Annals of Operations Research 67(1996)45–60

Risk-adjusted control charts for health care assessment

Farrokh Alemi

Health Care Administration Program, Cleveland State University, Cleveland, OH 44115, USA

Walter Rom

Production and Operations Management Department, Cleveland State University, Cleveland, OH 44115, USA

Eric Eisenstein

Outcomes Research and Assessment Group, Duke University Medical Center, Durham, NC 27710, USA

The recent Joint Commission on Accreditation of Healthcare Organization (JCAHO) requirement that hospital accreditation be based upon a Total Quality Management (TQM) approach has focused the attention of health care administrations on the use of techniques such as control charts. However, control charts are not typically adjusted for severity of illness. This adjustment is needed because, unlike industrial organizations, hospitals are not able to control all of their inputs and must accept variances in their patients. In this paper, we present a methodology for adjusting a health care organization's control charts to reflect their patient population's severity of illness during different time intervals. We then demonstrate that risk-adjusting expected patient outcomes can change our assessments of the relative quality of care offered by a health care organization in different time periods.

Keywords: Control chart, quality of care, risk adjustment, severity of illness, total quality management.

1 Background

1.1 *Health care quality management*

"Quality is all of the features and characteristics of a product or service that contribute to the satisfaction of a customer's needs (Besterfield, 1986)." Until recently, most health care administrators thought that the quality of their organization's products and services (patient care) was a clinical, rather than an administrative, concern. And, they relied upon health care professions to monitor the quality of care in their organizations. With the introduction of prospective payment systems, all components of the

health care industry have become far more competitive. Health care organizations have responded to these forces by becoming increasingly concerned about their quality of care and have sought effective means for managing it. Identifying patient outcomes and developing algorithms to estimate those outcomes under different sets of circumstances are a necessary prerequisite for effective quality of care management. It is because of these needs to relate a specific patient's outcomes to that patient's particular circumstances that quality control has become an integral part of health care quality management programs.

1.2 *Patient outcomes assessment*

Mortality (patient lives or dies) is frequently used to categorize outcomes for patients who have an acute, life-threatening illness such as myocardial infarction. However, morbidity or length of stay might be better outcome categories for patients who have chronic diseases such as hypertension or diabetes where death may be less immediately threatening. In this paper, we will use mortality (yes/no) as our outcome category in all of our examples. Within both the group of patients who live and the group that dies, some patients might receive good care while other patients in that same outcome category receive poor care. Likewise, some patients will be so ill that they cannot be saved by the best of care, while other patients will be saved despite the fact that they have received poor care. These interactions between patient outcomes, quality of care, and the severity of a patient's illness suggest that when a patient dies, his/her mortality might be attributed to either the severity of illness or to the quality of care he/she received (Gustafson et al., 1986; Hannan and Hartman, 1989). Because of the radically different implications of these attributions, it is important that health care organizations be able to separate the influences of quality of care from those of severity of illness when assessing patient outcomes. Without this ability, instances of poor and good quality of care cannot be determined. Risk-adjustment provides a means for making these distinctions and it has been used successfully to risk-adjust patient mortality as well as patient costs of care.

1.3 *Risk-adjustment studies*

Dubois et al. (1987) used a linear regression model to adjust the actual death rates in 93 hospitals. In this study, the death rates for hospitals ranged from 0.3 to 5.8 percent. After adjusting for different average patient risk levels, the predicted death rates range for these hospitals was reduced by 0.36 to 1.36 percent. Eleven (11.8%) of the hospitals in this study had death rates which were more than two standard deviations greater than their predicted death rates and nine (9.7%) of the hospitals had death rates which were more than two standard deviations less than their predicted death rates. Thus, a total of 20 of the 93 hospitals (21.5%) had important differences between their actual and predicted mortality rates which could be attributed to differences in patient risk levels.

Horn et al. (1986) demonstrated the importance of adjusting expected patient costs to reflect differences in severity of illness. In their study of 467 diagnosis-related groups (DRGs) in 15 hospitals, they found that a four-category severity of illness index explained more than 10% of the variability in costs for 94% of their DRGs and more than 50% of the variability in costs for 36% of their DRGs. Thus, in approximately one-third of the DRGs studied, severity of illness accounted for more than half of the variability in total patient costs. For their entire patient population, DRGs explained 28% of the variability in costs, whereas severity-adjusted DRGs explained 61% of the variability in costs. In contrast, after the effects of severity-adjusted DRGs were removed, a hospital variable only accounted for 6.3% more of the variability in patient costs. Thus, even a simple severity of illness index can have a significant impact in predicting patient costs and can have even a greater impact than the diagnoses themselves. Another study by the same authors (Horn et al., 1985) investigated the impact of risk adjustment on prospective payments in fifteen hospitals. This study found that risk adjusting prospective payments had as much as a 35% impact in total operating costs. Thus, these studies confirm that the individual characteristics of patients can be important in determining overall treatment costs for patients as well as their predicted mortality.

1.4 *Severity of illness measurement*

Many methods have been proposed for measuring severity of illness in patients. Some of these differences in methods are caused by different definitions of severity of illness and others are attributable to different assumptions regarding how/when severity of illness should be measured (Thomas and Lango, 1990; Fetter et al., 1991). Even for mortality, perhaps the simplest severity of illness concept, there are no commonly accepted modeling standards (Hadorn et al., 1993). The sections which follow briefly describe seven severity of illness measurement systems. The relative accuracies of these systems were compared in a previous study of myocardial infarction patients (Alemi et al., 1990).

Computerized Disease Staging (CDS) (Gonnella et al.,1984): This index is designed for a broad range of patients and is not limited to specific illnesses. Disease staging assumes that diseases are first localized and later spread to other body systems. As a disease advances to higher numbered stages, it is associated with increasing risk for the patient. The CDS system has four stages which can be subdivided into additional categories. Stage 1 includes conditions with no complications and minimal severity risk for the patient. Stage 2 includes problems which are contained in one organ or system. Stage 3 includes problems in multiple sites and general systemic problems. Stage 4 is death. As frequently implemented, CDS collects ICD-9-CM codes from discharge abstracts, does not include comorbidities or death adjustments, and uses an ordinal scale to represent severity of illness.

Patient Management Categories (PMC) (Young, 1985): This index is designed for a broad range of patients and is not limited to specific illnesses. Patient manage-

ment categories specify the costs for different care regimens. Since increasing costs are often associated with an increasing severity of illness, PMC's cost weights are frequently used as an indication of the patient's severity of illness. As frequently implemented, PMC collects ICD-9-CM codes from discharge abstracts, does not include procedures or death adjustments, and uses an interval scale to represent severity of illness.

Acute Physiological and Chronic Health Evaluation (APACHE II) (Knaus et al., 1985): This index is designed for a broad range of patients and is not limited to specific illnesses. The APACHE II score is the sum of three components. These are: (1) deviations from norm on 12 physiological variables, (2) an age index, and (3) a chronic conditions index. As frequently implemented, APACHE II collects physiological findings from chart reviews at admission and uses an interval scale to represent severity of illness.

Medisgroup (MDGRP) (Brewster et al., 1985): This index is designed for a broad range of patients and is not limited to specific illnesses. MDGRP scores have five levels, 0 to 4. At Level 0, there are no clinical findings. At Level 1, there are minimal abnormal findings. At Level 2, there are either acute findings with an unclear potential for organ failure or severe findings with a high probability of imminent organ failure. At Level 3, there is a high potential for imminent organ failure. And, at Level 4, organ failure is indicated. Unlike many of the other systems, there are sets of MDGRP indicators which are most relevant for different diagnoses. As frequently implemented, MEDISGRP collects physiological findings from chart reviews at admission and uses an ordinal scale to represent severity of illness.

Computerized Severity Index (CSI) (Backofen et al.,1987): This index is designed for a broad range of patients and is not limited to specific illnesses. As frequently implemented, CSI collects ICD-9-CM codes from discharge abstracts and physiological findings from chart reviews at admission. There is no death adjustment and an ordinal scale is used to represent severity of illness. CSI is the only severity of illness system described in this section which contains both diagnoses and physiological findings. CSI begins with the patient's principal diagnosis and uses physiological findings to adjust the diagnosis. CSI's scores range from 0 to 4.

Predictive Index for Myocardial Infarction (PIMI) (Pozen et al., 1980): This index was designed specifically for heart disease patients. Although it has sometimes been scored from chart reviews, it is intended to be used prospectively to estimate the probability of a myocardial infarction and to assist physicians in determining which patients might most benefit from intensive coronary care. The assumption in using PIMI for severity of illness measurement is that patients who could most benefit from intensive care are also those patients with the most severe illnesses. When used retrospectively, PIMI collects physiological findings from chart reviews at admission and uses an interval scale to represent severity of illness.

Ischemic Heart Disease Index (IHDI) (Gustafson et al., 1986): This index was designed specifically for heart disease patients. As frequently implemented, IHDI

produces a score ranging from 0 to 100. IHDI collects physiological findings from chart reviews at admission and uses an interval scale to represent severity of illness.

2 Control charts

1.1 *Introduction*

Many authors have suggested the use of industrial quality control methods for managing the quality of care in health care organizations (Ramsey and Cantrell, 1985; Berwick, 1989; Laffel and Blumenthal, 1989). The recent Joint Commission on Accreditation of Healthcare Organization (JCAHO) requirement that hospital accreditation be based upon a Total Quality Management (TQM) approach has focused the attention of health care administrators on the use of techniques such as control charts. However, control charts are not typically adjusted for severity of illness. In this paper, we present a methodology for doing so.

Control charts have traditionally been used to manage the quality of industrial processes by assisting administrators in determining whether the outputs of those processes are in or out of control. A process is defined as being out of control when its observed outcomes are different from those which would be expected to occur through random variations. When the outputs of the process are within the tolerances which would be expected by random variations, the process is defined as being in control. And, when it is out of those tolerances, the process is defined as being out of control. Usually, these tolerances are identified by two means: control limits attached to the observations and a combination of flags which look for trends, cycles, and other abnormal patterns in the observations.

One of the primary quality management differences between industrial organizations and hospitals is that industrial organizations are able to control their inputs, while hospitals must accept variances in their inputs (patients). Because industrial organizations are able to standardize their inputs, they can assume that any output variations must be caused by process variations. However, since hospitals are not able to control their inputs, they must separate their output variations into those components which are caused by differences in inputs (severity of illness) and those which are caused by differences in processes (quality of care). The remaining sections of this paper give examples both of how control charts and mortality rates have traditionally been applied to make quality of care comparisons as well as of how they should be used to adjust mortality rates to account for differences in severity of illness.

2.2 *Control charts without risk-adjustment*

Table 1 contains control chart information which we created for comparing mortality rates in different time periods (T_1 through T_8) for the same health care organization. Initially, this data is used to estimate p (the probability of mortality). Later, we can use the control chart's upper and lower control limits to assess whether measurements

Table 1

Unadjusted time period mortality rates.

Time period	No. cases	No. deaths	Observed mortality	90% *UCL*	90% *LCL*
T_1	186	49	0.2634	0.3022	0.1970
T_2	119	24	0.2017	0.3153	0.1838
T_3	111	25	0.2252	0.3177	0.1815
T_4	26	3	0.1154	0.3903	0.1089
T_5	39	15	0.3846	0.3645	0.1347
T_6	23	5	0.2174	0.3992	0.1000
T_7	61	16	0.2623	0.3414	0.1577
T_8	20	9	0.4500	0.4100	0.0891
Total	585	146	0.2496		

in subsequent time periods are from processes which may have control problems. Readers who are familiar with control charts will note that this table differs from traditional control chart information. Instead of sampling with a fixed sample size, this method includes in the sample all patients meeting the selection criterion who were seen at the health care organization during a particular time period. Thus, the number of cases varies between samples (time periods) and different sample sizes are used for each time period's mortality rate computation. Although a constant sample size may be preferable in industrial control charts, this is not possible when each sample includes all of the patients seen during a fixed time interval.

The columns 90% *UCL* and 90% *LCL* in table 1 represent the Upper Control Limits and Lower Control Limits, respectively, for each sample mortality rate. These two values are used to assess the significance of variations in individual time period mortality rates from the average mortality rate in all eight time periods. The reader should note that there are no standard percentages for the upper and lower control limits and that their determination is a matter of policy for each health care organization. When an individual time period's mortality rate is above the *UCL*, the processes which produced that measurement could be out of control and their output quality is of concern. Conversely, when a time period's mortality rate is below the *LCL* line, they could represent processes of exceptional quality. If a time period's mortality rate is between the *UCL* and *LCL* values, the processes which produced that measurement are most likely in control and any differences from other time period measurements are most likely due to random chance. However, the reader should look for trends and cycles in the measurements before concluding that these processes are in fact in control. Appendix A contains a more detailed explanation of control chart indicators and their usage in determining processes which are out of control.

The calculations in table 1 are not difficult. The number of cases and number of deaths in each time period are inputs to the formulas. The formulas for calculating the population mortality rate ($\overline{P}$), individual time period observed mortality rates (P_i), sample-adjusted population standard deviations (S_i), sample-adjusted upper control limits (UCL_i), and sample-adjusted lower control limits (LCL_i) are given below as equations (1) through (5). In these equations, d_i is the number of deaths for a time period, n_i is the number of cases for a time period, and $t_{\alpha/2}$ is the t-value.

The t-value for the UCL and LCL equations is determined by the time period's sample size (number of cases) and by the study's desired confidence level. A 90% confidence level was selected for this example. This confidence level means that 90% of the sample measurements (time period mortality rates) will normally be within the control chart's upper and lower control limits. Conversely, 10% of the sample measurements will normally be outside of the control chart's upper and lower control limits.

Although we use the t distribution in all of our examples, the reader should be cautioned that use of the symmetric t distribution as an approximation for the asymmetric binomial may yield poor approximations when there are either small probabilities or small sample sizes. In these instances, the reader should use software packages which can provide bounds computed directly from the binomial distribution. These limits may differ from the ones computed by using the t distribution.

$$\overline{P} = \frac{\sum_{i=1}^{m} d_i}{\sum_{i=1}^{m} n_i}, \tag{1}$$

$$P_i = \frac{d_i}{n_i}, \tag{2}$$

$$S_i = \sqrt{\frac{\overline{P} * (1 - \overline{P})}{n_i}}, \tag{3}$$

$$UCL_i = \overline{P} + t_{\alpha/2}(S_i), \tag{4}$$

$$LCL_i = \overline{P} - t_{\alpha/2}(S_i). \tag{5}$$

We can use these formulas to calculate the Population Mortality Rate and the Observed Mortality Rate, the Population Standard Deviation, and the Upper and Lower Control Limits for time period, T_1. These calculations are shown below in equations (6) through (10).

$$\overline{P} = \frac{146}{585} = 0.2496, \tag{6}$$

$$P_1 = \frac{49}{186} = 0.2634, \tag{7}$$

$$S_1 = \sqrt{\frac{0.2496 * (1 - 02496}{186}} = 0.0317, \tag{8}$$

$$UCL_1 = 0.2496 + (1.658 * 0.0317) = 0.3022, \tag{9}$$

$$LCL_1 = 0.2496 - (1.658 * 0.0317) = 0.1970. \tag{10}$$

Figure 1 is a control chart which graphically displays the data in table 1. This type of control chart is called a p chart because it measures the probability that an outcome will occur. In this chart, mortality rates (the probabilities that patients will die in specific time periods) are measured along the vertical axis and the different time periods are measured along the horizontal axis. There are three lines in figure 1 and a series of unconnected dots. The dots are the observed mortality rates for each of the time periods.

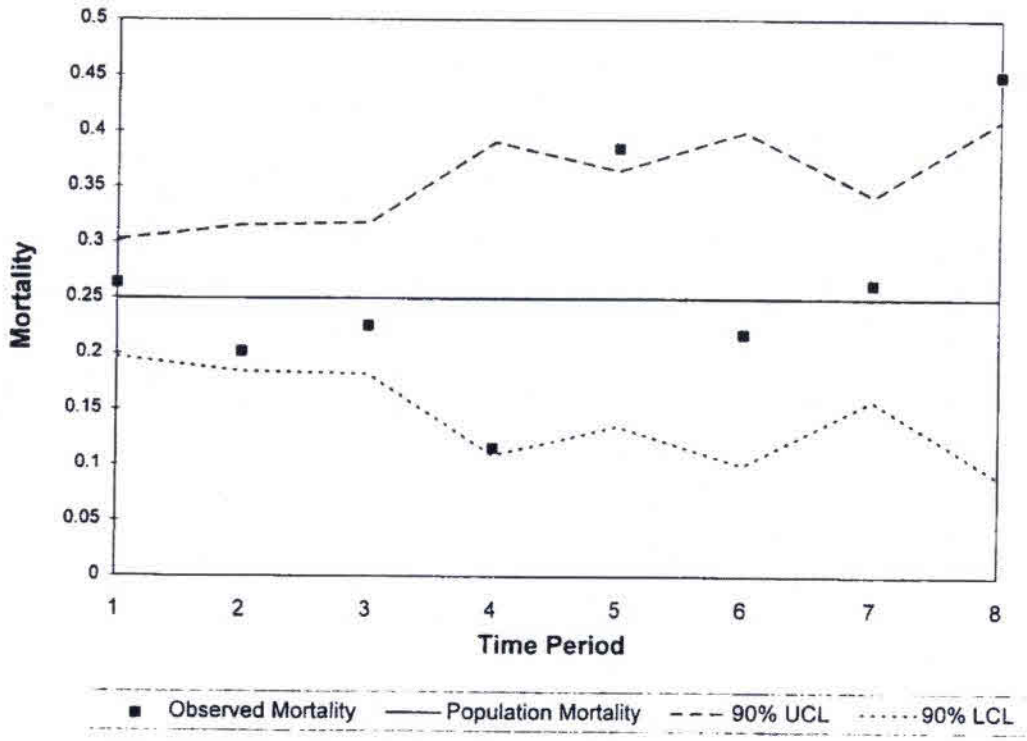

Figure 1. Unadjusted mortality rate control chart.

The middle line in figure 1 is parallel to the x axis at a 24.96% mortality rate and it is the Population Mortality Rate. Since this measurement is a weighted average, it will assign greater importance to the outcomes in the three time periods with the largest number of cases (T_1 through T_3) and will assign lesser importance to the outcomes in the time periods with fewer cases (T_4–T_8). Because of the large difference

in the number of cases in different time periods, this measurement has the effect of assessing the quality of care in the five smaller time periods according to the standard of care which is set by the three larger time periods. Even though this presents a problem, we would still recommend sampling health care organizations by fixed time intervals instead of using a fixed sample size (e.g. each 100 patients). This is because sampling at fixed intervals which correspond to the administrative cycles of individual organizations allows quality management to be more easily incorporated into the normal management processes of these organization.

The two lines on either side of the Population Mean Mortality rate are the 90% *UCL* and the 90% *LCL* for each time period. In this example, the 90% *UCL* and the 90% *LCL* for T_1 with 186 cases are at 30.22% and 19.70%, respectively, while the 90% *UCL* and the 90% *LCL* for T_8 with only 20 cases are at 41.00% and 8.91%, respectively. In reviewing control charts, it is important to remember that a larger number of cases decreases the value of the standard error of the mean and a smaller number of cases increases the value for the standard error of the mean. As a consequence, when the sample size increases, the *UCL* and *LCL* become closer to each other. Using the 90% control limits to define time period quality, we see that the mortality rates in time periods T_5 and in T_8 are beyond our control limits and we need to investigate why this is so. Since, at a 90% control limit, 10% or 0.8 time periods would be expected to be outside of the upper and lower control limits, this appears to be a situation in which one or both of these time periods has a significant quality deviation and they both should be investigated more thoroughly using the techniques in appendix A.

Although this discussion has evaluated all eight time periods retrospectively for illustrative purposes, a more realistic presentation would evaluate each time period as it occurred. Thus, the potentially out-of-control situation in time period T_5 would have been evaluated at the end of that time period and any remedial action would have been implemented during time period T_6. As a consequence, when the process was again potentially out of control in time period T_8, the remedial action would have taken effect and the health care organization's management team would evaluate other alternatives.

The above discussion summarizes the manner in which control charts and mortality rates have traditionally been used to make comparison between time periods. The major problem with this technique is that it assumes that all eight time periods in the study treat patients with the same average severity of illness. Specifically, since all deviations from the population mortality rate are attributed solely to differences in quality of care, there is an implicit assumption that the severity of illness in patients seen in all eight time periods is the same. While it might be permissible to make this assumption in rare instances, severity of illness generally varies considerably between patients and time periods. The section which follows demonstrates how control chart theory can be modified to accommodate different severity of illness levels in different time periods while still being able to make inter-time period mortality rate comparisons.

2.3 *Risk-adjusted control charts*

Table 2 contains risk-adjusted control chart data for the eight time periods in this study. Six of the eight columns in table 2 have identical titles to those in table 1. The only differences are Expected Mortality and $(\sum p * q)/n$. Expected Mortality is the risk-adjusted expected mortality rate, and $(\sum p * q)/n$ is the sum of the products of the probabilities for mortality and the probabilities for survival for each case in a time period divided by the number of cases. Four steps are required to risk-adjust control chart data. These steps are: (1) determine the number of expected deaths after risk-adjustment in each of the time periods, (2) calculate the expected mortality rate for each time period, (3) calculate the standard deviation of the expected mortality rate for each time period, and (4) calculate the risk-adjusted *UCL* and *LCL* values for each time period.

Table 2

Risk-adjusted time period mortality rates.

Time period	No. cases	No. deaths	Observed mortality	Expected mortality	$(\sum p * q)/n$	90% *UCL*	90% *LCL*
T_1	186	49	0.2634	0.2514	0.1274	0.2948	0.2080
T_2	119	24	0.2017	0.2315	0.1063	0.2811	0.1819
T_3	111	25	0.2252	0.2192	0.1090	0.2712	0.1672
T_4	26	3	0.1154	0.2230	0.0967	0.3241	0.1219
T_5	39	15	0.3846	0.2712	0.1173	0.3621	0.1803
T_6	23	5	0.2174	0.2166	0.1156	0.3341	0.0991
T_7	61	16	0.2623	0.1989	0.0955	0.2645	0.1333
T_8	20	9	0.4500	0.4052	0.0933	0.5184	0.2920
Total	585	146	0.2496				

We begin by predicting who is likely to die. Most statistical packages have a facility for generating such predictions. In a regression analysis, the model's dependent variable is the observed mortality (1 if the patient dies, 0 otherwise); the independent variables can be chosen from the severity of illness systems which were previously reviewed or they can be disease specific. In our example, our independent variables were Computerized Disease Staging (6 of 7 categories), Patient Management Categories (1 of 4 cost weights), Acute Physiological and Chronic Health Evaluation combined score, Medisgroup (4 of 5 levels), Ischemic Heart Disease Index score, sex, and age.

A regression model which includes all the patient data is used to predict outcomes for each patient. Once the cases are assessed, their predictions are summed by time period to yield the expected number of deaths that should occur in each of the eight time periods. Since the model only uses severity of illness variables to predict the

number of patients not discharged alive in each of the time periods, this model in effect produces a risk-adjusted estimate of the number of deaths that should have occurred in each time period.

Now that the number of risk-adjusted deaths has been calculated for each time period, the next steps are to calculate the risk-adjusted expected mortality rates ($\hat{P}_i$) and the risk-adjusted standard deviations ($\hat{S}_i$) for these same time periods as well as their risk-adjusted lower and upper control limits (UCL_i and LCL_i). The formulas for these computations are shown below as equations (11) through (14). In these equations, $\sum_{j=1}^{n_i} \hat{P}_{ij}$ is the expected number of deaths for a time period, n_i is the number of cases for a time period, and $t_{\alpha/2}$ is the t-value for a particular confidence level and sample size. Here, $\hat{P}_{ij}$ is the probability that patient j in time period i dies. By theorem 6.9.3, p. 207, of Fisz (1963), we know that the mortality rate has an asymptotically normal distribution provided that there is a sufficient sample size.

$$\hat{P}_i = \frac{\sum_{j=1}^{n_i} \hat{P}_{ij}}{n_i}, \tag{11}$$

$$\hat{S}_i = \frac{\sqrt{\sum_{j=1}^{n_i} (\hat{P}_{ij} * (1 - \hat{P}_{ij}))}}{n_i}, \tag{12}$$

$$UCL_i = \hat{P}_i + t_{\alpha/2}(\hat{S}_i), \tag{13}$$

$$LCL_i = \hat{P}_i - t_{\alpha/2}(\hat{S}_i). \tag{14}$$

We can use the above formulas to calculate the risk-adjusted expected mortality rate and expected standard deviation, and expected upper and lower control limits for time period, T_1. These are shown below in equations (15) through (18).

$$\hat{P}_1 = \frac{46.8}{186} = 0.2514, \tag{15}$$

$$\hat{S}_1 = \sqrt{\frac{0.1274}{186}} = 0.0262, \tag{16}$$

$$UCL_1 = 0.2514 + (1.658 * 0.0262) = 0.2948, \tag{17}$$

$$LCL_1 = 0.2514 - (1.658 * 0.0262) = 0.2080. \tag{18}$$

The risk-adjustment model increases and decreases the expected number of deaths in the eight time periods to compensate for patient risk levels which differ from the overall risk level in the entire patient population. This is demonstrated in table 2 by the fact that while the population's mortality rate is constant at 24.96%, the expected

mortality rate for individual time periods ranges from a minimum of 19.89% in T_7 to a maximum of 40.52% in T_8. Thus, the deviation of a time period's observed mortality rate from its risk-adjusted expected mortality rate is a more accurate measure of its quality of care than is achieved when the time period's observed mortality rate is compared with the population's mortality rate in the model which is not risk-adjusted.

Figure 2 is the risk-adjusted control chart for our eight time period example. Notice that both axes (time periods and mortality rates) are identical in the unadjusted and in the risk-adjusted control charts. Likewise, there are three lines and a series of unconnected dots in figures 1 and 2. However, despite these similarities, only the

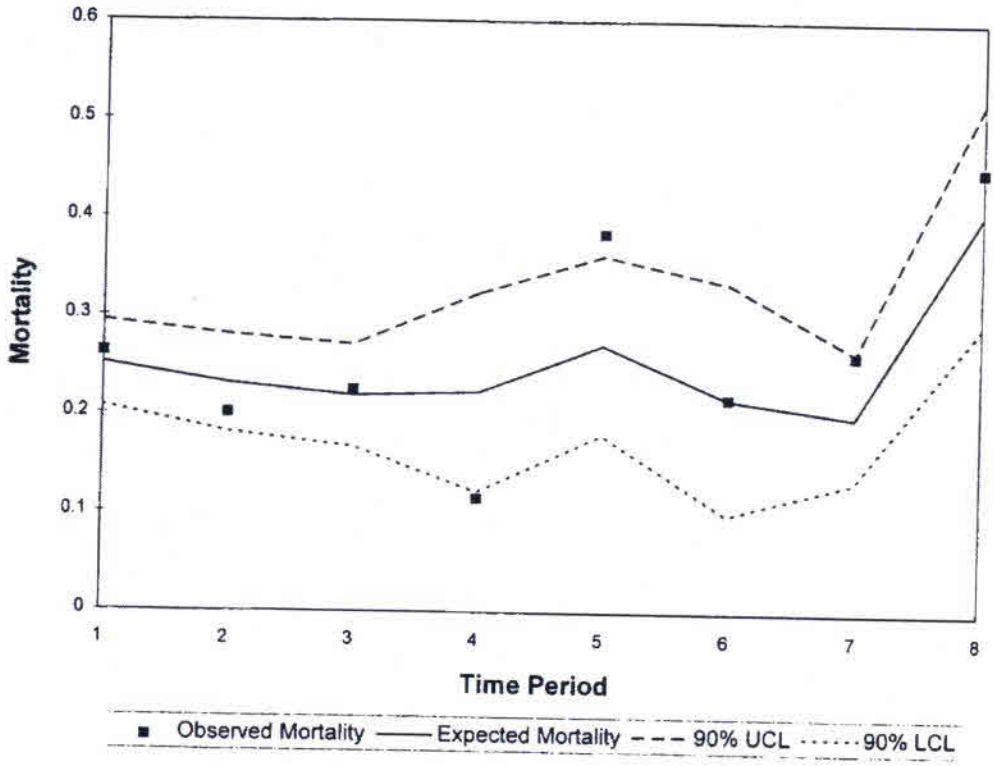

Figure 2. Risk-adjusted mortality rate control chart.

series of unconnected dots have the same appearance in both figures. This is because the series of dots represents the observed mortality rates in each of the time periods. The three lines have a different appearance because the expected mortality rates and upper and lower control limits have been adjusted in figure 2 to account for the different severity of illness levels experienced in each of the time periods.

The major change between figure 1 and figure 2 is the substitution of individual time period Expected Mortality Rate measurements ($\hat{P}_i$) for the observed Population Mortality Rate ($\overline{P}$) line as the standard by which observed time period mortality rates will be assessed. Thus, instead of a constant line at a 24.96% mortality rate, we now

have a line which bends to accommodate different severity of illness levels in each of our eight time periods. Since the Expected Mortality Rate is now the reference for calculating the upper and lower control limits, we can also expect that both of these lines are adjusted up or down as the Expected Mortality Rate moves up and down. Thus, by risk-adjusting the expected mortality rates to account for severity of illness differences, we have also risk-adjusted the control limit lines. These new control limits are termed the Expected Upper Control Limit (UCL_i) and the Expected Lower Control Limit (LCL_i).

In figure 1, two time periods (T_5 and T_8) had observed mortality rates which were above the 90% *UCL* and this was identified as a possible out-of-control situation. In figure 2, after adjusting for different time period risk levels, two time periods are outside of the upper and lower control limits. However, one of the out-of-control time periods in figure 2 is different from figure 1, and this time period, T_4, might actually be an example of exceptionally good quality of care. Had we acted on the basis of the unadjusted control chart, we also would have investigated an abnormal level of quality in T_8 when it did not exist. This is because what the unadjusted control chart identified as abnormal quality in time period T_8 turned out to be a patient severity of illness level which was much above the average population mortality rate and which led to a large difference between this time period's observed mortality rate and the population mortality rate. Thus, the value of risk-adjusting control chart data has been demonstrated.

3 Discussion

Traditional methods for measuring quality of care do not separate deviations from standard quality levels into their components which can be attributed to severity of illness and those other components which are attributable to other sources. This distinction is important because the deviations which are not related to severity of illness may be related to differences in quality of care. Risk-adjusted control charts provide a tool for making these distinctions. In this paper, we have demonstrated a methodology for establishing upper and lower control limits to monitor health care processes. Since the control charts in this example are used for general problem diagnosis, it is permissible to have a 90% confidence level. In other situations where control charts are used to take corrective action, it might be more appropriate to use a wider confidence interval (95% or 99%).

Small samples are a cause for concern, and when the sample size is less than 30, normality assumptions may not hold. However, health care administrators and researchers are faced with the choice of either excluding small samples from their studies or including the data and exercising judgement when acting upon the results. Again, the 90% confidence interval used in this paper's examples is intended only as a flag to denote situations which need further scrutiny with supplementary techniques. For example, is there any reason that the underlying processes in the health care

organization might be experiencing larger than normal deviations? If so, their cause might provide the supplementary analysis that is required.

In this paper, we have demonstrated that risk-adjusting time period mortality data can have a dramatic impact upon how we assess the comparative performance of a single health care organization in different time periods. We have also applied these methods to one of the TQM techniques, p charts. Other TQM control chart techniques (such as the $\bar{x}$ chart for process means and the $\bar{c}$ chart for the number of defectives) offer similar benefits for health care organizations and future researchers should devise techniques for modifying them to account for differing patient risk levels. With these modifications to the $\bar{x}$ and $\bar{c}$ charts, other quality management tools will become available to health care administrators and practitioners.

Although this paper has focused upon the use of risk-adjusted control charts as techniques for quality management within a single health care organization, we believe that these techniques can also be used by health care policy makers and purchasers (e.g. business groups, insurance companies, and individuals) to assess the relative qualities of care across health care organizations. Once it is determined that the processes in individual health care organizations are in control, the next step would be to compare risk adjusted outcomes between different health care organizations. If these types of comparisons are not preceded by risk-adjustment, those organizations with higher risk levels will be penalized in the analysis. However, with risk-adjustment, health care decision and policy makers will be able to assess which health care organizations really are providing the most effective care for their patients.

As we stated above, mortality is only one of the possible severity of illness measurements. For chronic illnesses, other severity measures such as length of stay or morbidity would be more appropriate. With these other severity measures, the computations would essentially be the same. The only difference would be that instead of computing the percentage of patients who died, one would be computing the percentage of patients who recovered from the illness or were discharged within a specified time period. Again, although there are many general severity measurements, it is often better to choose those measurements that are tailored for specific diseases.

We believe that control charts are a viable technique for health care quality analysis and management. They are easily computed, offer a visual representation of the problem, are adaptable to many types of information, and can be used to model a wide range of problems. It is because of these characteristics that they should be incorporated into the quality management programs of health care organizations.

Appendix A: Control chart indicators for out-of-control processes

In addition to looking to the appropriate control limit lines (*UCL* and *LCL*) to assess deviations from standard, Besterfield (1986) has suggested five primary indicators for out-of-control processes. The first indicator is a change or jump in the quality

measurement's level. This is often caused by a change in inputs (people, material, or equipment) or by a change in the process itself. The second indicator is a trend or steady change in the quality measurement's level without intervention. This is often due to gradual changes in process inputs. Besterfield's third indicator is recurring cycles. These occur in control charts when there are periodic high and low points. These are normally caused by seasonal effects in inputs or differences that occur on a daily or weekly basis. The fourth indicator is a large number of points near or outside the control limits. This often denotes that two or more populations are being measured which have large differences in their inputs or processes. The remedy in these situations is to construct multiple control charts. Besterfield's last indicator is errors is the quality measurements themselves.

Several authors have suggested the use of zone control charts with scoring rules to determine whether processes are out of control. Normally, zone control charts are divided into eight zones, four on each side of the mean. Zone A is the area between the mean and one standard deviation, zone B is the area between one and two standard deviations, zone C is the area between two and three standard deviations, and zone D is the area greater than three standard deviations. Evans (1993) proposes a simple set of scoring rules. These rules define a process as being potentially out of control when any one of the following situations occurs on a control chart:

1. Any point falling in zone D.
2. Two of three successive points falling in zone C on the same side of the mean.
3. Four of five successive points falling in zone B or C on one side of the mean.
4. Eight consecutive points falling in zone A, B or C on the same side of the mean.

These scoring rules assist in identifying out of control processes which do not exceed the upper and lower control limit criterion.

References

Alemi, F. and R. Hankin, Predicting in-hospital survival of myocardial infarction, Medical Care 28, 1990, 9.

Backofen, J., A. Ashworth and S.D. Horn, The computerized severity of illness index: A new sophisticated tool to measure hospital quality of care, Health Care Forum 2, 1987.

Berwick, D.M., Continuous improvement as an ideal in health care, The New England Journal of Medicine 320, 1989, 1.

Besterfield, D.H., *Quality Control*, Prentice-Hall, Englewood Cliffs, NJ, 1986.

Brewster, A.C., B.G. Karlin, L.A. Hyde, C.M. Jacobs, R.C. Bradbury and Y.M. Chae, A clinically based approach to classifying hospital patients at admission, Inquiry 12, 1985.

Dubois, R.W., R.H. Brook and W.H. Rogers, Adjusted hospital death rates: A potential screen for quality of medical care, American Journal of Public Health 77, 1987, 9.

Evans, J.R., *The Management and Control of Quality*, West Publ., St. Paul, MN, 1993.

Fetter, R.B., D.A. Grand and D. Gamache, *DRGs: Their Design and Development*, Health Administration Press, Ann Arbor, MI, 1991.

Fisz M., *Probability Theory and Mathematical Statistics*, Wiley, New York, 1963.
Gonnella, J.S., M.C. Hornbrook and D.Z. Louis, Staging of disease: A case mix measurement, Journal of the American Medical Association 251, 1984.
Gustafson, D.H., D.G. Fryback, J.H. Rose, V. Yick, C.T. Prokop and D.E. Detmer, A decision theoretic methodology for severity of illness development, Medical Decision Making 6, 1986.
Hadorn, D.C., E.B. Keeler, W.H. Rogers and R.H. Brook, Assessing the performance of mortality prediction models, RAND/UCLA/Harvard Center for Health Care Financing Policy Research, Santa Monica, CA, 1993.
Hannan, E.L. and T.W. Hartman, A critique of the 1987 HCFA mortality study based on New York State data, Department of Health, Office of Health Systems Management, Albany, NY, 1989.
Harrison, D.C., Cost containment in medicine: Why cardiology, American Journal of Cardiology 56, 1985, 5.
Horn, S.D., P.D. Sharkey, A.F. Chambers and R.A. Horn, Severity of illness within DRGs: Impact on prospective payment, American Journal of Public Health 75, 1985, 10.
Horn, S.D., R.A. Horn, P.D. Sharkey and A.F. Chambers, Severity of illness within DRGs: Homogeneity study, Medical Care 24, 1986, 3.
Knaus, W.A., E.A. Draper, D.P. Wagner and J.E. Zimmerman, APACHE II: A severity of disease classification system, Critical Care Medicine 13, 1985, 10.
Laffel, F. and D. Blumenthal, The case for using industrial quality management science in health care organizations, Journal of the American Medical Association 262, 1989, 20.
Pozen, M.W., R.B. D'Agostino, J.B. Mitchell, D.M. Rosenfeld, J.T. Guglielmino, M.L. Schwartz, N. Teebagy, J.M. Valentine and W.B. Hood, Jr., The usefulness of a predictive instrument to reduce inappropriate admission to coronary care unit, Annals of Internal Medicine 92, 1980, Pt 1.
Ramsey, L.P. and R.S. Cantrell, Investigating cost variances using control charts, Healthcare Financial Management, 1985.
Thomas, J.W. and D.R. Longo, Application of severity measurement systems for hospital quality management, Hospital and Health Services Administration 35, 1990, 2.
Weinstein, M.C., P.G. Coxson, L.W. Williams, T.M. Pass, W.B. Stason and L. Goldman, Forecasting coronary heart disease incidence, mortality, and most: The coronary heart disease policy model, American Journal of Public Health 77, 1987, 11.
Young, W.W., Measuring the cost of care using patient management categories: Final report of the Health Care Finance Administration, Baltimore, MD, Grant No. 18P97063/3, 1985.