USING QUALITY CONTROL TECHNIQUES TO MEASURE PERFORMANCE: A CASE STUDY OF A MIDWESTERN REHABILITATION INSTITUTE

Donna Neufelder, St. Mary's Rehabilitation Institute
Janet Raisor, St. Mary's Rehabilitation Institute
Mohammed Khayum, University of Southern Indiana
Dan Friesner, Weber State University

ABSTRACT

Over the past decade, health care providers have experienced increased pressure from consumers, insurers and regulatory organizations to demonstrate not only the fact that they provide high quality services, but also that they are continually working to improve upon the quality of those services. In particular, organizations such as the Joint Commission for the Accreditation of American Healthcare Organizations (JCAHO) have placed increased emphasis on the use of data-oriented methods of monitoring and improving an organization's performance. For many providers, particularly small and mid-sized providers, this presents two problems. First, small and mid-sized providers often have difficulty collecting the data necessary to document their process improvement activities. Secondly, even if these facilities are able to collect the appropriate data, a provider's staff often does not have the knowledge to appropriately apply the statistical techniques inherent in process improvement. The purpose of this paper is to present a case study that demonstrates how firms with limited resources and available data can employ some simple statistical techniques to measure performance and process improvement.

INTRODUCTION

In many ways, therapy is a science concerned with optimal performance. During each patient encounter, the objective is to strive for maximum results, whether the task at hand is teaching an 86-year-old stroke survivor to walk with a cane or doing passive range of motion exercises with patients to prevent contractures.

But in the pursuit of patient results, it is easy for practitioners to lose sight of the effectiveness of our own clinical approaches. As a result, many health care organizations are implementing performance improvement initiatives to quantify what they do and present it in a larger facility-wide context. The aggregated data helps practitioners make better clinical decisions as well as justify treatment plans.

Driving this movement further is demand from consumers, payers and regulatory agencies to prove clinical effectiveness and efficacy. From the regulatory perspective, the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) requires health care organizations
to assure quality in the delivery of care. In addition, gaining accreditation from the Commission on Accreditation of Rehabilitation Facilities (CARF) requires "information and outcomes management systems" and the use of comparative data. These movements drive the need to establish baseline measurements and then improve upon the processes.

There are two primary difficulties in measuring process improvement within a health care setting. The first is assimilating the data collection process into every day patient care tasks. Since the practitioner's primary objective is to treat his/her patient, therapists are limited in the amount of time they can spend collecting data. As a result, the lack of available data makes it difficult for facilities to conduct process improvement studies.

A second critical factor is data analysis. Many clinicians don't have the statistical skills to analyze complex data. Larger health care facilities can avoid this difficulty by adding bio-statisticians to its full-time staff or by hiring (part-time) consultants who specialize in productivity improvement. Similarly, teaching and research-oriented facilities may have clinicians who are trained in statistics and/or management science. However, smaller providers may not have the resources to hire consultants or attract research-oriented practitioners to their staff. Consequently, if these providers are to undertake process improvement studies, their staff must be able to complete the study using a set of relatively simple statistical tools.

Given these two difficulties, it becomes necessary for small and mid-sized health care providers (who are most likely to face data and skill limitations) to develop methods of measuring process improvement that do not require stringent data requirements or high powered statistical skills. The purpose of this paper is present some simple techniques that allow providers to measure process improvement given these constraints. We employ these techniques using data from a (mid-sized) Midwestern rehabilitation provider. As such, this

paper can be considered as a case study, whose findings may be applicable to a large number of health care providers that share both the mission and challenges faced by this institute.

The remainder of this paper proceeds in three steps. First, we describe the study's database and some of the limitations we encountered when using this data. Next, we describe some basic statistical techniques that we utilized to analyze the data. We conclude the paper by discussing the implications of our findings. In this section we also provide some suggestions for firms who face similar data and/or skill limitations.

Data

The data come from a major, nonprofit medical center in a medium sized city (with a population of approximately 130,000) in the Midwestern United States. The city serves as a regional health care center for a relatively large (approximately 80 miles in diameter) geographic area. It employs a range of specialized and general health care practitioners as well as a wide array of medical services, including physical therapy. The provider also experiences competition in almost all of its services from another, similarly sized (nonprofit) medical center that resides within the same city. The center offers physical therapy services on an inpatient basis at its 50 bed, acute care Rehabilitation Institute. Outpatient therapy services are offered at one of four different locations, which are strategically located throughout the city.[1] Most therapy sessions averaged 45 minutes in length, with a few sessions lasting as few as 30 minutes and as many as 60 minutes. Outpatients received one therapy session per day, while inpatients received therapy twice per day.

The data used in this study consist of all patients referred for inpatient physical therapy following total knee replacement surgery (TKR) during the fiscal year 2002. For each patient, data was collected on three different measures (both pre and post therapy) that the staff believed most efficiently characterized a patient's rehabilitation progress following TKR: knee extension (measured in degrees) while in a supine position and knee flexion (again, measured in degrees) in both a sitting and a supine position. The staff's a priori expectations were that, if a patient successfully completed rehabilitation following TKR, flexion should increase, while extension should decrease. As a result, the staff chose to focus on the difference between the pre and post measurements for each of these variables. For the flexion variables, the difference was created by subtracting the pre-TKR measurement from the post measurement (i.e. post – pre). The extension measurement was created in a reverse fashion (i.e., pre – post). As a result, a positive value for each of these differenced variables indicates an improvement in the patient's condition, while a negative value indicates a regression in the patient's condition.

The staff also chose to collect data on a fourth variable that could potentially impact rehabilitation following TKR. Specifically, the girth of the surgically repaired knee was measured (in inches) and compared to the girth of the patient's other (non-invasive) knee. The intuition behind this metric is that, immediately following surgery, the repaired knee experiences swelling, which may inhibit mobility and retard rehabilitation. Assuming symmetry, the difference between the two girth measurements (i.e., the TKR measurement – the non-invasive measurement) gives a normalized metric of the amount of swelling. As such, the girth difference provides a very rough measure of a patient's initial illness severity. Patients with a larger girth difference would subsequently be expected to take longer to heal, and thus require additional therapy.

Lastly, data was collected on a number of supporting variables. For example, at the conclusion of treatment, patients were asked to evaluate their perceived pain using a 0 (no pain) to 10 (maximum pain) rating scale. Other data included the physician who performed the TKR, the length of stay, age and sex. A total of 122 patients were included in this study; however, the staff was not able to collect a complete set of information for all patients in the study. As a result, there are some missing values. Table 1 contains the names and definitions of all relevant variables used in the analysis, while Table 2 presents some basic descriptive statistics for each of these variables.

Statistical Analysis

Our analysis of the data proceeds in two steps. Our first approach is to analyze the trends in the data using simple descriptive statistics, correlations and hypothesis tests. This approach is quite useful because of its computational ease as well as the fact that most practitioners are familiar with these basic tools, and can consequently interpret the findings within the context of the practice. A drawback to this approach is that it does not always give a clear and concise conclusion about whether or not the facility's production process is operating efficiently with respect to patient outcomes.

Our second approach applies control chart theory to the data in an effort to assess the effectiveness

[1] The provider also offers a limited number of outpatient physical therapy services at its two assisted living sites.

of the provider's rehabilitation process. The benefit to this approach is that it allows the provider to determine whether or not there is a fundamental problem with their practice methods, which may not be inferred from the descriptive statistics. These charts are also easily created using tools such as Microsoft Excel, and can be presented in a graphical format which can be interpreted quite easily by those with a limited statistical background. However, while control charts are fairly easy to read and interpret, creating control charts implicitly assumes a more detailed understanding of statistical methods. Additionally, since rehabilitation is a fundamentally slow process, the sampling methods used to create the control charts may not be directly applicable to medical processes. In this paper, we present a brief review of control chart theory, and also describe a simple method (again, which can be implemented in a "point and click" fashion using a spreadsheet application such as Excel) by which one can obtain the information necessary to construct the control chart[2].

Descriptive Statistics and Hypothesis Tests

Summary statistics for the original data used in this study are reported in Table 2. Of the 122 patients considered in the study, 70 percent are female and 30 percent are male. The mean age of patients is 70.1 years and the average length of stay is 7.6, with thirteen different physicians performing the knee replacement procedures[3]. The pre-treatment and post-treatment mean values of the three process performance measures are significantly different and in each case the difference coincides with performance improvement (see Table 3). On average, patients reduced extension by about 4.5 degrees and gained 22 and 25 degrees of additional sitting and supine flexion, respectively.

The average amount of girth difference in the data set is 3.6 inches, which is also statistically different from zero. It is important to note that there is some ambiguity concerning the minimum amount of swelling necessary to limit mobility. For example, a one-inch increase in knee girth may or may not be enough swelling to restrict mobility, and thus reduce the effectiveness of therapy. However, one can demonstrate that (at a 95% level of confidence) our average girth difference of 3.6 is statistically different from any postulated value less than three. So if three inches of swelling or less is enough to retard mobility, then we are 95% sure that, on average, patients are experiencing reduced mobility due to post-TKR swelling.

Examination of Table 2 also shows that the variances of the performance measures decreased between the pre-treatment and post-treatment situations. From a process-review perspective, these results are encouraging because they show that not only are patients improving after completing rehabilitation, but the dispersion among patients is also decreasing. That is, the patients are becoming more similar, and are (hopefully) experiencing a baseline range of motion and flexibility that are commensurate with a healthy, normal lifestyle.

Given the fact that three different process measures were utilized, and that each of the measures indicated patient improvement during rehabilitation, it also becomes interesting to examine correlations across these measures, which are contained in Table 4. Because our girth measurement may impact the effectiveness of rehabilitation, Table 4 also presents correlations between our girth and process improvement measures. Correlations varied among the performance measures between the pre-treatment and post-treatment conditions, although most of the results agree with conventional wisdom. Test results from the pre-treatment data indicate a strong negative correlation between supine flexion and supine extension, and a strong positive correlation between supine flexion and sitting flexion. These results make sense, particularly since (pre-rehabilitation) lower (greater) flexion and greater (lower) extension both coincide with a more (less) severe loss of motion. Additionally, while taken in different positions, the flexion measurements essentially measure much the same thing, and so should be positively correlated. In the post-treatment data, the test results also indicate a strong positive correlation between supine flexion and sitting flexion. Additionally, the post-treatment data showed a negative correlation between TKRGirth and the two flexion measures. This supports our a priori expectation that an increase in knee girth reduces therapy effectiveness.

[2] In what follows, we assume that the reader is familiar with the tools in a spreadsheet application such as Excel. As such, we will focus primarily on the steps in the application, and give secondary emphasis to the actual Excel commands and applications used to undertake the analysis. For those readers who are familiar with control chart theory, one may want to only briefly skim pages 9-13 of the manuscript, which review this theory. Normally, we would place such information in the appendix of the paper. However, because one of the premises of this paper is to *explain* as well as implement some techniques commonly used in process improvement, we felt that it would be more beneficial to leave this information in the body of the paper.

[3] Of the 122 patients and 13 physicians, 100 of the patients were treated by only 5 of the physicians.

We were also interested in determining whether there was a significant relationship between our (differenced) mobility and girth measures and our demographic variables. Tables 5 – 9 of the paper present some frequency tables[4] (with corresponding chi-square tests of independence) that address this issue[5]. Examining Tables 5 – 7, we find no significant evidence that our age, sex and physician variables are related to any of our four process variables. Not only are the frequencies evenly distributed throughout all of the categories, but the chi-square probabilities are all very high, indicating that we fail to reject the null hypothesis of independence (i.e., no significant relationship) between these variables. However, in Tables 8 and 9, we see that the length of stay is positively and significantly related to a patient's improvement in flexibility[6]. The trends in the table indicate that patients who completed a smaller length of stay (and hence a smaller number of therapy sessions) also increased their flexibility (whether in a supine or sitting position) by a smaller amount. Additionally, we see that patients with lower pain scores also exhibited a significantly greater increase in supine flexibility. These last two results are intuitive, because they indicate (but do not necessarily prove) that patients who work harder at rehabilitation make more progress in gaining flexibility. And as individuals gain greater flexibility and range of motion, they are able to rehabilitate, and so experience less pain from the knee injury.

Control Chart Analysis – A Brief Introduction

As mentioned earlier, descriptive statistics and basic hypothesis tests are quite useful in determining whether a production process is operating efficiently because they are relatively simple to implement and provide results that are easily interpreted (particularly by those with a limited statistical background) within the context of the practice. However, a drawback is that these methods may not provide a complete characterization of a firm's production process. As such, we turn to a slightly more technical alternative known as statistical process (or quality) control.

Statistical process control refers to the use of statistical techniques to detect changes in process performance and to identify assignable (special) causes of variation in process performance. By process we mean operations or a combination of procedures through which inputs are used to attain desired outcomes. Processes are typically repetitive and the outcomes can be measured and recorded. Control is defined as a feedback loop based on the following steps: (1) establishment of a standard, (2) measurement of actual performance, (3) comparison of actual performance with some standard, and (4) corrective action, if needed, addressing the discrepancy between actual and the standard (Alwan, 2000). In general, process control, therefore, can be defined as these steps when applied to a measurable process outcome. In this paper the process under consideration is the effectiveness of inpatient physical therapy on patients following TKR.

[4] Tables 5 – 9 present absolute frequencies. For readers who prefer to read and interpret relative frequencies, we have re-created these tables in a relative frequency format and placed them in the appendix of this paper.

[5] Our primary intent is merely to express, in a simple, complete and concise framework, some of the trends inherent in the data. As a result, our discussion focuses primarily on the trends in each of the frequency tables. We provide the chi-square statistic solely as a tangential piece of information to confirm or deny whether the results are significant. The chi-square tests operate under the null hypothesis of no relationship (or "independence") between the two variables of interest. Concomitantly, rejecting the null hypothesis is evidence that the two variables are significantly related. The test is one sided, with an upper bound, so that larger chi-square tests statistic values (and hence probability values below .05) would lead us to reject the null hypothesis (at a 95% level of confidence). For more information about the chi-square test of independence, see Anderson et al (2002) or Aczel (2002). Also note that we could have examined these trends on a marginal basis through the use of regression analysis. However, we chose not to utilize this approach because there are certainly other exogenous factors besides age, sex, physician, length of stay and perceived pain that influence our variables of interest. As such, it is inappropriate to utilize regression analysis because omitted variables would bias the results. Additionally, since one of our goals is to analyze the data in a simple, yet efficient manner, we felt that frequency tables would tell much the same story, yet do so in a much simpler manner.

[6] A technical aspect of the chi-square test should be noted here. Several of the entries in each of the frequency tables for the number of sessions contain values of two or less. This may artificially inflate the chi-square statistic so that the results appear to be significant, even though they are not. We checked for this possibility (by pooling categories so that there are very few entries with small frequencies and re-calculating the statistic) and did not find it to be a significant concern. Also, note that test statistic inflation is only a concern if we reject the null hypothesis (i.e., for the sessions variable), because if the test statistic is inflated, and we still fail to reject the null hypothesis, then we would also fail to reject it even if the test statistic were not inflated.

An underlying premise of statistical process control is that processes exhibit variability. Generally, this variability is assigned to one of two causes: *random causes* or *assignable causes*. Variation due to random causes reflects the natural variation inherent in every process. However, variation due to assignable causes can often be traced to some identifiable source or event. In this context, the main objective of statistical process control is to determine whether variations in process performance are due to assignable causes or to random causes. Whenever assignable causes are detected, the process is deemed to be out of control. Under such circumstances corrective action will be taken to bring the process back in line with an acceptable level of quality.

The primary focus in control chart applications is the detection of change in the level and/or dispersion of the process under investigation. The established convention is to construct two control charts, one to monitor the level of the process (often referred to as an $\bar{x}$ chart) and the other to monitor the dispersion of the process (known as an R chart). These charts are used to analyze and monitor individual measurements on a quality characteristic or on subgroups or samples taken on the quality characteristic. A starting point in standard control chart applications is that the quality characteristic, denoted by X, is a random variable with a population mean μ and standard deviation σ. Using standard statistical sampling theory, on could collect a random sample, or subgroup, of n observations on the characteristic X. Denote the subgroup mean as $\overline{X}$. Statistical theory shows that the sampling distribution of $\overline{X}$ has an expected value of μ and a standard deviation $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. Given general assumptions about n, the central limit theorem can be applied to subgroup means and be used to show that the sampling distribution of $\overline{X}$ is normal. As a result, one can use information from the random sample to build a confidence interval estimate around the population mean, μ. For example, the two-sided, 99.72% confidence interval is given by the following upper and lower bounds (or control limits):

$$\text{Upper Control Limit } (UCL) = \mu + 3\frac{\sigma}{\sqrt{n}} \qquad (1)$$

$$\text{Lower Control Limit } (LCL) = \mu - 3\frac{\sigma}{\sqrt{n}} \qquad (2)$$

The connection between a confidence interval and statistical process control lies in the interpretation of the confidence interval. If a process is "in control", or operating efficiently, then any variation in the outcomes of that process must be due to randomness. If that is the case, then (using the confidence interval above) we are 99.72% sure all outcomes from the production process fall within the confidence interval. However, if the production process is inefficient, then there is some assignable cause (i.e., inefficiency) that increases outcome variation, and thus causes the process' outcomes to lie outside of the confidence interval. Statistical control is the process of building these confidence intervals and examining them to determine whether or not the data are contained within the confidence interval. Most often, this is accomplished by creating a graph which shows every data point and its relation to the center line and each control limit. This graph is referred to as the *control chart*.

Since μ and σ $(\text{or } \sigma_{\overline{X}})$ are generally unknown, they have to be estimated. So denoting the estimates for μ and $\sigma_{\overline{X}}$ as $\hat{\mu}$ and $\hat{\sigma}_{\overline{X}}$ respectively, the empirical control limits become

$$UCL \cong \hat{\mu} + 3\hat{\sigma}_{\overline{X}} \qquad (3)$$

$$LCL \cong \hat{\mu} - 3\hat{\sigma}_{\overline{X}} \qquad (4)$$

In a production setting, these estimates are most commonly obtained through repeated random sampling. That is, one collects a random sample of size n, and uses that sample to calculate the sample mean, $\bar{x}$. This process is repeated m – 1 times. As a result, we are left with a collection of m sample means: $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$. The estimate for μ (which is also sometimes denoted as $\bar{\bar{x}}$) is then calculated as the average of these m values:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{m} \bar{x}_i}{m} \qquad (5)$$

This overall average $\bar{\bar{x}}$ is referred to as the *grand mean* or the *grand average*.

While there are several possible means through which the standard deviation of the sampling distribution can be estimated, the most common approach is to calculate the grand range. As one might infer, the grand range is calculated in a manner similar to the grand average. Specifically, one collects a random sample of size n, and uses that sample to calculate the sample range, R. This process is repeated m – 1 times. As a result, we are left with a collection of m sample ranges: $R_1, R_2, \ldots, R_m$. The grand range is then calculated as the average of these m values:

$$\overline{R} = \frac{\sum_{i=1}^{m} R_i}{m} \quad (6)$$

Statistical theory has shown that if R is multiplied by a factor, called A_2, which is a function of the subgroup size n, then a reasonable estimate of the term $3\sigma/\sqrt{n}$ found in (1) and (2) can be obtained[7]. Based on the above, the center-line (CL, or the estimate for the population mean) and control limits for the $\bar{x}$ chart for subgroup means are:

$$UCL = \bar{\bar{x}} + A_2\overline{R} \quad (7)$$

$$CL = \bar{\bar{x}} \quad (8)$$

$$LCL = \bar{\bar{x}} - A_2\overline{R} \quad (9)$$

The $\bar{x}$ chart focuses on detecting average changes in the process level. If one were interested in looking at process control from a variability standpoint, one could employ an R chart. The idea behind an R chart is virtually identical to that of a x chart, except now we are building the confidence interval in terms of dispersion, instead of averages. The centerline and control limits of the R chart are as follows:

$$UCL = D_4\overline{R} \quad (10)$$

$$CL = \overline{R} \quad (11)$$

$$LCL = D_3\overline{R} \quad (12)$$

where D_3 and D_4 are general constants based on number of observations in the sample used to determine the range. As with the constant A, most statistics and operations management textbooks contain tables whereby, if one has information on the sample size, n, the constants D_3 and D_4 can be obtained from those tables.

Control Charts – Technical Issues in Health Care

When employing control charts to measure process control, there are a number of technical concerns that must be addressed, particularly when employing these techniques in health care. First, one must determine an appropriate sample size (n) as well as an appropriate number of re-samplings (m). Aquilano et al (1995) suggest a sample size of 4 or 5 observations, and also suggest that at least 25 re-samplings should be utilized. Based on these guidelines, we chose to employ a sample size of 5 and a re-sampling size of 35.

An additional issue of concern when applying control chart theories in health care, and particularly in rehabilitation studies, is that the production process is very slow. In manufacturing or other related industries, it is not uncommon to produce hundreds or thousands of finished products per day. As such, it is relatively easy to collect a sample of n = 5 units, and then repeat this process n = 25 times, where each re-sampling constitutes a different "batch" or finished products (i.e., there is sampling *without* replacement). But rehabilitation is a lengthy process, where it may take weeks, or even months to create a finished product (i.e., rehabilitate an individual's injury). For example, this mid-sized provider completed only 122 inpatient TKR rehabilitations over an entire fiscal year. In these cases, it is unclear as to how to implement the re-sampling process. Our approach is to randomly select 5 individuals *with replacement* from the 122 in the data set. As a result, we randomly draw 5 individuals from the data set (for a given variable), find the sample average and sample range, and then place these individuals back into the data set before moving on to the next sampling. The benefit of this approach is that it allows us a reliable means to conduct the random re-sampling for the control chart analysis. The cost is that individuals may be collected and placed in multiple samples. Given that we are using the sample to find means and ranges, this should not substantially affect our results[8]. We repeat

[7] Statistical tables have been created that give the value of A for every possible sample size. These tables can be found in most introductory statistics and operations management textbooks.

[8] Several comments are in order here. First, random sampling with replacement is commonly used in situations where one attempts to collect a random sample from a small population or a very slow production process (Anderson et al, 2002). As a result, our approach is not without precedent. Second, most spreadsheet packages have built in applications that randomly select a sample (whether with or without replacement) from a data set. In our case, we used the "Sampling" option in Excel's Analysis Tool Pack macro. Third, there are alternative methodologies that have been used to perform control chart analyses of slow production processes. For example, Aczel (2002) provides a very brief discussion of how to use individual data, as opposed to sample average and ranges, to construct control charts. However, many of these approaches (including the one presented by Aczel) are ad hoc, and employ very unrealistic assumptions, particularly when calculating the control limits. We chose this approach because we believed it provided a method of collecting the data that was reliable, and could be easily implemented by practitioners. Lastly,

this process separately for each of our four process improvement variables. As a result, the control charts implicitly assume that these four factors are separate phenomenon[9].

Applying Control Charts to the Rehabilitation Data

Since all of the outcome indicators exhibit significant improvement after the physical therapy sessions, the control chart applications in this study examine process performance in terms of the difference between admission and discharge measurements for each of the indicators. In particular, our focus will be to identify the random sources of variation in process performance and to use those results to determine benchmarks for future performance.

Given the fact that our girth measurement was significantly correlated with our process improvement measures, it is also of interest to build control charts for our girth variable. However, our interpretation of the latter is slightly different than for our three outcome variables. Specifically, while excess variation in the girth measurement is indicative of an "out-of-control" process, it is not a process that is directly under the control of the facility's staff. Instead, it represents excess variability in an initial condition, which the staff must subsequently deal with. As such, it is still of interest to know if the girth difference is out of control, and staff may want to adjust their practices in accordance with the results.

Our initial step is to study the variability of the impacts associated with the inpatient physical therapy sessions. To do so, we construct $\bar{x}$ and R charts for each of the three performance indicators measured in terms of the difference between the pre-treatment and post-treatment data values, as described earlier in this paper. We also construct these same charts using the girth difference variable. The charts are based on the values associated with 35 randomly selected subgroups, each of size $n = 5$. In order to determine whether there is a shift in the process mean, the $\bar{x}$ chart is constructed based on (7), (8) and (9) for each of the four variables. With subgroup size $n = 5$, $\bar{R} = 11.114$, $\bar{\bar{x}} = 4.074$, and $A_2 = 0.577$, the centerline and control limits for the supine extension $\bar{x}$ chart are:

$$\text{UCL} = 4.074 + 0.577*11.114 = 10.487$$
$$\text{CL} = 4.074$$
$$\text{LCL} = 4.074 - 0.577*11.114 = -2.338$$

The corresponding $\bar{x}$ chart, which is a sequence plot of the subgroup means with the control limits superimposed, is shown in Figure 1. The control chart shows that the subgroup means vary well within the limits with no strong indications of special causes. Based on this finding, the subgroup means are in a state of statistical control for supine extension improvement, and the retrospective limits can be projected out for prospective control.

Using the data for the subgroups on supine extension, we are also able to calculate the bounds of the R-chart. For subgroup size $n = 5$, the values of D_3 and D_4 are 0 and 2.114, respectively. Using (10) - (12), we include the centerline and compute the control limits for the supine extension difference R chart as:

$$\text{UCL} = 2.114*11.114 = 23.496$$
$$\text{CL} = 11.114$$
$$\text{LCL} = 0*11.114 = 0$$

The R chart shown in Figure 2 indicates that approximately 54 percent of the supine extension difference values are above the center line, while all 35 data points fall within the upper and lower control limits. In addition, there is no evidence of a sustained shift in the dispersion of measurements of the change in supine extension between admission and discharge from a stable process mean.

R and $\bar{x}$ charts constructed for two of the other performance measures, namely, changes in supine flexion and in sitting flexion also indicate that the all data points vary within the relevant control limits, as evidenced by Figures 3-6. Additionally, the data are clustered relatively evenly around the center line, with approximately half of the data values above and below each center line, respectively. As such, the control charts indicate that the flexion variables are also in a state of statistical control.

However, the R and $\bar{x}$ charts for the girth difference (Figures 7 and 8) indicate a number of out-of-control points that are candidates as special-cause signals. Thus, the preliminary indication is that the process is out of control, and so the staff is treating patients whose girth measurements (and hence knee swelling after TKR) is significantly different than the norm. An examination of these out-of-control signals

---

because our data set (122 observations) is very large in comparison to the size of the sample we are selecting (5 observations) and the number of re-samplings (35), random sampling with replacement should not noticeably impact our results, since one would expect (but could not guarantee) that each data point would be drawn only 1 or 2 times.

[9] Note that, if one wanted to account for the possibility of an inter-relationship between these four variables, one could use a multivariate control chart (Alwan 2000). Since multivariate control charts are much more complicated than their univariate counterparts, we employ the latter for simplicity.

showed that they arose from the same subgroup data points (3rd, 24th, 32nd) in both the R and $\bar{x}$ charts. Moreover, a review of the within-subgroup observations for those data points indicated that the source of the process variability was due to a single, isolated outlier that (negatively) distorted both the mean and the range in each of these subgroups.[10] The out-of-control subgroups were subsequently discarded and the limits and centerlines for the R and $\bar{x}$ charts were recomputed. Inspection of the revised control charts (Figures 9 and 10) indicate that, without the presence of these 3 subgroup data points, the girth performance measure is in a state of statistical control.

Up to this point, control charts have been used retrospectively, in the sense of looking back on the performance of the physical therapy treatments. Based on the analysis, the process is found to be in statistical control for the various performance indicators. Consequently, these control limits can also be used to project into the future so that the process can be monitored and appropriately adjusted to ensure that the process can be maintained in a state of statistical control. That is, over time, as the practitioners increase the size of the existing data set, they can repeat the above analysis to continually check to see if the process is operating efficiently.

The use of control charts as a means of assuring quality standards and performance improvement generates added challenges for decision-makers. One set of challenges is related to control chart maintenance. In particular, because processes evolve over time, control limits established using historical data eventually become obsolete, and must be revised. As such, the staff must use logic and experience to determine how often these revisions take place. One important factor in this decision is whether there have been any discernible changes in process performance. For example, if the staff changed its methods of providing care in a way that reduced process variability, control limits would need to be revised to reflect the new process realities. As a general rule it is also good practice to consider characteristics such as the speed of the production process when determining the frequency of control limit revisions.

Evaluation of Future Process Performance

A key challenge in process evaluation is to determine whether future variability in performance is due to random or assignable causes. There is interest, therefore, in extending control chart theory to create *process capability intervals* for a random process. In this section we explore the construction of process capability intervals based on the idea that a random process occurs with a certain probability.

The primary difference between process control and process capability is how one defines (and interprets) the acceptable upper and lower limits of the process. Process control uses a confidence interval approach to define the upper and lower limits, and interprets the results in the context of actual process performance. Alternatively, process capability examines how well a process performs in relation to limits that are constructed based on an organization's goals, customer feedback, industry standards or benchmarks. That is, process capability establishes limits based not on actual performance, but how well the process should be performing. It is possible that process capability (or lack of) does not imply statistical behavior in terms of being in control or out of control, particularly if the process capability limits are larger than the control limits. In that case, for example, a process can be within the capability limits but not necessarily within control limits. The alternative, of course, is where the capability limits are more restrictive than the control limits. In this case, a process can be in control but not within the capability limits. In either event, the worse case scenario would be where a process is neither in control nor capable. This would call for immediate steps to stabilize and improve the process by identifying and eliminating the effects of underlying special causes.

A prerequisite for conducting process capability analyses is that process behavior, at least historically, is in a state of statistical control, so that special causes are not detected from past control charts and process behavior is random. Given that process outcomes are normally distributed, lower specification limits (LSL) and upper specification limits (USL), represent intervals within which observed outcomes are likely to fall due to random causes. Thus, given a relatively small probability, α, the following statement can be made: there is a $(1-\alpha)$ probability that an observed process outcome will randomly fall within the specified limits. Within this framework, a major purpose of process capability analysis is to monitor process performance in relation to desirable outcomes and to determine the likelihood that the process can meet certain quality requirements. Therefore, by estimating the specification limits and discussing the results in terms of probabilities (as opposed to actualities), process capability analyses

[10] As a follow-up, a process investigation is necessary to determine the root cause of the outlier and to decide whether any process changes are warranted.

provide a basis for making predictions about future performance.

In the context of this paper, the specification limits for a performance measure is the confidence interval $\bar{\bar{x}} \pm z_{\alpha/2}\hat{\sigma}_{\bar{x}}$ where $(1-\alpha)$ is the confidence coefficient and $z_{\alpha/2}$ is the z value providing an area $\alpha/2$ in the upper tail of the standard normal probability distribution. The idea behind process capability is that the staff uses its goals, experiences and common sense to pick a value for α, which in turn determines the z value and the upper and lower limits of the confidence interval. Then one collects data and examines the proportion of data values that fall within the confidence interval. If this proportion is greater than or equal to than 1 - α, the process is within its capabilities. However, if this is not the case, then the process needs to be adjusted.

As with standard control charts, we need to establish a means of estimating the bounds of the confidence interval. Statistical theory shows that the following formulas can be applied to generate these bounds:

$$UCL = \bar{\bar{x}} + z_{\alpha/2}\left(\frac{\bar{R}}{d_2\sqrt{n}}\right) \quad (13)$$

$$LCL = \bar{\bar{x}} - z_{\alpha/2}\left(\frac{\bar{R}}{d_2\sqrt{n}}\right) \quad (14)$$

Similarly, if one wanted to compute the process capability equivalent of the R chart, the estimated limits are given by:

$$UCL = D_{1-\alpha/2}\left(\frac{\bar{R}}{d_2}\right) \quad (15)$$

$$LCL = D_{\alpha/2}\left(\frac{\bar{R}}{d_2}\right) \quad (16)$$

where d_2 = 2.326 for a subgroup size of $n = 5$ and pairs of values for

$(D_{0.025}, D_{0.975}), (D_{0.005}, D_{0.995}), (D_{0.001}, D_{0.999})$ for $3 \le n \le 10$ are given as

$(0.85,4.20),(0.55,4.89),(0.37,5.48)$ respectively.

Table 10 provides and example of control limit estimates and specification limits for the four variables used in this paper, given α = 0.002. Inspection of Table 10 indicates that the use of specification limits narrows the boundaries for evaluating future performance. However, all data points for each of the three performance measures are well within the upper and lower bounds. Thus, the proportion of data within the bounds is one which is greater than 1 - α = 0.998. Therefore, the processes are also in capability control.

It is important to note that the control chart application used in this paper has some limitations that are found in standard subgroup charts. There is the danger that basing the analysis only on subgroup statistics lead to control limits being too wide or too narrow since the control limits are based only on within-subgroup variation. Therefore, it is important to check whether within-subgroup variation is significantly different from the between-subgroup variation. The typical cause of this inconsistency is the presence of nonrandom variability within the subgroups, which can only be discovered through an analysis of the individual observations.

Conclusions and Suggestions for Facilities with Similar Problems

The focus of this paper is to present a case study that demonstrates how firms with limited resources and available data can employ some simple statistical techniques to measure performance and process improvement. We focus on the use of statistical process control charts and nonparametric tests to determine whether past performance was in control, and also to identify some potential areas for improvement. We implemented these tools using data from a Midwestern rehabilitation provider, and tracked the progress of patients who received physical therapy after inpatient, total knee replacement surgery (TKR). Control chart applications related to four measures of performance demonstrated that the rehabilitation process was largely effective in terms of providing quality care. Given the in-control characteristics of these performance measures, probability limits were generated based on estimates of the process mean and standard deviation to provide benchmarks for evaluating future performance.

Based on the methods and results of this paper, a set of procedures is suggested for the ongoing evaluation of physical therapy treatments of patients who undergo total knee replacement surgery (TKR). First, rehabilitation units that provide physical therapy services need to gather relevant information to track outcomes using a standardized data collection process. A critical consideration is to make the data collection process integral to the treatment session and to the evaluation of rehabilitation progress. The benefit of this approach is that the staff can easily retrieve and aggregate data without having to go back to do extensive chart reviews at another time. However, implementing

this type of arrangement requires effective communication with the staff involved in the treatment sessions. That is, the staff needs to be informed about the purpose and the scope of the data collection process. In addition, focusing on the long-term benefits to patients and to the success of the organization can help provide greater understanding of the need for these efforts. Staff training can also be very useful. The formation of a performance improvement team, which acts as a resource and training unit, allows for consistent administration of the data collection instruments. The data then needs to be compiled regularly (e.g. quarterly) and the results shared at departmental meetings.

A critical success factor, which can be a stumbling block for many facilities, is data analysis. Practitioners at small or mid-sized institutions who do not have access to an on-site data analyst may have to find creative methods of analyzing the data. One possibility is to analyze the data using some simple, yet powerful statistical tools that can be employed by the practitioners themselves. Another possibility is that there may be faculty in local universities who specialize in applied statistical analysis, and might be interested in assisting the staff with the data analysis. The appeal to a faculty member would be the potential for publishing research papers and working on community service projects that are consistent with the university's mission.

As clinicians, it is easy to lose sight of the larger picture in the daily tasks of helping patients achieve their functional goals. But by undertaking performance and quality improvement studies, it is possible to create benchmarks that can be used to highlight strengths and weaknesses. And by acting to correct those weaknesses and improve upon those strengths, it is possible to continually provide optimal, and efficient care.

TABLE 1: Variable Definitions

Variable	Definition
NIGirth	Girth measurement (in inches) for a patient's non-invasive knee.
TKRGirth	Girth measurement (in inches) for a patient's surgically repaired knee.
DifGirth	Difference between the TKRGirth and the NIGirth measurements.
PreSupExt	Supine extension measurement (in degrees) upon admission to therapy.
PostSupExt	Supine extension measurement (in degrees) upon completion of therapy.
DifSupExt	Difference between pre and post-therapy extension measurements.
PreSupFlex	Supine flexion measurement (in degrees) upon admission to therapy.
PostSupFlex	Supine flexion measurement (in degrees) upon completion of therapy.
DifSupFlex	Difference between post and pre-therapy supine flexion measurements.
PreSitFlex	Sitting flexion measurement (in degrees) upon admission to therapy.
PostSitFlex	Sitting flexion measurement (in degrees) upon completion of therapy.
DifSitFlex	Difference between post and pre-therapy sitting flexion measurements.
Phys	Proxy variable indicating the physician who performed the TKR.
Sex	Takes a value of 1 if the subject is female and 0 if the subject is male.
Age	The age of each patient (in years).
DCscore	Patient pain perception score upon exiting treatment.
Los	The length of stay for each patient.

TABLE 2a: Descriptive Statistics

Variable	Mean	Standard Deviation	Number of Observations
NIGirth	42.26	4.56	109
TKRGirth	45.96	5.15	114
DifGirth	3.60	3.52	106
PreSupExt	8.31	4.98	121
PostSupExt	3.73	3.76	122
DifSupExt	4.59	4.44	121
PreSupFlex	64.57	16.07	121
PostSupFlex	89.97	10.56	122
DifSupFlex	25.26	13.49	121
PreSitFlex	70.99	13.17	122
PostSitFlex	93.00	10.20	122
DifSitFlex	22.01	10.04	122
Sex	0.70	0.46	122
Age	70.14	8.72	122
DCscore	2.87	2.10	118
Los	7.61	3.43	122

TABLE 2b: Frequency Table for Physicians

MD Indicator	Frequency of TKR's
A	32
B	24
C	13
D	18
E	13
All Others	22

TABLE 3: T-Tests for Significance of the Difference Variables

Variable	Mean	Standard Error	T-Ratio	Prob.
DifGirth	3.60	0.342	10.52	0.000
DifSupExt	4.59	0.404	11.36	0.000
DifSupFlex	25.26	1.23	20.60	0.000
DifSitFlex	22.01	0.909	24.21	0.000

TABLE 4: Paired Sample Pearson Correlations

Variable 1	Variable 2	Pearson Correlation	Prob.	No. Observations
NIGirth	TKRGirth	0.730	0.000	106
PreSupExt	PostSupExt	0.515	0.000	121
PreSupFlex	PostSupFlex	0.553	0.000	121
PreSitFlex	PostSitFlex	0.657	0.000	122
NIGirth	PreSupExt	0.053	0.583	108
NIGirth	PreSupFlex	0.038	0.694	108
NIGirth	PreSitFlex	0.058	0.548	109
TKRGirth	PreSupExt	0.090	0.340	114
TKRGirth	PreSupFlex	-0.030	0.749	113
TKRGirth	PreSitFlex	-0.077	0.418	114
PreSupExt	PreSupFlex	-0.287	0.001	120
PreSupExt	PreSitFlex	-0.120	0.188	121
PreSupFlex	PreSitFlex	0.745	0.000	121
NIGirth	PostSupExt	0.053	0.581	109
NIGirth	PostSupFlex	-0.090	0.351	109
NIGirth	PostSitFlex	-0.076	0.431	109
TKRGirth	PostSupExt	-0.019	0.844	114
TKRGirth	PostSupFlex	-0.187	0.046	114
TKRGirth	PostSitFlex	-0.196	0.037	114
PostSupExt	PostSupFlex	-0.088	0.337	122
PostSupExt	PostSitFlex	-0.060	0.511	122
PostSupFlex	PostSitFlex	0.887	0.000	122

TABLE 5: Frequency Tables for the Age Variable and the Outcome Variables

		Age				
		under 60	60's	70's	80+	Total
	0 or negative	1	9	2	0	12
Girth	0.1 - 2.99 inches	3	5	14	4	26
Difference	3.0 - 4.99 inches	4	9	11	6	30
	5.0 + inches	5	12	17	4	38
	Total	13	35	44	14	106

Chi-Square Statistic	14.159
Probability	0.117

		Age				
		under 60	60's	70's	80+	Total
Sitting	< 15 Degrees	2	12	9	5	28
Flexion	15-29 Degrees	10	18	30	9	67
Difference	30+ Degrees	3	7	14	3	27
	Total	15	37	53	17	122

Chi-Square Statistic	4.665
Probability	0.587

		Age				
		under 60	60's	70's	80+	Total
	<= 0 Degrees	4	2	13	5	24
Supine	1 - 4 Degrees	5	16	13	4	38
Extension	5 - 8 Degrees	2	12	19	3	36
Difference	9+ Degrees	4	7	7	5	23
	Total	15	37	52	17	121

Chi-Square Statistic	13.770
Probability	0.131

		Age				
		under 60	60's	70's	80+	Total
Supine	< 15 Degrees	2	9	10	2	23
Flexibion	15-29 Degrees	11	17	24	8	60
Difference	30+ Degrees	2	11	19	6	38
	Total	15	37	53	16	121

Chi-Square Statistic	5.28
Probability	0.508

TABLE 6: Frequency Tables for the Sex Variable and the Outcome Variables

		Sex		
		Female	Male	Total
	0 or negative	9	3	12
Girth	0.1 - 2.99 inches	21	5	26
Difference	3.0 - 4.99 inches	18	12	30
	5.0 + inches	29	9	38
	Total	77	29	106
Chi-Square Statistic		3.568		
Probability		0.312		

		Sex		
		Female	Male	Total
Sitting	< 15 Degrees	21	7	28
Flexion	15-29 Degrees	43	24	67
Difference	30+ Degrees	22	5	27
	Total	86	36	122
Chi-Square Statistic		3.125		
Probability		0.21		

		Sex		
		Female	Male	Total
	<= 0 Degrees	18	6	24
Supine	1 - 4 Degrees	25	13	38
Extension	5 - 8 Degrees	27	9	36
Difference	9+ Degrees	15	8	23
	Total	85	36	121
Chi-Square Statistic		1.288		
Probability		0.732		

		Sex		
		Female	Male	Total
Supine	< 15 Degrees	18	5	23
Flexion	15-29 Degrees	43	17	60
Difference	30+ Degrees	24	14	38
	Total	85	36	121
Chi-Square Statistic		1.678		
Probability		0.432		

TABLE 7: Frequency Tables for the Physician Variable and the Outcome Variables

		Physician						
		A	B	C	D	E	Others	Total
	0 or negative	1	3	1	2	3	2	12
Girth	0.1 - 2.99 inches	12	5	2	4	0	3	26
Difference	3.0 - 4.99 inches	6	8	4	3	2	7	30
	5.0 + inches	10	7	5	6	3	7	38
	Total	29	23	12	15	8	19	106

Chi-Square Statistic	15.29
Probability	0.431

		Physician						
		A	B	C	D	E	Others	Total
Sitting	< 15 Degrees	9	7	3	4	3	2	28
Flexion	15-29 Degrees	17	10	8	11	7	14	67
Difference	30+ Degrees	6	7	2	3	3	6	27
	Total	32	24	13	18	13	22	122

Chi-Square Statistic	5.426
Probability	0.861

		Physician						
		A	B	C	D	E	Others	Total
	<= 0 Degrees	7	2	3	3	3	6	24
Supine	1 - 4 Degrees	5	11	7	4	4	7	38
Extension	5 - 8 Degrees	15	7	3	4	2	5	36
Difference	9+ Degrees	5	4	0	6	4	4	23
	Total	32	24	13	17	13	22	121

Chi-Square Statistic	19.910
Probability	0.175

		Physician						
		A	B	C	D	E	Others	Total
Supine	< 15 Degrees	9	6	1	3	2	2	23
Flexion	15-29 Degrees	11	14	10	6	7	12	60
Difference	30+ Degrees	12	4	1	9	4	8	38
	Total	32	24	12	18	13	22	121

Chi-Square Statistic	15.82
Probability	0.105

TABLE 8: Frequency Tables for the Length of Stay Variable and the Outcome Variables

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
	0 or negative	4	8	0	0	12
Girth	0.1 - 2.99 inches	8	16	1	1	26
Difference	3.0 - 4.99 inches	8	18	4	0	30
	5.0 + inches	14	19	2	3	38
	Total	34	61	7	4	106

Chi-Square Statistic	7.764
Probability	0.558

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
Sitting	< 15 Degrees	16	10	2	0	28
Flexion	15-29 Degrees	16	47	2	2	67
Difference	30+ Degrees	3	12	9	3	27
	Total	35	69	13	5	122

Chi-Square Statistic	38.034
Probability	0.000

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
	<= 0 Degrees	9	12	3	0	24
Supine	1 - 4 Degrees	8	27	3	0	38
Extension	5 - 8 Degrees	11	17	5	3	36
Difference	9+ Degrees	7	12	2	2	23
	Total	35	68	13	5	121

Chi-Square Statistic	9.750
Probability	0.371

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
Supine	< 15 Degrees	11	12	0	0	23
Flexion	15-29 Degrees	17	37	5	1	60
Difference	30+ Degrees	6	20	8	4	38
	Total	34	69	13	5	121

Chi-Square Statistic	17.833
Probability	0.007

TABLE 9: Frequency Tables for the Pain Variable and the Outcome Variables

		DCScore (0 -10 basis)		
		0-4	5+	Total
	0 or negative	7	2	9
Girth	0.1 - 2.99 inches	19	6	25
Difference	3.0 - 4.99 inches	26	4	30
	5.0 + inches	25	13	38
	Total	77	25	102

Chi-Square Statistic	3.987
Probability	0.263

		DCScore (0 -10 basis)		
		0-4	5+	Total
Sitting	< 15 Degrees	17	9	26
Flexion	15-29 Degrees	49	17	66
Difference	30+ Degrees	23	3	26
	Total	89	29	118

Chi-Square Statistic	3.848
Probability	0.146

		DCScore (0 -10 basis)		
		0-4	5+	Total
	<= 0 Degrees	18	6	24
Supine	1 - 4 Degrees	24	11	35
Extension	5 - 8 Degrees	29	6	35
Difference	9+ Degrees	18	5	23
	Total	89	28	117

Chi-Square Statistic	2.042
Probability	0.564

		DCScore (0 -10 basis)		
		0-4	5+	Total
Supine	< 15 Degrees	12	10	22
Flexion	15-29 Degrees	43	15	58
Difference	30+ Degrees	33	4	37
	Total	88	29	117

Chi-Square Statistic	8.953
Probability	0.011

TABLE 10a: X-Bar Control Limits and Specification Limits

of the Difference Variables

Variable	UCL	LCL	USL	LSL
DifGirth	7.909	-1.158	8.042	-1.292
DifSupExt	10.487	-2.338	10.677	-2.529
DifSupFlex	38.697	6.583	39.173	6.107
DifSitFlex	34.460	8.248	34.848	7.859

TABLE 10b: R Control Limits and Specification Limits

of the Difference Variables

Variable	UCL	LCL	USL	LSL
DifGirth	16.610	0	18.510	1.249
DifSupExt	23.496	0	26.184	1.768
DifSupFlex	58.830	0	65.562	4.427
DifSitFlex	48.018	0	53.514	3.613

Figure 1: X-Bar Chart for Supine Extension Difference

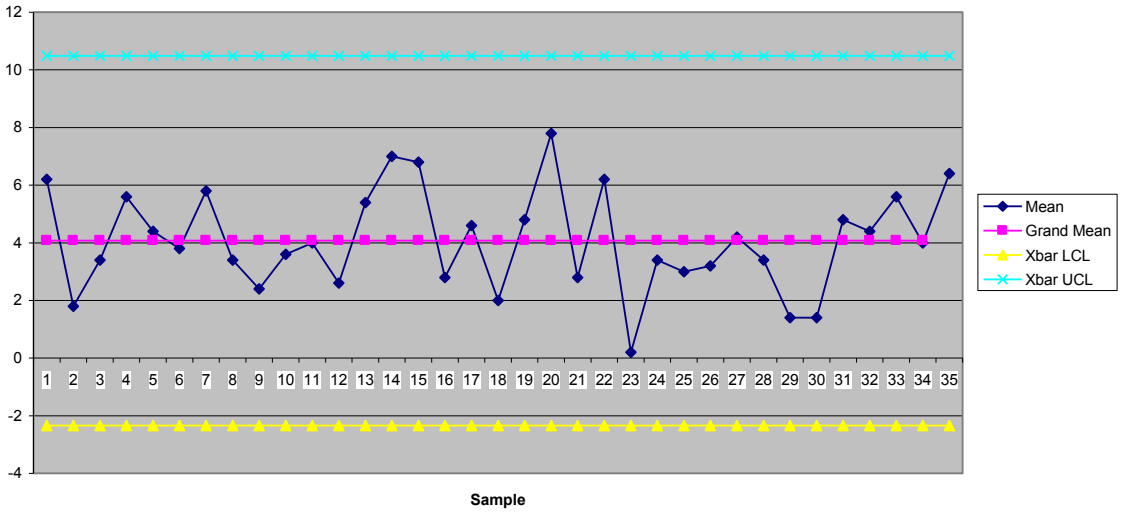


Figure 2: R-Chart for Supine Extension Difference

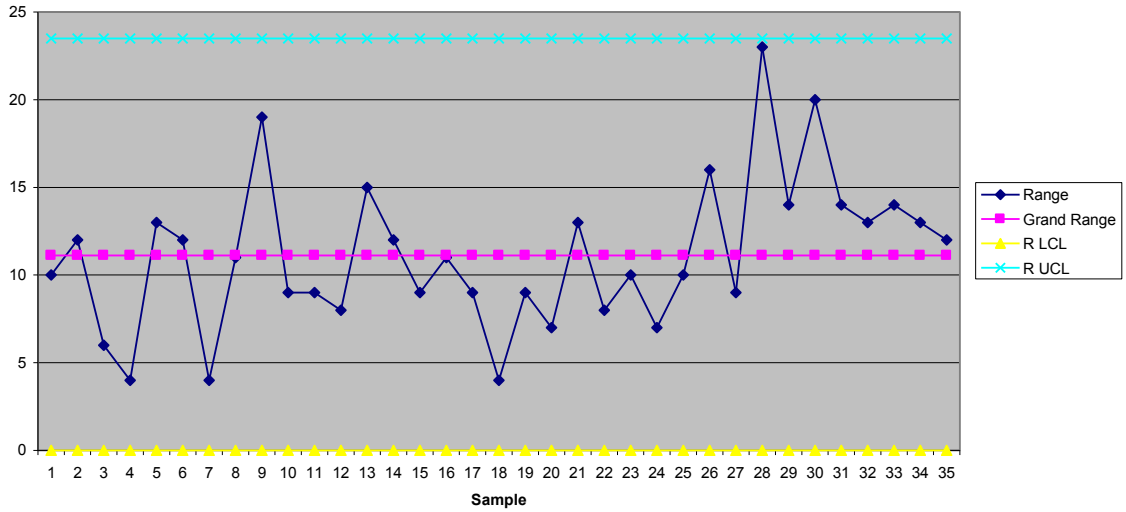


Figure 3: X-bar Chart for Supine Flexion Difference

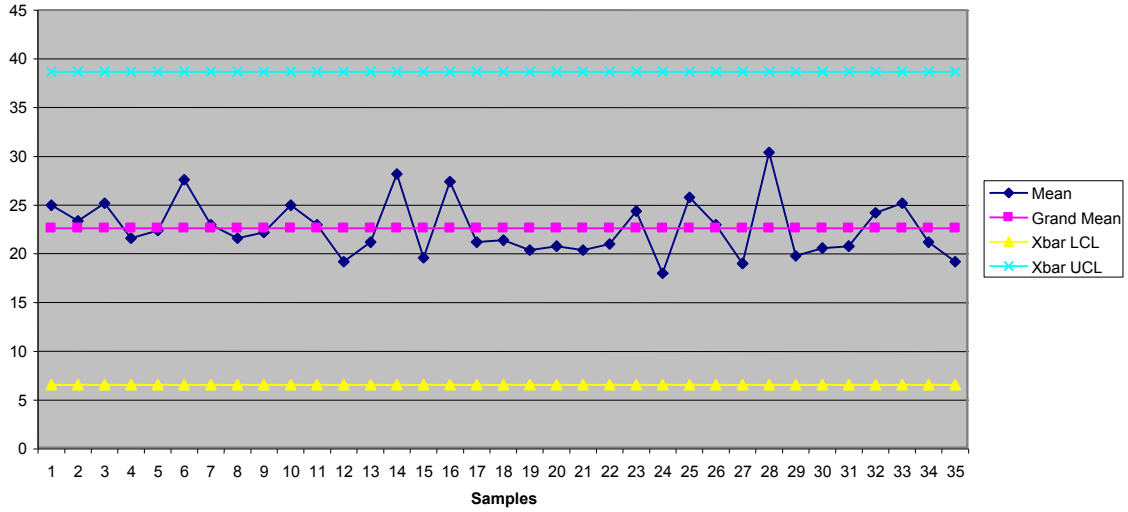


Figure 4: R-Chart for Supine Flexion Differences

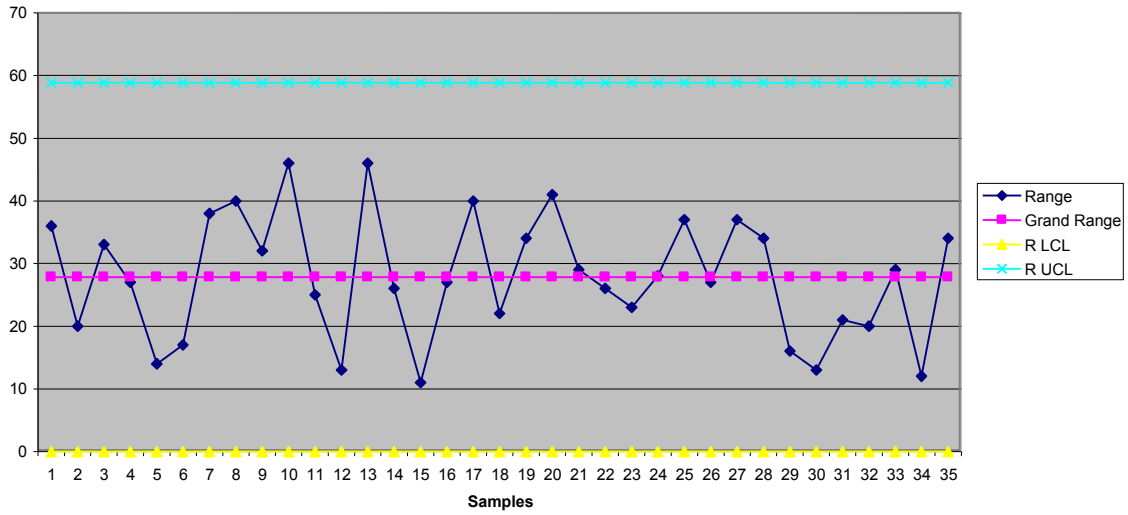


Figure 5: X-bar Chart for Sitting Flexion Differences

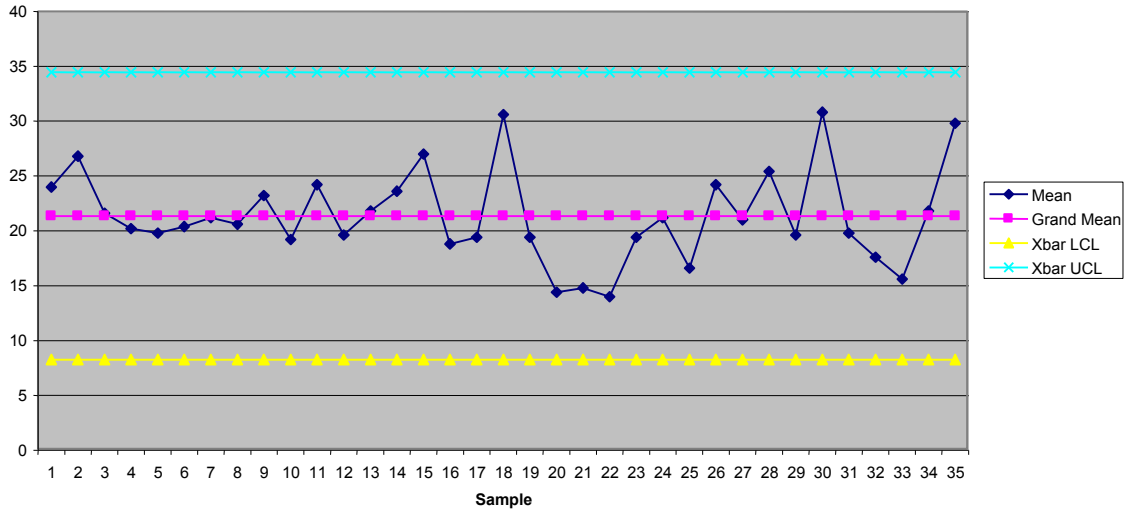


Figure 6: R-Chart for Sitting Flexion Differences

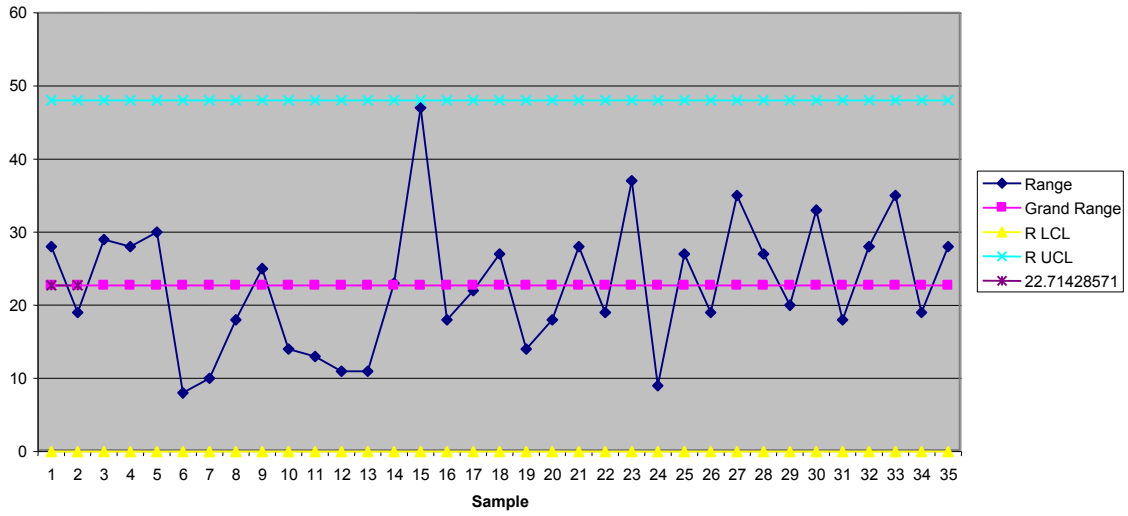


Figure 7: X-bar Chart for Girth Differences

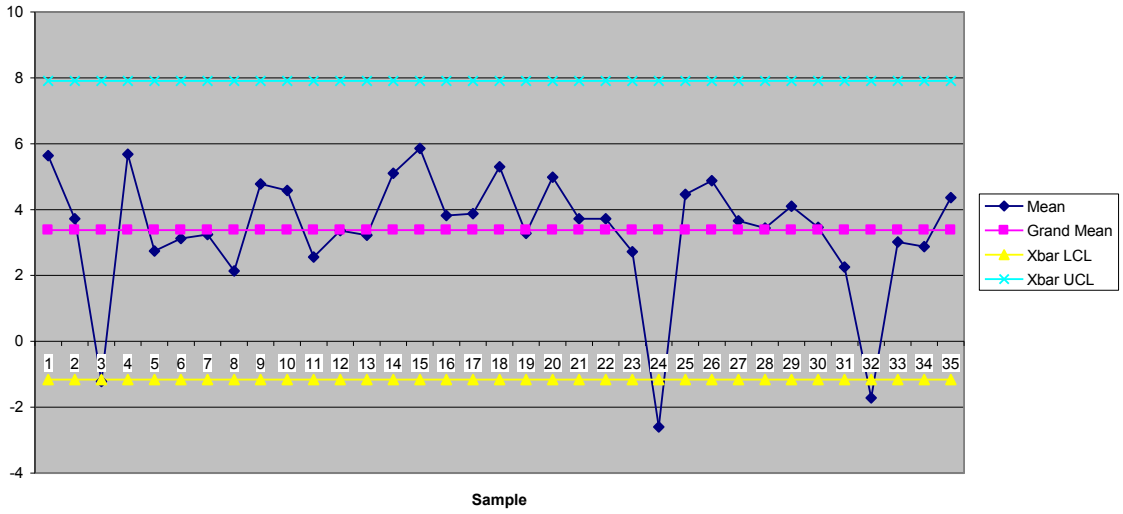


Figure 8: R-Chart for Girth Differences

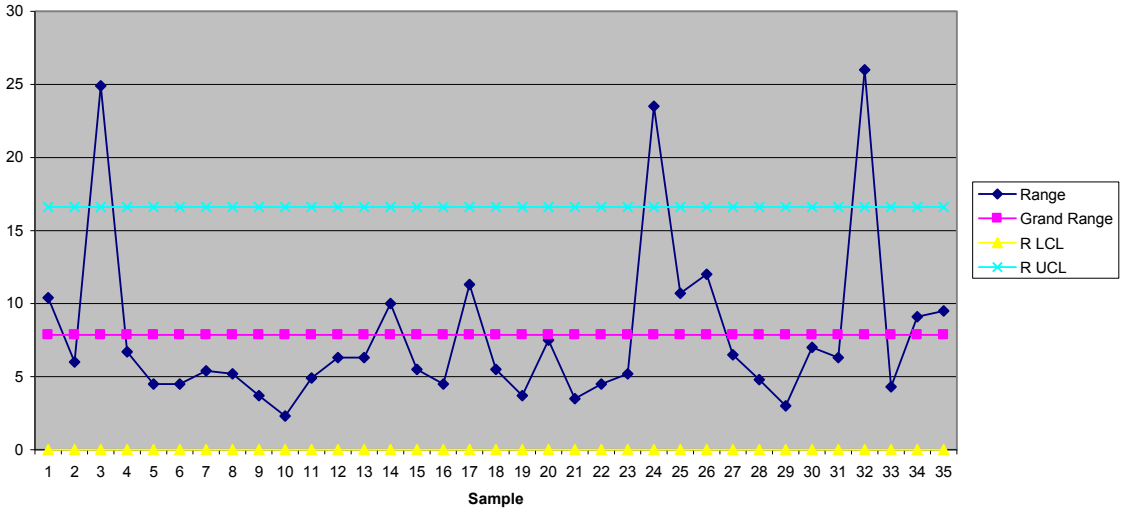


Figure 9: Revised X-Bar Chart for Girth

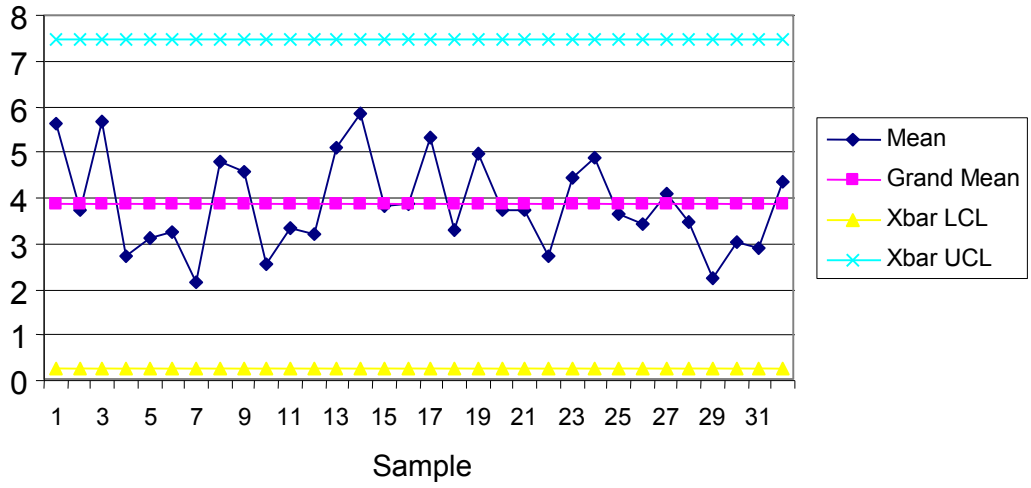
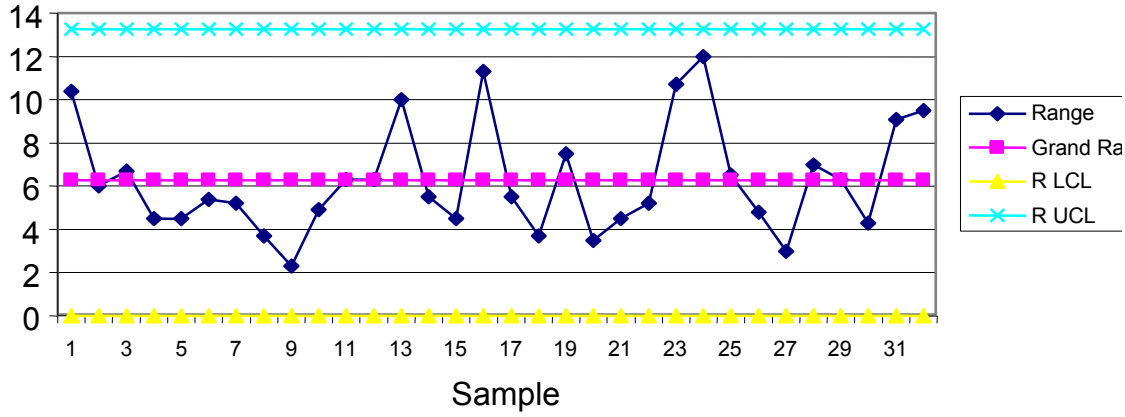


Figure 10: Revised R-Chart for Girth Difference



Appendix: Relative Frequency Tables

TABLE 5a: Frequency Tables for the Age Variable and the Outcome Variables

		Age				
		under 60	60's	70's	80+	Total
	0 or negative	0.0094	0.0849	0.0189	0.0000	0.1132
Girth	0.1 - 2.99 inches	0.0283	0.0472	0.1321	0.0377	0.2453
Difference	3.0 - 4.99 inches	0.0377	0.0849	0.1038	0.0566	0.2830
	5.0 + inches	0.0472	0.1132	0.1604	0.0377	0.3585
	Total	0.1226	0.3302	0.4151	0.1321	1.0000
Chi-Square Statistic		14.159				
Probability		0.117				

		Age				
		under 60	60's	70's	80+	Total
Sitting	< 15 Degrees	0.0164	0.0984	0.0738	0.0410	0.2295
Flexion	15-29 Degrees	0.0820	0.1475	0.2459	0.0738	0.5492
Difference	30+ Degrees	0.0246	0.0574	0.1148	0.0246	0.2213
	Total	0.1230	0.3033	0.4344	0.1393	1.0000
Chi-Square Statistic		4.665				
Probability		0.587				

		Age				
		under 60	60's	70's	80+	Total
	<= 0 Degrees	0.0331	0.0165	0.1074	0.0413	0.1983
Supine	1 - 4 Degrees	0.0413	0.1322	0.1074	0.0331	0.3140
Extension	5 - 8 Degrees	0.0165	0.0992	0.1570	0.0248	0.2975
Difference	9+ Degrees	0.0331	0.0579	0.0579	0.0413	0.1901
	Total	0.1240	0.3058	0.4298	0.1405	1.0000
Chi-Square Statistic		13.770				
Probability		0.131				

		Age				
		under 60	60's	70's	80+	Total
Supine	< 15 Degrees	0.0165	0.0744	0.0826	0.0165	0.1901
Flexion	15-29 Degrees	0.0909	0.1405	0.1983	0.0661	0.4959
Difference	30+ Degrees	0.0165	0.0909	0.1570	0.0496	0.3140
	Total	0.1240	0.3058	0.4380	0.1322	1.0000
Chi-Square Statistic		5.28				
Probability		0.508				

TABLE 6a: Frequency Tables for the Sex Variable and the Outcome Variables

		Sex		
		Female	Male	Total
	0 or negative	0.0849	0.0283	0.1132
Girth	0.1 - 2.99 inches	0.1981	0.0472	0.2453
Difference	3.0 - 4.99 inches	0.1698	0.1132	0.2830
	5.0 + inches	0.2736	0.0849	0.3585
	Total	0.7264	0.2736	1.0000
Chi-Square Statistic		3.568		
Probability		0.312		

		Sex		
		Female	Male	Total
Sitting	< 15 Degrees	0.1721	0.0574	0.2295
Flexion	15-29 Degrees	0.3525	0.1967	0.5492
Difference	30+ Degrees	0.1803	0.0410	0.2213
	Total	0.7049	0.2951	1.0000
Chi-Square Statistic		3.125		
Probability		0.21		

		Sex		
		Female	Male	Total
	<= 0 Degrees	0.1488	0.0496	0.1983
Supine	1 - 4 Degrees	0.2066	0.1074	0.3140
Extension	5 - 8 Degrees	0.2231	0.0744	0.2975
Difference	9+ Degrees	0.1240	0.0661	0.1901
	Total	0.7025	0.2975	1.0000
Chi-Square Statistic		1.288		
Probability		0.732		

		Sex		
		Female	Male	Total
Supine	< 15 Degrees	0.1488	0.0413	0.1901
Flexion	15-29 Degrees	0.3554	0.1405	0.4959
Difference	30+ Degrees	0.1983	0.1157	0.3140
	Total	0.7025	0.2975	1.0000
Chi-Square Statistic		1.678		
Probability		0.432		

TABLE 7a: Frequency Tables for the Physician Variable and the Outcome Variables

		Physician						
		A	B	C	D	E	Others	Total
	0 or negative	0.0094	0.0283	0.0094	0.0189	0.0283	0.0189	0.1132
Girth	0.1 - 2.99 inches	0.1132	0.0472	0.0189	0.0377	0.0000	0.0283	0.2453
Difference	3.0 - 4.99 inches	0.0566	0.0755	0.0377	0.0283	0.0189	0.0660	0.2830
	5.0 + inches	0.0943	0.0660	0.0472	0.0566	0.0283	0.0660	0.3585
	Total	0.2736	0.2170	0.1132	0.1415	0.0755	0.1792	1.0000

Chi-Square Statistic	15.29
Probability	0.431

		Physician						
		A	B	C	D	E	Others	Total
Sitting	< 15 Degrees	0.0738	0.0574	0.0246	0.0328	0.0246	0.0164	0.2295
Flexion	15-29 Degrees	0.1393	0.0820	0.0656	0.0902	0.0574	0.1148	0.5492
Difference	30+ Degrees	0.0492	0.0574	0.0164	0.0246	0.0246	0.0492	0.2213
	Total	0.2623	0.1967	0.1066	0.1475	0.1066	0.1803	1.0000

Chi-Square Statistic	5.426
Probability	0.861

		Physician						
		A	B	C	D	E	Others	Total
	<= 0 Degrees	0.0579	0.0165	0.0248	0.0248	0.0248	0.0496	0.1983
Supine	1 - 4 Degrees	0.0413	0.0909	0.0579	0.0331	0.0331	0.0579	0.3140
Extension	5 - 8 Degrees	0.1240	0.0579	0.0248	0.0331	0.0165	0.0413	0.2975
Difference	9+ Degrees	0.0413	0.0331	0.0000	0.0496	0.0331	0.0331	0.1901
	Total	0.2645	0.1983	0.1074	0.1405	0.1074	0.1818	1.0000

Chi-Square Statistic	19.910
Probability	0.175

		Physician						
		A	B	C	D	E	Others	Total
Supine	< 15 Degrees	0.0744	0.0496	0.0083	0.0248	0.0165	0.0165	0.1901
Flexion	15-29 Degrees	0.0909	0.1157	0.0826	0.0496	0.0579	0.0992	0.4959
Difference	30+ Degrees	0.0992	0.0331	0.0083	0.0744	0.0331	0.0661	0.3140
	Total	0.2645	0.1983	0.0992	0.1488	0.1074	0.1818	1.0000

Chi-Square Statistic	15.82
Probability	0.105

TABLE 8a: Frequency Tables for the Length of Stay Variable and the Outcome Variables

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
	0 or negative	0.0377	0.0755	0.0000	0.0000	0.1132
Girth	0.1 - 2.99 inches	0.0755	0.1509	0.0094	0.0094	0.2453
Difference	3.0 - 4.99 inches	0.0755	0.1698	0.0377	0.0000	0.2830
	5.0 + inches	0.1321	0.1792	0.0189	0.0283	0.3585
	Total	0.3208	0.5755	0.0660	0.0377	1.0000

Chi-Square Statistic	7.764
Probability	0.558

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
Sitting	< 15 Degrees	0.1311	0.0820	0.0164	0.0000	0.2295
Flexion	15-29 Degrees	0.1311	0.3852	0.0164	0.0164	0.5492
Difference	30+ Degrees	0.0246	0.0984	0.0738	0.0246	0.2213
	Total	0.2869	0.5656	0.1066	0.0410	1.0000

Chi-Square Statistic	38.034
Probability	0.000

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
	<= 0 Degrees	0.0744	0.0992	0.0248	0.0000	0.1983
Supine	1 - 4 Degrees	0.0661	0.2231	0.0248	0.0000	0.3140
Extension	5 - 8 Degrees	0.0909	0.1405	0.0413	0.0248	0.2975
Difference	9+ Degrees	0.0579	0.0992	0.0165	0.0165	0.1901
	Total	0.2893	0.5620	0.1074	0.0413	1.0000

Chi-Square Statistic	9.750
Probability	0.371

		Length of Stay				
		5 or less	6-10'	11-15'	15+	Total
Supine	< 15 Degrees	0.0909	0.0992	0.0000	0.0000	0.1901
Flexion	15-29 Degrees	0.1405	0.3058	0.0413	0.0083	0.4959
Difference	30+ Degrees	0.0496	0.1653	0.0661	0.0331	0.3140
	Total	0.2810	0.5702	0.1074	0.0413	1.0000

Chi-Square Statistic	17.833
Probability	0.007

TABLE 9a: Frequency Tables for the Pain Variable and the Outcome Variables

		DCScore (0 -10 basis)		
		0-4	5+	Total
	0 or negative	0.0686	0.0196	0.0882
Girth	0.1 - 2.99 inches	0.1863	0.0588	0.2451
Difference	3.0 - 4.99 inches	0.2549	0.0392	0.2941
	5.0 + inches	0.2451	0.1275	0.3725
	Total	0.7549	0.2451	1.0000
Chi-Square Statistic		3.987		
Probability		0.263		

		DCScore (0 -10 basis)		
		0-4	5+	Total
Sitting	< 15 Degrees	0.1441	0.0763	0.2203
Flexion	15-29 Degrees	0.4153	0.1441	0.5593
Difference	30+ Degrees	0.1949	0.0254	0.2203
	Total	0.7542	0.2458	1.0000
Chi-Square Statistic		3.848		
Probability		0.146		

		DCScore (0 -10 basis)		
		0-4	5+	Total
	<= 0 Degrees	0.1538	0.0513	0.2051
Supine	1 - 4 Degrees	0.2051	0.0940	0.2991
Extension	5 - 8 Degrees	0.2479	0.0513	0.2991
Difference	9+ Degrees	0.1538	0.0427	0.1966
	Total	0.7607	0.2393	1.0000
Chi-Square Statistic		2.042		
Probability		0.564		

		DCScore (0 -10 basis)		
		0-4	5+	Total
Supine	< 15 Degrees	0.1026	0.0855	0.1880
Flexion	15-29 Degrees	0.3675	0.1282	0.4957
Difference	30+ Degrees	0.2821	0.0342	0.3162
	Total	0.7521	0.2479	1.0000
Chi-Square Statistic		8.953		
Probability		0.011		

REFERENCES

Aczel, A., and Sounderpandian, J. Complete Business Statistics. Fifth Edition, McGraw-Hill Irwin Publishing, Boston, MA, 2002.

Alwan, L.C. Statistical Process Control. McGraw-Hill Irwin Publishing, Boston, MA, 2002.

Anderson, D., Sweeney, D. and Williams, T. Statistics for Business and Economics. Eighth Edition, Southwestern Publishing, Cincinnati, OH, 2002.

Aquilano, N., Chase, R. and Davis, M. Fundamentals of Operations Management. Second Edition, Irwin Publishing, Chicago, IL, 1995.

Neufelder, D., Raisor, J. and Friesner, D. "Determinants of Perceived Pain in Physical Therapy Patients: A Case Study of a Midwestern Health Care Provider." *Midwest Business Economics Association Conference Proceedings*, 2002.

Neufelder, D., Raisor, J. and Friesner, D. "Measuring Results: Outcomes Can Help You Improve Performance and Enhance Patient Care," *Advance for Directors in Rehabilitation*, July 2002, pp. 51-52.