness the deteriorating nature of this phenomenon on the Shewhart chart in the last rows of Table 2; without any shift in the process parameters, the chart declared the process out-of-control. On the other side, the proposed technique is capable to reduce the impact of falsified responses and capable to detect the shift in the process parameter.

### Table 2: Average RL for various combinations of shift size and different values of $p$.

| $p$ | Model      | Shift size $\delta$ |
|-----|------------|----------------------|
|     |            | 0.00 | 0.15 | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 |
| 0.0 | B-L        | 369.8204 | 342.9505 | 324.4787 | 259.5459 | 195.4470 | 145.3163 | 78.6172 | 45.2577 |
| 0.1 | G-B        | 370.3704 | 398.3633 | 384.5378 | 331.1972 | 266.5245 | 195.9070 | 107.5266 | 62.0638 |
| 0.2 | G-A        | 371.3979 | 355.6773 | 333.2773 | 266.6183 | 202.8503 | 155.3758 | 85.9351 | 49.3702 |
| 0.3 | G-B        | 372.5152 | 351.1763 | 333.5666 | 271.5980 | 214.9162 | 160.5205 | 93.1897 | 53.4576 |
| 0.4 | G-A        | 373.6495 | 356.1011 | 350.4887 | 293.7171 | 237.4126 | 182.4637 | 101.7029 | 67.346 |
| 0.5 | G-B        | 375.7280 | 378.5445 | 380.1446 | 386.9596 | 375.3932 | 364.0923 | 325.8341 | 282.0948 |
| 0.6 | G-A        | 370.9040 | 368.6180 | 355.5431 | 303.6942 | 254.1390 | 194.8039 | 117.1545 | 72.4191 |
| 0.7 | G-B        | 374.0855 | 373.9863 | 369.7926 | 322.6122 | 265.3373 | 207.8775 | 130.0364 | 81.5487 |
| 0.8 | G-A        | 373.4111 | 382.4115 | 383.5503 | 344.4638 | 278.3104 | 225.1947 | 142.9661 | 91.5245 |
| 0.9 | G-B        | 373.0704 | 378.3108 | 377.6799 | 348.8476 | 294.1774 | 238.7937 | 141.83 | 75.3392 |
| 1.0 | G-A        | 370.4920 | 389.5094 | 393.5571 | 353.9593 | 290.8406 | 239.8486 | 152.9354 | 101.600 |

6. An Application

In this section, we consider a data set from Wild and Seber [33] and we randomize these individual observations using the model proposed by Bar-Lev et al. [13] (in fact, any other model may be considered). The masked and unmasked data...
Table 3: SD of RL for various combinations of shift size and different values of $p$.

| $p$ | Model | Shift size $\delta$ | 0.00 | 0.15 | 0.25 | 0.50 | 0.75 | 1.00 | 1.50 | 2.00 |
|-----|-------|---------------------|------|------|------|------|------|------|------|------|
| 0.0 | B-L   | 372.09              | 339.742 | 321.813 | 258.143 | 197.926 | 142.765 | 78.6001 | 44.1666 |
|     | G-B   | 471.5541            | 476.5019 | 474.7691 | 467.3532 | 462.7542 | 453.5529 | 427.5192 |
|     | G-A   | 393.87              | 375.5852 | 380.7511 | 336.6681 | 283.7831 | 239.263 | 185.4578 |
| 0.1 | B-L   | 372.24              | 352.019 | 330.247 | 267.846 | 204.118 | 154.636 | 86.0165 | 49.4626 |
|     | G-B   | 381.0626            | 385.9171 | 386.5853 | 373.0959 | 328.6268 | 266.7557 | 344.6503 |
|     | G-A   | 369.87              | 373.2505 | 368.2201 | 328.723 | 269.6875 | 211.3197 | 117.1727 |
| 0.2 | B-L   | 375.32              | 348.609 | 328.835 | 269.888 | 214.983 | 162.289 | 93.4574 | 53.0225 |
|     | G-B   | 381.0626            | 385.9171 | 386.5853 | 373.0959 | 328.6268 | 266.7557 | 344.6503 |
|     | G-A   | 369.87              | 373.2505 | 368.2201 | 328.723 | 269.6875 | 211.3197 | 117.1727 |
| 0.3 | B-L   | 360.3               | 352.365 | 342.078 | 279.144 | 224.088 | 169.749 | 100.512 | 59.309 |
|     | G-B   | 369.87              | 375.5852 | 375.5363 | 363.3669 | 338.9533 | 318.8717 | 276.1298 |
|     | G-A   | 369.87              | 373.6151 | 399.827 | 341.0555 | 281.6622 | 211.3972 | 126.568 |
| 0.4 | B-L   | 370.36              | 361.48 | 349.226 | 307.06 | 250.373 | 195.474 | 119.042 | 76.6496 |
|     | G-B   | 375.2276            | 378.9434 | 378.3174 | 373.0098 | 363.7104 | 337.0251 | 276.1298 |
|     | G-A   | 369.87              | 373.6151 | 399.827 | 341.0555 | 281.6622 | 211.3972 | 126.568 |
| 0.5 | B-L   | 379.83              | 368.302 | 372.48 | 321.424 | 264.616 | 206.768 | 128.65 | 81.3299 |
|     | G-B   | 375.2276            | 378.9434 | 378.3174 | 373.0098 | 363.7104 | 337.0251 | 276.1298 |
|     | G-A   | 369.87              | 373.6151 | 399.827 | 341.0555 | 281.6622 | 211.3972 | 126.568 |
| 0.6 | B-L   | 372.62              | 361.48 | 349.226 | 307.06 | 250.373 | 195.474 | 119.042 | 76.6496 |
|     | G-B   | 379.83              | 375.5363 | 375.56 | 363.3669 | 338.9533 | 318.8717 | 276.1298 |
|     | G-A   | 369.87              | 380.0005 | 378.9083 | 335.1135 | 274.0873 | 213.9339 | 126.8937 |
| 0.7 | B-L   | 372.39              | 374.482 | 390.75 | 343.233 | 275.894 | 224.354 | 141.136 | 91.4099 |
|     | G-B   | 379.83              | 377.5398 | 375.3722 | 349.39 | 293.673 | 241.1474 | 139.4597 | 75.6884 |
|     | G-A   | 369.87              | 405.9803 | 432.2446 | 388.1164 | 320.4707 | 273.755 | 143.3654 | 85.9176 |
| 0.8 | B-L   | 366.338             | 381.036 | 390.31 | 354.569 | 284.41 | 236.836 | 153.799 | 100.455 |
|     | G-B   | 369.87              | 374.9014 | 371.9819 | 322.1747 | 245.773 | 190.582 | 93.30615 | 43.7664 |
|     | G-A   | 369.87              | 395.0287 | 394.5415 | 361.0619 | 286.6851 | 228.0623 | 126.2924 | 77.97229 |
| 0.9 | B-L   | 368.42              | 404.066 | 397.187 | 382.3 | 319.119 | 253.742 | 172.752 | 113.99 |
|     | G-B   | 369.87              | 370.268 | 357.3045 | 276.661 | 190.143 | 122.5217 | 47.72949 | 19.11604 |
|     | G-A   | 369.87              | 390.3073 | 410.1353 | 367.7251 | 280.2753 | 213.2582 | 120.1642 | 66.91047 |
| 1.0 | B-L   | 377.155             | 337.864 | 279.979 | 155.518 | 81.9042 | 43.8392 | 14.4668 | 5.97265 |
|     | G-B   | 369.87              | 331.2344 | 279.4253 | 160.1547 | 84.03973 | 42.47246 | 14.45013 | 5.613285 |
|     | G-A   | 369.87              | 338.2211 | 280.1716 | 155.7074 | 82.09768 | 43.85211 | 14.40269 | 5.878676 |
| $\infty$ | B-L | 0.1787 | 0.1577 | 0.16111 | 0.1355 | 0.1145 | 0.11332 | 0.0896 | 0.0705 |

![Figure 1: Continued.](a) (b) (c)
are given in Table 4, for different values of $p = 0, 0.1, 0.5, 0.8, 0.9, \text{and } 1.0$. The thickness of paint given in the first row of Table 1 is unmasked; that is, the measurements corresponding to $p = 1$ are the original measurements. We believe that due to sensitive situations, the producer may falsify these observations before presenting these to the

**Table 4: Masked and unmasked thickness of paint data on refrigerators for five refrigerators from twenty shifts.**

| SN | P | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.0 | 2.7 | 2.8 | 2.4 | 2.7 | 2.6 | 2.3 | 2.2 | 2.2 | 2.2 | 2.6 | 2.8 | 2.4 | 2.6 | 3.1 | 2.4 | 2.1 | 2.1 | 2.2 | 2.4 | 3.1 | 2.9 | 1.9 | 2.3 | 1.8 |
| 0.9 | 2.6 | 2.6 | 2.4 | 2.4 | 2.6 | 2.3 | 2.1 | 2.7 | 2.2 | 2.6 | 2.8 | 2.3 | 2.3 | 2.8 | 3.2 | 2.8 | 3 | 2.6 | 2.4 | 1.4 | 1.6 | 2.6 | 2.8 |
| 0.8 | 2.6 | 2.6 | 2.4 | 2.3 | 2.5 | 2.3 | 2.6 | 2.6 | 2.8 | 2.3 | 2.3 | 2.8 | 3.2 | 0 | 3 | 2.6 | 2.4 | 1.6 | 2.6 | 2.8 |
| 0.5 | 2.4 | 2.4 | 2.6 | 2.1 | 2.2 | 2.2 | 2.7 | 2.2 | 2.5 | 2.8 | 2.9 | 2.6 | 2.2 | 2.5 | 2.8 | 1.3 | 3.3 | 2.8 | 2 |
| 0.1 | 2.7 | 2.8 | 2.7 | 2.8 | 2.6 | 2.6 | 2.3 | 2.5 | 2.3 | 2.4 | 3.1 | 2.5 | 2.8 | 2.4 | 2.1 | 2.1 | 1.8 | 3.3 | 3.2 | 2.9 |

*Figure 1: Comparison of proposed charts for different shifts, $\delta$, and masking, $p$. (a) $\delta = 0.15$. (b) $\delta = 0.25$. (c) $\delta = 0.5$. (d) $\delta = 0.75$. (e) $\delta = 1.5$. (f) $\delta = 2.0$.)*
quality inspectors. Our procedure allows the producer to mask these measurements through the randomized response model discussed in the previous sections. By doing so, the inspector will not be able to trace the original measurements and hence the privacy of the measurements will be ensured. Here, we considered negative binomial distribution with mean 3 and variance 7.5 in masking the observations for different value of the masking parameter $P$. And it is evident from the values that the true measurements are masked; furthermore, the intensity of masking increases when the value of $P$ decreases, and hence, for $P = 0.0$, the data are fully masked; that is, each measurement is very likely to be different from the actual one. In Figure 2, we showed that after masking the observation, the monitoring team is still capable to declare the process out-of-control which in fact is out-of-control.

7. Conclusion and Recommendations

Data privacy is nowadays a serious issue, and we believe that if privacy is not protected, then it is very likely that the data owners (producers and patients) have no choice left with but to falsify the data. It has been established that falsified responses lead to misleading conclusions. To handle this situation, a win-win strategy is offered in this work. The producer is allowed to do the masking, through a probabilistic way, up to any extent, to insure the privacy, and to share the value of the masking parameter, $P$, with the monitoring team. The empirical study showed that as the extent of masking increases, then the difference between actual and masked observation increases, and at the highest extent, the chart is still capable to detect the actual (out-of-
control) status of the process. The proposed technique is established for the Shewhart-type control chart to monitor the process mean. Hence, there is a room for extending the proposed strategy to memory control charts and also applicable to attributes. On the masking mechanism side, there are a lot of advanced models but the basic models were considered to avoid the complications for the practitioners. Therefore, a comparative study with different RR models will be interesting.

**Data Availability**

The data used to support the findings of this study are included within the article.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**References**

[1] M. A. Mohammed, P. Worthington, and W. H. Woodall, "Plotting basic control charts: tutorial notes for healthcare practitioners," *Quality and Safety in Health Care*, vol. 17, no. 2, pp. 137–145, 2008.

[2] C. Sherlaw-Johnson and M. Barsdale, *Monitoring Change in Health Care Through Statistical Process Control Methods*, The Nuffield Trust, London UK, 2016.

[3] A. Bottle and P. Aylin, "Intelligent information: a national system for monitoring clinical performance," *Health Services Research*, vol. 43, no. 1p1, pp. 10–31, 2008.

[4] B. E. Rogers, L. M. Severe, and P. N. Eggers, "Method and apparatus for monitoring and communicating with an implanted medical device," U.S. Patent No. 6,957,107, 2005.

[5] J. Curran Celentano, J. D. Burke, and B. R. Hammond Jr, "In vivo assessment of retinal carotenoids: macular pigment detection techniques and their impact on monitoring pigment status," *The Journal of Nutrition*, vol. 132, no. 3, pp. 5355–539S, 2002.

[6] C. Sherlaw-Johnson, A. P. R. Wilson, B. Keogh, and S. Gallivan, "Monitoring the occurrence of wound infections after cardiac surgery," *Journal of Hospital Infection*, vol. 65, no. 4, pp. 307–313, 2007.

[7] O. A. Grigg, D. J. Spiegelhalter, and H. E. Jones, "Local and marginal control charts applied to methicillin resistant-Staphylococcus aureus bacteremia reports in UK acute National Health Service trusts," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 172, no. 1, pp. 49–66, 2009.

[8] M. Coory, S. Duckett, and K. Sketcher-Baker, "Using control charts to monitor quality of hospital care with administrative data," *International Journal for Quality in Health Care*, vol. 20, no. 1, pp. 31–39, 2007.

[9] I. R. Smith, B. Garlick, M. A. Gardner, R. D. Brighouse, K. A. Foster, and J. T. Rivers, "Use of graphical statistical process control tools to monitor and improve outcomes in cardiac surgery," *Heart, Lung and Circulation*, vol. 22, no. 2, pp. 92–99, 2013.

[10] K. W. Linna and W. H. Woodall, "Effect of measurement error on Shewhart control charts," *Journal of Quality Technology*, vol. 33, no. 2, pp. 213–222, 2001.

[11] D. G. Horvitz, B. G. Greenberg, and J. R. Abernathy, "Randomized response: a data-gathering device for sensitive questions," *International Statistical Review/Revue Internationale de Statistique*, vol. 44, no. 2, pp. 181–186, 1976.

[12] B. H. Eichhorn and L. S. Hayre, "Scrambled randomized response methods for obtaining sensitive quantitative data," *Journal of Statistical Planning and Inference*, vol. 7, no. 4, pp. 307–316, 1983.

[13] S. K. Bar-Lev, E. Bobovitch, and B. Boukai, "A note on randomized response models for quantitative data," *Metrika*, vol. 60, no. 3, pp. 255–260, 2004.

[14] S. L. Warner, "Randomized response: a survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[15] B. G. Greenberg, A.-L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz, "The unrelated question randomized response model: theoretical framework," *Journal of the American Statistical Association*, vol. 64, no. 326, pp. 520–539, 1969.

[16] N. S. Mangat and R. Singh, "An alternative randomized response procedure," *Biometrika*, vol. 77, no. 2, pp. 439–442, 1990.

[17] N. S. Mangat, "An improved randomized response strategy," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 56, no. 1, pp. 93–95, 1994.

[18] A. Y. C. Kuk, "Asking sensitive questions indirectly," *Biometrika*, vol. 77, no. 2, pp. 436–438, 1990.

[19] R. F. Boruch, "Assuring confidentiality of responses in social research: a note on strategies," *The American Sociologist*, vol. 6, pp. 308–311, 1971.

[20] J. A. Fox and P. E. Tracy, *Randomized Response: A Method for Sensitive Surveys*, SAGE, New York, NY, USA, 1986.

[21] S. F. Shah, S. A. Cheema, Z. Hussain, and E. A. Shah, "Masking data: a solution to social desirability bias in paired comparison experiments," *Communications in Statistics—Simulation and Computation*, pp. 1–19, 2020.

[22] G. Blair, K. Imai, and Y.-Y. Zhou, "Design and analysis of the randomized response technique," *Journal of the American Statistical Association*, vol. 110, no. 511, pp. 1304–1319, 2015.

[23] W. Du and Z. Zhan, "Using randomized response techniques for privacy-preserving data mining," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 505–510, Washington, DC, USA, August 2003.

[24] S. F. Shah, Z. Hussain, and S. A. Cheema, "Combining answers to direct and indirect questions: an implementation of Kuk’s randomized response model," *Communications in Statistics - Theory and Methods*, vol. 49, no. 16, pp. 3933–3949, 2020.

[25] W. A. Shewhart, "Application of statistical method in mass Production," in *Proceedings of the Industrial Statistics Conference*, Pitman Publishing Corporation, New York, NY, USA, 1939.

[26] I. A. Ragi, M. H. Lee, M. Riaz, M. A. R. Abujiya, and N. Abbasi, "Outliers detection models in shewhart control charts: an application in photolithography; a semiconductor manufacturing industry," *Mathematics*, vol. 8, no. 5, p. 857, 2020.

[27] G. Celano and S. Chakraborti, "A distribution-free Shewhart-type Mann-Whitney control chart for monitoring finite horizon productions," *International Journal of Production Research*, pp. 1–18, 2020.

[28] R. Mehmood, M. S. Qazi, and M. Riaz, "On the performance of X control chart for known and unknown parameters supplemented with runs rules under different probability distributions," *Journal of Statistical Computation and Simulation*, vol. 88, no. 4, pp. 675–711, 2018.
[29] S. Haridy, A. Maged, S. Kaytbay, and S. Araby, "Effect of sample size on the performance of Shewhart control charts," *The International Journal of Advanced Manufacturing Technology*, vol. 90, no. 1-4, pp. 1177–1185, 2017.

[30] B. G. Greenberg, R. R. Kuebler Jr, J. R. Abernathy, and D. G. Horvitz, "Application of the randomized response technique in obtaining quantitative data," *Journal of the American Statistical Association*, vol. 66, no. 334, pp. 243–250, 1971.

[31] S. Gupta, J. Shabbir, and S. Sehra, "Mean and sensitivity estimation in optional randomized response models," *Journal of Statistical Planning and Inference*, vol. 140, no. 10, pp. 2870–2874, 2010.

[32] K. H. Pollock and Y. Beck, "A comparison of three randomized response models for quantitative data," *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 884–886, 1976.

[33] C. J. Wild and G. A. F. Seber, *Chance Encounters: A First Course in Data Analysis and Inference*, John Wiley & Sons Inc., Hoboken, NJ, USA, 2000.