Identifying Latent Outcome Measures in Inpatient Physical Therapy

Donna Neufelder, St. Mary's Rehabilitation Institute
Janet Raisor, St. Mary's Rehabilitation Institute
Mohammed Khayum, University of Southern Indiana
Daniel Friesner, Gonzaga University

ABSTRACT
Like all health care practitioners, physical and occupational therapists provide services to their patients that are intended to increase patient health and well-being. But because patient health and well-being are inherently latent, therapists are often forced to utilize indirect methods for measuring the effectiveness of their practices. These indirect methods usually involve gathering data on multiple indicators of therapy effectiveness. The problem with this approach is that there are usually multiple indicators for each latent phenomenon, and each of those indicators may give contradictory findings. As a result, the exact relationship between each indicator and the underlying phenomenon is often unclear. In this paper, we apply factor analysis to data on inpatient physical therapy following total knee replacement surgery (TKR) at a Midwestern rehabilitation institute. Our goal is to not only determine how many different latent measures of patient well-being are inherent in this type of therapy, but also to identify which (and how much) each of the empirical indicators commonly used to measure patient progress contribute to each of these underlying factors. The results of our analysis indicate that there are two underlying rehabilitation factors. One of these underlying measures is explained primarily by the flexion measures taken by the staff, while the other latent factor is explained by the remaining empirical indicators. Thus, it appears that patient progress in post-TKR therapy is really a two-fold phenomenon. As such, process improvement initiatives intended to quantify and increase the quality of care provided to patients must be adapted to take this finding into account.

INTRODUCTION

Like all health care practitioners, physical and occupational therapists provide services to their patients that are intended to increase patient health and well being. But because patient health and well-being are inherently latent, therapists are often forced to utilize indirect methods for measuring the effectiveness of their practices. These indirect methods usually involve gathering data on multiple indicators of therapy effectiveness. The indicators are then aggregated in some fashion (whether implicitly or explicitly) and used to make inferences about the practice's effectiveness.

While using empirical indicators to document practice effectiveness is very common in health care, there are several problems with this approach. First, because patient health and well being is such a broad definition, it is unclear as to exactly how many distinct, latent aspects of patient well being actually exist in rehabilitation therapy. For example, it may be the case that patient well-being is a single phenomenon. Alternatively, many studies (for example, Stanley and Cheek, 2003; Mossberg and McFarland, 2001; Weale, et al 2001; Rissanen et al 2000) assume that patient well-being can be decomposed into to multiple, distinct aspects, including (but certainly not limited to) physical well-being and mental well-being. The appropriateness of this assumption certainly depends on the type and severity of medical treatment. Physical and occupational rehabilitation, for example, is both corporeal and psychological in nature (for example, Jerosch and Floren, 2000; Jackson et al, 1998), implying that decomposing patient well-being into multiple factors is appropriate. However, there is little consensus in the literature about the appropriate number of latent factors to be analyzed (Stanley and

Cheek, 2003). And if the appropriate number of distinct outcomes is not identified correctly, process improvement studies (which inherently rely on this assumption) may be of little use, since the results generated will not appropriately characterize therapy effectiveness.

Another problem is that there are usually multiple indicators for each latent phenomenon. On the surface, utilizing a large number of empirical indicators for each latent factor may appear beneficial, as it provides a more robust examination of the unobservable factor. Unfortunately, this approach is not without its consequences. In the worst-case scenario, the indicators may give contradictory results, and the staff must decide which of the indicators are reliable (i.e., better indicators of the true latent variable) and which are not. Alternatively, each of the indicators may give the same general conclusion (i.e., the process is effective or not effective), but differ in the degree to which the process is effective or not effective. In this case, the practice's staff must still make judgments about the relative effectiveness of each of the empirical indicators.

In this paper, we apply factor analysis (a statistical technique commonly used by statisticians and management scientists to characterize latent processes) to data on inpatient physical therapy following total knee replacement surgery (TKR) at a Midwestern rehabilitation institute. Our goal is to not only determine how many different latent measures of patient well-being are inherent in this type of therapy, but also to determine which (and how much) each of the empirical indicators commonly used to measure progress contribute to each of these underlying factors. In doing so, our study will provide empirical evidence which other therapy practices can utilize to improve the effectiveness of their process improvement initiatives.

The remainder of this paper proceeds in three steps. First, we describe the database used in our study. Next, we describe some basic statistical techniques (including factor analysis) 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. Patients referred for outpatient therapy receive one session per day, while those admitted for inpatient therapy receive two sessions 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 the physical aspects of 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 measurement should increase, while extension measurement should approximate zero. 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-operative) 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-operative 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.

Data was also collected on a fifth variable intended to measure the psychological effectiveness of therapy. At the conclusion of treatment, patients were asked to evaluate their perceived pain using a 0 (no pain) to 10 (maximum pain) rating scale.[2] Lastly, data was collected on a number of supporting variables, including 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 of these patients. As a result, there are some missing values, leaving us with a working data set of 100 observations. 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 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 identify the true latent outcomes of the process, not does it provide the relative contribution of each of these indicators to the latent outcomes.

Our second approach applies factor analysis to the data in an effort to identify each of the latent outcomes that characterize patient rehabilitation. The benefit to this approach is that it allows the provider a more detailed method of characterizing and improving their practice methods, which may not be inferred from the descriptive statistics. This type of analysis is also easily conducted using tools such as SPSS or SAS, and can be presented in a format that is easily interpreted by those with a limited statistical background. However, while factor analysis results are fairly easy to read and interpret, conducting the analysis implicitly assumes a more detailed understanding of statistical methods. As such, we present a brief review of factor analysis theory prior to discussing our results[3].

DESCRIPTIVE STATISTICS AND HYPOTHESIS TESTS

Summary statistics for the original data used in this study are reported in Table 2. Of the 100 patients considered in the study, 71 percent are female and 29 percent are male. The mean age of patients is 70.03 years and the average length of stay is 7.4, with twelve different physicians performing the knee replacement procedures[4]. 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 4.96 degrees and gained 21.7 and 25.2 degrees of additional sitting and supine flexion, respectively.

The average amount of girth difference in the data set is 3.97 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 3.97 is statistically different from any postulated value less than 3.4. So if 3.4 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.

AN INTRODUCTION TO FACTOR ANALYSIS

The first step in conducting factor analysis is to determine whether the data are, in fact, appropriate for the analysis. More specifically, factor analysis is a data reduction technique that relies upon (and hence assumes) the fact that the variables of interest are empirical indicators for some common, latent process(es). If this is the case, then these variables should be sufficiently correlated with each other. If they are not sufficiently correlated, then there is no statistical relationship between the indicators, and thus no common latent process relating these variables. Statisticians have developed several heuristic measures to determine whether the data are, in fact, appropriate for factor analysis. In this paper, we will discuss and employ three of the most commonly used measures (Sharma 1996).[5] The first is to examine the size and significance of the correlation coefficients between each of the empirical indicators. If these correlations are strong and statistically significant, then factor analysis is more likely to be appropriate for the data. Conversely, weak correlations indicate that factor analysis is not appropriate.

A second measure is the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. This is essentially a statistic that is bounded between zero and one. The higher the value of the statistic, the more likely it is that factor analysis can be applied to the data. Generally, this measure must be above 0.5, and values higher than 0.8 are preferred. The final statistic is the Bartlett test of sphericity. This is a chi-square test intended to determine whether the correlations among the variables are jointly statistically significant. Since this test follows the chi-square distribution, it takes values between zero and infinity, with larger test statistic values indicating rejection of the null hypothesis of no joint correlation among the empirical indicators.[6]

Having described the assumptions of factor analysis, it is possible to discuss the intuition behind the technique using a simple example. Suppose we have an inpatient physical therapy process for patients recovering from TKR surgery. Also suppose that the staff uses two empirical indicators for patient rehabilitation: flexion (F) and extension (E). If there is one underlying measure of patient progress (P), then the relationship between the empirical indicators and the underlying factor can be expressed mathematically as:

$$F = \beta_1 * P + u_F \tag{1}$$

$$E = \beta_2 * P + u_E \tag{2}$$

where β_1 and β_2 represent the coefficients that relate the mean (marginal) impact of the latent variable (P) on the empirical indicators E and F; and u_F (u_E) represent that part of the empirical indicator E (F) that is unrelated to the latent factor. The β's are commonly referred to as *pattern loadings*, while the u's are referred to as the *unique factors*. In general, if one has information on P, then (1) and (2) can be interpreted as regression equations and the β's can be estimated. In that case, P provides information about the expected determinants of each empirical indicator, while the u's represent the stochastic error terms for each regression equation. Moreover, as in regression analysis, the overall (or total) variance of each regression equation can be decomposed into two parts: that part of the overall variance explained by the latent factor, and that part explained by the error term:

$$\sigma^2_{Total} = \sigma^2_{Explained} + \sigma^2_u \tag{3}$$

The part of the variance that is explained by the latent factor is commonly known as the *communality* of the equation while portion explained by the error term is known as the *unique variance*.

The previous example assumes that there is only one latent process, and that both empirical indicators measure that process. However, it may be the case that there are multiple latent processes.[7] In the case of two empirical indicators, there can be as many as two latent processes. Call these processes P_1 and P_2. Now equations (1) and (2) can be re-written as:

$$F = \beta_{11} * P_1 + \beta_{12} * P_2 + u_F \tag{1b}$$

$$E = \beta_{21} * P_1 + \beta_{22} * P_2 + u_E \tag{2b}$$

As before, assuming that P_1 and P_2 can be determined, (1b) and (2b) can be considered as (multiple) regression equations, the variance of which can also be decomposed into a form analogous to (3).[8] Examination of the β's provide information about the number of latent processes as well as the extent to which each of the empirical indicators measure those processes. For example, if $\beta_{12} = \beta_{22} = 0$ and $\beta_{11}, \beta_{21} \neq 0$, then there is only one significant latent process (P_1), and both indicators measure that process. Alternatively, if $\beta_{12} = \beta_{21} = 0$ and $\beta_{11}, \beta_{22} \neq 0$, then there are two latent processes and each indicator is a distinct indicator of only one latent process. The purpose of factor analysis is to identify how many latent processes exist in the data, and to subsequently characterize the marginal relationship between the empirical indicators and each of those processes.

There are several methods for determining (or *extracting*) the number of latent processes and the communalities, the two most common of which are *principal components factoring* and *principal axis factoring*. In this paper, we use the former approach.[9] Principal components factoring operates by normalizing the total variation of each indicator such that it takes a value of one. It then utilizes linear algebra to isolate the eigenvalues of the data matrix, which represent the maximum number of latent processes in the data.[10] Once the maximum number of latent processes has been determined, a set of general guidelines is used to determine how many of the eigenvalues are relevant to (or significant enough to be used in) the analysis. The most common guideline (which we employ in this analysis) is the eigenvalue-greater-than-one rule, which (obviously) states that any eigenvalue taking a value greater than one represents a significant latent process.[11] Having identified the correct number of latent factors as well as their values, regression analysis can be used to identify an initial solution to the pattern loadings.

A problem with these initial results is that they are not unique; that is, the results produced are only one of several possible (optimal) outcomes. As such, an additional step is required, which is often known as *rotating* the initial factor loadings. Essentially, this entails conducting an iterative procedure based on a pre-defined set of criteria to produce a single, optimal set of factor loadings. There are a number of different factor rotation procedures used in the literature. For simplicity, we use the two most commonly employed in the literature: the VARIMAX rotation method and the QUARTIMAX rotation method (Hair et al 1998, Sharma 1996). The former employs a criterion that attempts to find a solution where each empirical indicator has factor loadings that are very high for one latent factor, and very close to zero for all other latent factors. That is, VARIMAX attempts to assign each empirical indicator to a single latent factor. In our previous example, this is equivalent to attempting to setting β_{12}, $\beta_{21} \approx 0$ and β_{11}, $\beta_{22} > 0$. The QUARTIMAX rotation method is slightly less stringent: it attempts to assign each empirical indicator *primarily* (but not completely) to a single latent factor. Under QUARTIMAX, the factor loadings are more likely to be more evenly distributed across each of the factors. Thus, using our previous example, we might expect to find that β_{12} and β_{21} are somewhat larger in absolute value under QUARTIMAX than under VARIMAX. However, it should be noted that both rotation methods often provide very similar results (Sharma 1996).

APPLYING FACTOR ANALYSIS TO THE REHABILITATION DATA

Based on the information provided above, it was decided that five different variables should be included as empirical indicators in the factor analysis: the three differenced measurements in knee flexion and extension, the girth difference measurement and the perceived pain score at discharge from therapy. The first three measures are natural choices, as they represent the physical gains in mobility (or reductions in extension) that result from completing therapy. The fourth measure was chosen because of its strong a priori expected impact on therapy effectiveness. That is, increases in the knee girth measurement should significantly retard therapy effectiveness. The last indicator was chosen because it is hypothesized to represent the psychological aspects of therapy effectiveness.

Table 4 provides information that determines whether the data are appropriate for factor analysis. The results generally support the use of factor analysis, although the different indicators provide mixed evidence about the strength of this support. In Table 4, we see that there are some significant correlations among the different indicators, particularly between the two flexion measurements, as well as between the two measurements taken in the supine position. The Pearson correlations also indicate a significant relationship between the girth measurement and the extension difference measure. Interestingly, there is no significant correlation between the perceived pain score and any of the other empirical indicators.[12] The KMO measure of sampling adequacy is greater than the minimum of 0.5, but again, only marginally so. The Bartlett test result provides the only strong indication in favor of analyzing the data with factor analysis. The test statistic is 39.6, which rejects the null hypothesis (of no significant joint correlation) at better than a one percent level of significance. Given this finding, along with the fact that there were strong correlations across four of the five variables, we decided to proceed with the factor analysis.[13]

Table 5 presents the results of our principal components factoring. Of the 5 possible eigenvalues (or latent factors), only two have values greater than one. Additionally, those two factors jointly explain almost 60% of the variance in the data. From a practitioner's standpoint, these results are interesting, because they imply that there are actually two significant measures of patient progress in inpatient therapy following TKR. The results in Tables 6 and 7 show whether and how each of empirical indicators relate to each of those factors. Table 6 gives the extracted communalities for each indicator. We see that the two latent factors explain as little as 41 percent of the variation in the perceived pain score variable, but as much as 75 percent of the gain in flexion. Given the fact that the pain variable was not correlated with the other outcome measures, as well as the fact that the supine flexion measure was highly correlated with several of the other outcome measures, these results are not surprising.

Tables 7a and 7b give the final (rotated) factor loadings. As expected, the results from both the QUARTIMAX and VARIMAX rotations are very similar, both in sign and magnitude. The DifSupFlex and DifSitFlex variables load very highly on the first factor, but only minimally on the second factor. The remaining three outcome measures load very highly on the second factor and marginally on the first factor. The implication of this finding is quite intuitive: the two latent outcomes are the gain in mobility that results from completing therapy, and all other outcome factors over the course of therapy.

We can also go further to examine the marginal relationship between each indicator and its latent process. The DifSupFlex and DifSitFlex variables are positively associated with the first latent factor. The magnitudes of these loadings are also very similar. This indicates that gains in flexion (whether taken in a sitting or supine position) are positively related to the latent factor measuring the mobility gains from therapy, and that each indicator provides an equally accurate indication of patient progress. From a process improvement standpoint, this implies that each of these measures is an (approximately) equal indicator of mobility gains. As such, they should be given the same amount of consideration in process improvement studies.

The signs and magnitudes of the primary factor loadings for the remaining empirical indicators are slightly less intuitive. All three factors are positively associated with the second latent outcome, although the magnitude of this (marginal) relationship varies across each of the empirical indicators. A one-unit increase in the latent factor causes a 0.61 unit increase in the perceived pain score at discharge, a 0.63 unit increase in the reduction of supine extension, and a 0.72 unit increase in the girth difference measure. What is particularly troubling (and at the same time intriguing) is the fact that a change in the latent factor has marginal impacts on the empirical indicators that are counter-intuitive. For example, a one-unit change in the latent factor is likely to increase the girth and perceived pain measures, implying that the latent outcome should be identified as *reductions* in patient wellness that result from feeling higher levels of pain and having a lower initial illness severity.

However, the final outcome is also positively related to reductions in extension. Based on this criterion, one would be inclined to identify the latent factor as the *increase* in patient wellness that results from reducing extension. This last finding obviously conflicts with a priori expectations, since an increase in patient wellness should result in *lower pain and girth scores, but higher reductions in extension* (and analogously if the latent factor is interpreted as a reduction in well-being). That is, we would expect the factor loading for the extension measure and the remaining measures to be of opposite sign.

One possible explanation for this discrepancy lies in the interpretation of the factor loadings as regression coefficients. Specifically, one must interpret each of these factor loadings *holding constant the impacts of the other specified explanatory (or latent) variables*. Thus, the counterintuitive factor loading for the extension measure, for example, may come from the fact that we are holding constant the latent factor determining flexion-based patient well-being, possibly encompassing all physical aspects of the physical mobility gains from therapy. And so having controlled for these factors, the extension measure (and its factor loading) may be picking up other aspects of the therapy progress, possibly the initial illness severity aspects of this treatment.[14] It remains to be seen from future research whether these assertions are borne out.

DISCUSSION AND CONCLUSIONS

The purpose of this paper is to present an exploratory empirical analysis using data on inpatient physical therapy following total knee replacement surgery (TKR) at a Midwestern rehabilitation institute. Our goal was to not only determine how many different latent measures of patient well-being are inherent in this type of therapy, but also to identify which (and how much) each of the empirical indicators commonly used to measure patient progress contribute to each of these underlying factors. The results of our analysis indicate that there are two underlying rehabilitation factors. One of these factors is explained primarily by the flexion measures taken by the staff, while the other is explained by the remaining empirical indicators.

Our findings present several implications for health care practitioners. First, it appears that patient progress in post-TKR therapy is really a two-fold phenomenon. As such, process improvement initiatives intended to quantify and increase the quality of care provided to patients must be adapted to take this finding into account. Our findings also indicate that the two flexion measurements commonly taken by therapists contribute relatively evenly to one underlying aspect of patient progress and well-being, while the remaining factors primarily contribute to the second (or "catch-all") underlying factor, with each empirical indicator contributing in varying degrees. Thus, practitioners may want to give both flexion measures equal consideration in evaluating therapy effectiveness. However, when evaluating the second latent measure of patient progress, practitioners may want to pay a little more (or a little less, depending on the interpretation of the factor) attention to the girth difference measurement versus the extension and perceived pain indicators.

While our study provides an initial analysis of latent factors in inpatient physical therapy following TKR, our findings are preliminary, and should be viewed with caution. However, these limitations also provide some suggestions for future research. One drawback to our study is that our data come from a single, nonprofit health care provider during a single fiscal year. And while our data form the basis for an interesting case study, other health care providers of different size, profit status, or socio-economic distinction may find different results. Analyses of patients completing therapy following different procedures (other than TKR) may also obtain different findings. Thus, replications of our study that utilize a larger, and more general sample of patients and (initial) procedures would provide a valuable additional to our understanding of the determinants of patient well-being and progress from therapy.

Another limitation of our study is one that also plagues the entirety of the factor analysis literature; namely, that factor analysis identifies how many latent processes there are, but does not specifically identify what those processes represent. Our study, for example, found two major latent factors, one of which was closely associated with our flexion measures, and one that was associated with the remaining empirical indicators. In the former, one can use experience and intuition to appropriately "label" the meaning of the factor as "well-being or progress attributable to gains in flexion". However, in the latter, it is not clear what the meaning of this "catch-all" factor really is. And unless the researcher and/or the practitioners can intuitively identify what that factor is, changing practice patterns based on those findings should be done with caution.

TABLE 1: Variable Definitions

Variable	Definition
NOGirth	Girth measurement (in inches) for a patient's non-operative knee.
TKRGirth	Girth measurement (in inches) for a patient's surgically repaired knee.
DifGirth	Difference between the TKRGirth and the NOGirth 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
NOGirth	42.02	4.067
TKRGirth	45.99	4.016
DifGirth	3.971	2.569
PreSupExt	8.650	4.963
PostSupExt	3.690	3.530
DifSupExt	4.960	4.134
PreSupFlex	63.95	16.58
PostSupFlex	89.14	10.38
DifSupFlex	25.19	13.78
PreSitFlex	70.57	13.30
PostSitFlex	92.28	10.06
DifSitFlex	21.71	9.575
Sex	0.710	0.456
Age	70.03	8.556
DCscore	2.880	2.110
Los	7.430	3.325
Number of Observations		100

TABLE 2b: Frequency Table for Physicians

MD Indicator	Frequency of TKR's
A	29
B	23
C	15
D	10
E	6
F	6
All Others	11

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

Variable	Mean	Std. Error of the Mean	T-Ratio	Prob.
DCscore	2.88	0.211	13.649	0.000**
DifGirth	3.97	0.257	15.459	0.000**
DifSupExt	4.96	0.413	11.998	0.000**
DifSupFlex	25.19	1.378	18.285	0.000**
DifSitFlex	21.71	0.958	22.673	0.000**

* indicates statistical significance at the 10% level
** indicates statistical significance at the 5% level

TABLE 4a: Pearson Correlations

Variable 1	Variable 2	Correlation Coefficient	Prob. (two-tailed)
DCscore	DifGirth	0.127	0.209
DCscore	DifSupExt	0.100	0.321
DCscore	DifSupFlex	-0.114	0.258
DCscore	DifSitFlex	-0.009	0.927
DifGirth	DifSupExt	0.201	0.045**
DifGirth	DifSupFlex	-0.029	0.772
DifGirth	DifSitFlex	0.074	0.464
DifSupExt	DifSupFlex	0.129	0.202
DifSupExt	DifSitFlex	0.144	0.154
DifSupFlex	DifSitFlex	0.505	0.000**

TABLE 4b: Spearman (Nonparametric) Correlations

Variable 1	Variable 2	Correlation Coefficient	Prob. (two-tailed)
DCscore	DifGirth	0.147	0.145
DCscore	DifSupExt	0.070	0.490
DCscore	DifSupFlex	-0.078	0.441
DCscore	DifSitFlex	-0.010	0.923
DifGirth	DifSupExt	0.148	0.143
DifGirth	DifSupFlex	-0.059	0.561
DifGirth	DifSitFlex	0.045	0.658
DifSupExt	DifSupFlex	0.176	0.080*
DifSupExt	DifSitFlex	0.054	0.595
DifSupFlex	DifSitFlex	0.566	0.000**
KMO Measure		0.521	
Bartlett Chi-Square Test Statistic (10 dof)		39.566	0.000**

* indicates statistical significance at the 10% level
** indicates statistical significance at the 5% level

TABLE 5: Initial Eigenvalues and Percent of Variance Explained

Component	Total	Percent of Variance	Cumulative Percent
1	1.587	31.748	31.748
2	1.275	25.498	57.246
3	0.876	17.520	74.766
4	0.785	15.691	90.457
5	0.477	9.543	100.000

TABLE 6: Extracted Communalities using Principal Components

Variable	Extracted Communalities
DCscore	0.414
DifGirth	0.514
DifSupExt	0.480
DifSupFlex	0.750
DifSitFlex	0.714

TABLE 7a: Factor Matrices using Varimax Rotation

	Unrotated Factor Loading Matrix		**Rotated Factor Loading Matrix**	
Variable	Factor 1	Factor 2	Factor 1	Factor 2
DCscore	-0.051	0.642	-0.215	0.607
DifGirth	0.205	0.687	0.021	0.717
DifSupExt	0.442	0.533	0.290	0.629
DifSupFlex	0.810	-0.305	0.862	-0.085
DifSitFlex	0.831	-0.117	0.833	0.101

Transformation Matrix

Component	1	2
1	0.966	0.258
2	-0.258	0.966

TABLE 7b: Factor Matrices using Quartimax Rotation

	Unrotated Factor Loading Matrix		Rotated Factor Loading Matrix	
Variable	Factor 1	Factor 2	Factor 1	Factor 2
DCscore	-0.051	0.642	-0.217	0.606
DifGirth	0.205	0.687	0.018	0.717
DifSupExt	0.442	0.533	0.287	0.631
DifSupFlex	0.810	-0.305	0.862	-0.082
DifSitFlex	0.831	-0.117	0.833	0.105

Transformation Matrix

Component	1	2
1	0.965	0.261
2	-0.261	0.965

References

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

Hair, J., Anderson, R., Tatham, R. and Black, W. Multivariate Data Analysis. Fifth Edition, Prentice Hall Press, Upper Saddle River, NJ, 1998.

Jackson, J., et al. "Occupation in Lifestyle Redesign: The Well Elderly Occupational Therapy Program." *American Journal of Occupational Therapy*, Vol. 52, No. 5, 1998, pp. 326-336.

Jerosch, J., and Floren, M. "Quality of Life Improvement (SF-36) after Implementation of a Knee Endoprosthesis." *Unfallchirurg*, Vol. 103, No. 5, 2000, pp. 371-374.

Johnson, Richard, and Wichern, Dean. Applied Multivariate Statistical Analysis. Fifth Edition, Prentice Hall Press, Upper Saddle River, NJ, 2002.

Mossberg, K., and McFarland, C. "A Patient-Oriented Health Status Measure in Outpatient Rehabilitation." *American Journal of Physical Medicine and Rehabilitation*, Vol. 80, No. 12, 2001, pp. 896-902.

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.

Neufelder, D., et al. "Using Quality Control Techniques to Measure Performance: A Case Study of a Midwestern Rehabilitation Institute." *Midwest Business Economics Association Conference Proceedings*, 2003.

Neufelder, D., et al. "Benchmarking Patient Improvement in Inpatient Physical Therapy with Data Envelopment Analysis." Working Paper, University of Southern Indiana, 2003.

Rissanen, P., et al. "Do QOL Instruments Agree? A Comparison of the 15D (Health-Related Quality of Life) and NHP (Nottingham Health Profile) in Hip and Knee Replacements." *International Journal of Technology Assessment in Health Care*, Vol. 16, No. 2, pp. 696-705.

Sharma, Subhash. Applied Multivariate Techniques. John Wiley and Sons Press, New York, NY, 1996.

Stanley, M., and Cheek, J. "Well-Being and Older People: A Review of the Literature." *Canadian Journal of Occupational Therapy*, Vol. 70, No. 1, 2003, pp. 51-59.

Weale, A.E., et al. "Perceptions of Outcomes after Uni-compartmental and Total Knee Replacements." *Clinical Orthopedics and Related Research*, Vol. 382, 2001, pp. 143-153.

Endnotes

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

[2] One limitation of our study is that this measure was not taken at the beginning of therapy. As a result, we are not able to obtain a net reduction in pain over the course of therapy.

[3] In what follows, we assume that the reader is familiar with one or more of the statistical packages (such as SPSS, MINITAB or SAS) commonly used to conduct factor analyses. As such, we will focus primarily on the steps in the application of this technique. For those readers who are familiar with factor analysis theory, one may want to only briefly skim pages 7-11 of the manuscript, which review this theory. Normally, we would place such information in the appendix of the paper. However, because practitioners may not be familiar with this technique, and because one of the purposes of this paper is to demonstrate how factor analysis can be used to improve process performance, we felt that it would be more beneficial to leave this information in the body of the paper. Even so, our discussion abstracts from most of the technical details, and refer the interested reader to Sharma (1996), Hair et al (1998) or Johnson and Wichern (2002) for the full discussion of these issues.

[4] Of the 100 patients included in the study, 89 of the patients were treated by only 6 of the 12 physicians. This finding is also consistent with the entire data set: 108 of the 122 observations were treated by these same six physicians.

[5] Because these measures are primarily heuristic, there is no hard and fast rule with which a researcher can make a concrete decision about whether factor analysis can be applied to the data. As such, the researcher must apply his/her intuition and expertise to these measures in order to make a reasonable decision.

[6] A potential drawback to the Bartlett test is that it is sensitive to the sample size employed in the analysis. Specifically, larger sample sizes may inflate the test statistic, thereby increasing the chances that the null hypothesis will be rejected. However, given the fact that our data set is relatively small ($n = 100$), this is not a significant concern.

[7] In general, the number of latent processes can be as few as one and as many as the number of empirical indicators employed in the analysis.

[8] If there are multiple latent processes, then we can further decompose the explained variance into partial explained variances, or communalities, specific to each latent factor.

[9] Principal axis factoring is essentially equivalent to principal components factors. The difference between the two is that principal components factoring normalizes the total variance of each empirical indicator to one. As such, the estimated communalities should be interpreted as the proportion of total variance that is explained by the latent process. Principal axis factoring takes the results generated by principal components and uses an iterative procedure to reverse the normalization, so that the communalities can be expressed in absolute (as opposed to relative) terms. The drawback to principal axis factoring is that the

iterative procedure may not converge for some data sets, including the one we employ in this study.

[10] By definition, there will be one eigenvalue for every empirical indicator used in the analysis. As such, this technique always finds the maximum number of possible latent factors.

[11] Other guidelines employed in the literature include the scree-plot guideline, the parallel analysis and the minimum average partial correlation analysis (Sharma 1996). We utilize the eigenvalue-greater-than-one rule not only because it is the most popular of these rules, but also because it is the rule commonly employed by statistical packages such as SPSS and SAS.

[12] We also re-ran the analysis excluding the perceived pain score and found only minor differences in the results. As such, we decided to include this variable for the sake of completeness.

[13] We note in passing that, should our decision to proceed be incorrect, one would expect to find that each of our empirical indicators load almost completely onto its own factor. So if we find that the indicators load very highly onto only one or two factors, it would support (but not necessarily prove) the assertion that our decision was correct.

[14] Keep in mind that the extension measure is the difference between the pre and post-therapy extension measurements. So this last possibility may be particularly true if the pre-therapy contribution to this indicator dominates the post-therapy contribution, and thus all correlations between the differenced extension measure and the other four empirical indicators.