Rate-Quality Control Method of Identifying Hazardous Road Locations

ROBERT W. STOKES AND MADANIYO I. MUTABAZI

A brief historical perspective on the development of the rate-quality control method and its use in the identification of hazardous roadway locations is presented. The evolution of the formulas used in the rate-quality control method from their origin in the late 1950s to their present form is traced. The derivation of the basic formulas used in the method is also presented and discussed. It is suggested that, contrary to assertions in the literature, the accuracy of the equations used in the rate-quality method is not improved by eliminating the normal approximation correction factor from the original equations. The need for the correction factor is particularly apparent at higher probability levels. Charts are provided for determining an appropriate correction factor for those who may wish to incorporate these factors into the equations.

The rate-quality control method is used by many transportation agencies to identify hazardous road locations. This method uses a statistical test to determine whether the traffic accident rate for a particular intersection or roadway segment is abnormally high when compared with the rate for other locations with similar characteristics. The statistical test is based on the assumptions that traffic crashes are rare events and that the probability of their occurrence can be approximated by the Poisson distribution (1). The critical accident rate is determined statistically as a function of the systemwide average accident rate for the category of highway and the vehicle exposure (vehicles or vehicle kilometers) at the location being studied. If the actual (observed) accident rate for a particular roadway location is equal to or greater than the critical rate, the situation is caused by chance and the more likely that the process needs correction of some kind.

The critical rate for particular location (accidents per million vehicles or accidents per million vehicle-km), the upper and lower control limits on the overall accident rate are established for each control section. By applying the following equations, the upper and lower control limits on the overall accident rate are established for each control section.

\[
R_c = \lambda + k \sqrt{\frac{\lambda}{m}} + \frac{1}{2m} \tag{1}
\]

where

\[
R_c = \text{critical rate for particular location (accidents per million vehicles or accidents per million vehicle-km)},
\]

\[
\lambda = \text{average accident rate for all road locations of like characteristics (accidents per million vehicles or million vehicle-km)},
\]

\[
m = \text{number of vehicles traversing particular road section (millions of vehicle-km) or number of vehicles entering particular intersection (millions of vehicles) during the analysis period, and}
\]

\[
k = \text{probability factor determined by the level of statistical significance desired for } R_c.
\]

This paper provides a brief historical perspective on the development of the rate-quality control method and its use in the identification of hazardous roadway locations. The paper traces the evolution of the formulas used in the rate-quality control method from their origin in the late 1950s to their present form. The derivation of the basic formulas used in the method is also presented and discussed.

HISTORICAL PERSPECTIVE

Statistical quality control techniques were originally developed as a means to dynamically control the quality of industrial production. By setting upper and lower control limits on the amount of variability permitted in a particular process and by periodically sampling product quality, these techniques can provide a means of verifying that a process is in control. The control limits and the results of the periodic samples of product quality can be plotted on a control chart and any sample measures of product quality that fall outside the critical values established by the control limits are said to be out of control. The greater the difference between the observed sample values and the critical control limits, the less likely that the out-of-control situation is caused by chance and the more likely that the process needs correction of some kind.

In 1956, a method was proposed to analyze accident data for highway sections based on statistical quality control techniques (2). A procedure was described for determining the amount of variability in the accident rate that could be expected as a result of chance for any highway control section. By applying the following equations, the upper and lower control limits on the overall accident rate are established for each control section.

\[
\text{UCL} = \lambda + 2.576 \sqrt{\frac{\lambda}{m}} + \frac{0.829}{m} + \frac{1}{2m} \tag{2}
\]

\[
\text{LCL} = \lambda - 2.576 \sqrt{\frac{\lambda}{m}} + \frac{0.829}{m} - \frac{1}{2m} \tag{3}
\]

where

\[
\text{UCL} = \text{upper control limit},
\]

\[
\lambda = \text{average accident rate for all road sections of like characteristics (accidents per 10 million vehicles-mi)},
\]

\[
m = \text{number of vehicles traversing road section during analysis period (10 millions of vehicle-mi)},
\]

\[
\text{LCL} = \text{lower control limit}.
\]

In Equations 2 and 3, the first two terms result from the normal approximation to the Poisson distribution, the third term is a correction to the normal approximation, and the last term is a correction factor necessary because only integer values are possible for
the observed number of accidents [i.e., a correction for continuity necessary when using the normal distribution to approximate a discrete distribution (3, pp. 84–85)]. The coefficient of the second term (2.576) is based on a probability level of 0.995 for each control limit.

The third term in Equations 2 and 3 is particularly interesting. The normality correction factor in Equations 2 and 3 is a function of the desired probability level and an assumed mean accident frequency. As discussed subsequently, this factor appears to have been the subject of some confusion in later efforts to refine the original equations proposed by Norden et al. (2). It appears that some early researchers may have viewed Equations 2 and 3 as generic formulas, when in fact they were based on a specific probability level and accident frequency.

The underlying statistical theory for Equations 2 and 3 is presented in a subsequent section of this paper. That presentation is preceded by a continuation of the discussion about the evolution of the formulas originally developed by Norden et al.

In 1962 and 1967, modified forms (4,5) of the formulas developed by Norden et al. (2) in the analysis of highway crash data were applied. The effects of different probability levels were tested (4) by varying the $k$-values in Equations 2 and 3. However, the normality correction factors in Equations 2 and 3 were not adjusted to reflect the new $k$-values (4).

In 1967, it was suggested (5) that the “validity of the equations is improved if the correction term (0.829) as it appears in the original equations (Equations 2 and 3) is omitted.” Equations 4 and 5 reflect proposed revisions (5) to the original equations.

$$
UCL = \lambda + k \sqrt{\lambda \frac{1}{m} + \frac{1}{2m}} \tag{4}
$$

$$
LCL = \lambda - k \sqrt{\lambda \frac{1}{m} + \frac{1}{2m}} \tag{5}
$$

The recommendation (5) to eliminate the correction factor in the original equations was based on a comparison of the errors in the expected number of accidents ($\lambda m$) as estimated from Equations 2 and 4. The comparison was for cases where the average number of accidents varied from about 0.3 to 13 accidents for the 90 and 95 percent probability levels (see Table 1). A more equitable comparison would have been to compare the estimation errors from Equations 2 and 4 at the probability level used to determine the original correction factor (i.e., the 99.5 percent level). Alternatively, new correction factors could have been calculated based on the 90 and 95 percent probability levels and the accident frequencies used in the comparison (5). These points will be discussed further later in this paper.

Equations 4 and 5 represent the formulas most widely used to establish the upper and lower control limits for the rate-quality control method. Equations 4 and 5 are cited in descriptions of the rate-quality control method (1,6,7). FHWA (8), however, provides the following equation for calculating critical accident rates:

$$
R = \lambda + k \frac{\sqrt{\lambda \frac{1}{m} + \frac{1}{2m}}}{2m} \tag{6}
$$

Note that Equations 4 and 6 are identical except for the sign of the last term. Equations 4 and 6 illustrate a subtle difference in interpretation that results when the continuity correction factor is added to rather than subtracted from the equation. Equation 4 provides an expression for the upper control limit that has a probability $1 - P$ of being equaled or exceeded by chance, where $P = a$ given probability level. Equation 6 provides an expression for the upper control limit that has a probability $1 - P$ of being exceeded by chance.

### TABLE 1 Comparison of Equations 2 and 4 (5)

<table>
<thead>
<tr>
<th>Probability Level</th>
<th>Accident Frequency</th>
<th>% Error in Eq. 2</th>
<th>% Error in Eq. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0.325</td>
<td>33</td>
<td>-8.5</td>
</tr>
<tr>
<td>0.95</td>
<td>1.970</td>
<td>12</td>
<td>-4.6</td>
</tr>
<tr>
<td>0.95</td>
<td>11.638</td>
<td>3</td>
<td>-1.4</td>
</tr>
<tr>
<td>0.90</td>
<td>0.550</td>
<td>40</td>
<td>-1.8</td>
</tr>
<tr>
<td>0.90</td>
<td>1.103</td>
<td>26</td>
<td>-1.7</td>
</tr>
<tr>
<td>0.90</td>
<td>12.820</td>
<td>4</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

### ANOTHER LOOK AT RATE-QUALITY CONTROL METHOD

The following sections present a derivation of the basic equations used to establish the control limits on accident rates. The derivation provides a useful format to discuss the underlying statistical theory and to elaborate on several points of controversy alluded to earlier. The derivation closely parallels the approach taken by Norden et al. (2) in 1956.

#### Derivation of Control Limit Formulas

The occurrence or nonoccurrence of highway accidents may be modeled by a Bernoulli sequence, which, when stated in terms of the present problem, is based on the following assumptions: (a) each trial (vehicle-km) has only two possible outcomes (i.e., the occurrence or nonoccurrence of an accident), (b) the probability of occurrence of an accident in each veh-km is constant, and (c) the trials (veh-km) are statistically independent. If the probability of occurrence of an accident in each veh-km is $\lambda$ (and the probability of nonoccurrence is $1 - \lambda$), then the probability of exactly $x$ accidents in $m$ veh-km in a Bernoulli sequence is given by the binomial probability mass function (PMF) as follows (9):

$$
p(x) = \frac{m!}{x!(m-x)!} \lambda^x (1 - \lambda)^{m-x} \tag{7}
$$

Equation 7, although appealing in its simplicity, is readily applicable only for integer values of $m$ (veh-km). However, when $\lambda$ is small and $m$ is large such that $\lambda m$ remains fixed, it can be shown that a good approximation to $p(x)$ can be obtained from Equation 8 (10):

$$
p(x) = \frac{e^{-\lambda x} (\lambda m)^x}{x!} \tag{8}
$$

where $e$ is the base of the natural logarithms. Equation 8 is, of course, the familiar Poisson PMF. If the product $\lambda m$ (the expected number of accidents in $m$ veh-km) is replaced with $F_a$, Equation 8 can be rewritten as...
p(x) = \frac{e^{-\mu} \mu^x}{x!} \quad (8a)

Equation 8a describes the frequency of accidents as a Poisson process with mean and variance \( \mu \). The corresponding process for accident rates (\( R_x \)) is also Poisson with mean and variance \( \mu / m \).

Equation 8a can be used to formulate upper and lower control limits (confidence intervals) in terms of accident frequencies or accident rates. Because of the inherent intractability of evaluating the \( x/m \) factorial for the accident rate process, it is convenient to initially formulate the control limits in terms of accident frequency. The resulting control limits can then be converted to reflect accident rates by simply dividing by \( m \). The basic approach is illustrated in the following summary of the procedure used by Norden et al. (2) to estimate Equations 2 and 3.

A table of the Poisson distribution was used (2) to obtain an upper and lower limit, \( U \) and \( L \), on \( \mu \), (the expected number of accidents), such that \( p(X = U) = 0.005 \) and \( p(X = L) = 0.005 \), where \( X \) is the observed number of accidents along the test sections of a particular highway [the name or length of the highway is not specified (2)]. The resulting upper and lower limits on number of accidents were divided by \( m \) (veh-mi) to obtain the corresponding limits for the observed accident rate. Control charts showing the observed accident rate, the upper and lower control limits on accident rates, and the central value (assumed accident rate) were then plotted for each of the 18 highway intervals considered.

The formulation of control limits from a table of the Poisson distribution requires a double interpolation (for \( \mu \) and for \( x \)) for each road interval. It was reported (2) that the normal approximation to the Poisson provided an excellent approximation to the control limits without the need for tedious interpolations from the table of the Poisson distribution.

If the mean and variance of the Poisson distribution to be approximated are used to specify the mean and variance of the approximating normal distribution, then the general form of the equation for the confidence interval for \( \mu \) is

\[
\frac{U}{L} = \mu \pm Z \sqrt{\mu}
\]

(9)

where \( Z \) is the standard normal variate corresponding to the required confidence level.

All that remains to complete the derivation of the general formulas suggested (2) for computing the upper and lower limits on accident frequency is to insert the appropriate \( Z \)-score (2.576 for a 99.5 percent probability level for each limit), the normal approximation correction factor (0.829), and the continuity correction factor (0.5) into Equation 9. Incorporating these additional factors into Equation 9 and dividing by \( m \) (veh-mi) lead to the formulas for the upper and lower control limits on the overall accident rate given by Equations 2 and 3.

The need for the continuity correction factor in Equations 2 and 3 is well documented elsewhere (3, pp. 84–85) and is not recounted in this paper. The basis for the normal approximation correction factor in Equations 2 and 3, however, warrants additional discussion.

Normal Approximation Correction Factor

The normal approximation correction factor proposed (2) is examined in this section. The discussion focuses on the calculation of the correction factors and their effects on the accuracy of the equations. The discussion is limited to the correction factors for the upper control limits on accident frequency.

The normal approximation correction factor is simply the difference between the true and approximate (estimated) upper limits on accident frequency, as computed from Equations 10 and 11, respectively.

\[
p = \sum_{x=0}^{\infty} \frac{e^{-\mu} \mu^x}{x!}
\]

(10)

\[
U = F_a + Z \sqrt{F_a} + 0.5
\]

(11)

where

- \( p \) = prescribed probability level,
- \( U \) = true upper control limit on accident frequency, and
- \( U_a \) = estimated upper control limit on accident frequency.

Figures 1 through 3 show the normal approximation correction factors \( U - U_a \) for a range of frequencies from 0 to 30 accidents for standard probability levels of 0.90, 0.95, and 0.995, respectively. The correction factors in Figures 1 through 3 were calculated by selecting values for \( \mu \) that correspond to the specified probability levels. This approach was used (5) in constructing Table 1. [The authors’ calculations indicate that the first entry in the second column of Table 1 should be 0.352 instead of 0.325, as reported by Morin (5).]

Note in Figures 1 through 3 that the curves flatten noticeably for frequencies in the range of about five or more accidents. For the 90 and 95 percent probability levels the correction factors are relatively small and probably have little practical significance in terms of improving the accuracy of the equations. The correction factors for the 99.5 percent probability level (Figure 3), however, are of sufficient magnitude to substantially affect the accuracy of the equations. Note also that for the 99.5 percent probability level the correction factors for frequencies greater than about five accidents are generally consistent with the value of 0.829 suggested elsewhere (2) in Equation 2.
Tables 2 through 4 present comparisons of the accuracy of the formulas for the upper control limits with and without the normal approximation correction factors. As shown in the tables, the incorporation of the appropriate correction factors results in consistently better estimates of the upper control limits on accident frequencies. However, as noted, the correction factors for the 90 and 95 percent probability levels are relatively small and for practical purposes could be omitted from the equations.

In a previous section of this paper it was noted that Morin (5) suggested that the "validity of the equations is improved if the 'correction term' (0.829/m) as appears in the original equations (Equations 2 and 3) is omitted." That recommendation (5) to eliminate the correction factor in the original equations was based on a comparison of the errors in the expected number of accidents as estimated from Equations 2 and 4. The comparison was for cases where the average number of accidents varied from about 0.3 to 13 accidents for the 90 and 95 percent probability levels (see Table 1). A more equitable comparison would have been to compare the estimation errors from Equations 2 and 4 at the probability level used to determine the original correction factor (i.e., the 99.5 percent level). Table 4 shows that comparison. Alternatively, new correction factors could have been calculated (5) based on the 90 and 95 percent probability levels and the accident frequencies used in the comparison. Tables 2 and 3 show the results of those comparisons. As shown in Tables 2 through 4, the accuracy of the control limits is improved if the appropriate normal approximation correction factors are included in the equations. The improvement is particularly noteworthy at the 99.5 percent probability level.

**TABLE 2** Accuracy of Formula for Upper Control Limit With and Without Normal Approximation Correction Factor for 90 Percent Probability Level

<table>
<thead>
<tr>
<th>Accident Frequency</th>
<th>z-value</th>
<th>Percent Error w/correction</th>
<th>Percent Error w/o correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.105</td>
<td>1.282</td>
<td>0.64</td>
<td>2.04</td>
</tr>
<tr>
<td>0.532</td>
<td>1.282</td>
<td>-1.35</td>
<td>1.65</td>
</tr>
<tr>
<td>1.102</td>
<td>1.282</td>
<td>-0.57</td>
<td>-1.74</td>
</tr>
<tr>
<td>2.432</td>
<td>1.282</td>
<td>0.23</td>
<td>-1.37</td>
</tr>
<tr>
<td>11.135</td>
<td>1.282</td>
<td>-0.02</td>
<td>0.54</td>
</tr>
<tr>
<td>12.222</td>
<td>1.282</td>
<td>-0.01</td>
<td>-0.45</td>
</tr>
</tbody>
</table>

Note: Correction factors are from Figure 1.

**TABLE 3** Accuracy of Formula for Upper Control Limit With and Without Normal Approximation Correction Factor for 95 percent Probability Level

<table>
<thead>
<tr>
<th>Accident Frequency</th>
<th>z-value</th>
<th>Percent Error w/correction</th>
<th>Percent Error w/o correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.355</td>
<td>1.645</td>
<td>0.09</td>
<td>-8.24</td>
</tr>
<tr>
<td>0.818</td>
<td>1.645</td>
<td>0.03</td>
<td>-6.47</td>
</tr>
<tr>
<td>1.366</td>
<td>1.645</td>
<td>0.02</td>
<td>-5.28</td>
</tr>
<tr>
<td>1.57</td>
<td>1.645</td>
<td>0.00</td>
<td>-4.42</td>
</tr>
<tr>
<td>11.634</td>
<td>1.645</td>
<td>0.00</td>
<td>-1.42</td>
</tr>
<tr>
<td>12.442</td>
<td>1.645</td>
<td>0.00</td>
<td>-1.35</td>
</tr>
</tbody>
</table>

Note: Correction factors are from Figure 2.

**TABLE 4** Accuracy of Formula for Upper Control Limit With and Without Normal Approximation Correction Factor for 99.5 percent Probability Level

<table>
<thead>
<tr>
<th>Accident Frequency</th>
<th>z-value</th>
<th>Percent Error w/correction</th>
<th>Percent Error w/o correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.338</td>
<td>2.576</td>
<td>-0.18</td>
<td>-22.15</td>
</tr>
<tr>
<td>0.672</td>
<td>2.576</td>
<td>-0.01</td>
<td>-17.91</td>
</tr>
<tr>
<td>1.557</td>
<td>2.576</td>
<td>0.01</td>
<td>-12.82</td>
</tr>
<tr>
<td>2.038</td>
<td>2.576</td>
<td>0.02</td>
<td>-11.21</td>
</tr>
<tr>
<td>11.792</td>
<td>2.576</td>
<td>0.00</td>
<td>-3.92</td>
</tr>
<tr>
<td>12.521</td>
<td>2.576</td>
<td>0.00</td>
<td>-3.76</td>
</tr>
</tbody>
</table>

Note: Correction factors are from Figure 3.
SUMMARY AND CONCLUSIONS

This paper provides a brief historical perspective on the development of the rate-quality control method and its use in the identification of hazardous roadway locations. The paper traces the evolution of the formulas used in the rate-quality control method from their origin in the late 1950s to their present form. The derivation of the basic formulas used in the method is also presented and discussed.

It is suggested that, contrary to assertions in the literature, the accuracy of the equations used in the rate-quality method is not improved by eliminating the normal approximation correction factor from the original equations (Equations 2 and 3). The need for the correction factor is particularly apparent at higher probability levels.

Equation 12 can be used by the analyst who wishes to consider the normal approximation correction factor when calculating the upper control limits on accident frequency.

\[ \text{UCL} = F_0 + Z \sqrt{F_0} + c + 0.5 \]  

(12)

where \( c \) is the normal approximation correction factor. The appropriate \( c \)-value can be determined from Figure 1, 2, or 3.

Similarly, Equations 2 and 3 can be generalized to reflect the upper and lower control limits for accident rates at the desired probability level by replacing the coefficients of the second terms (2.576) and the numerators of the third terms (0.829) with the appropriate \( Z \)- and \( c \)-values, respectively.

REFERENCES


*Publication of this paper sponsored by Committee on Statistical Methodology and Statistical Computer Software in Transportation Research.*