THE COST-EFFECTIVENESS OF A TREATMENT-BASED CLASSIFICATION (TBC) APPROACH COMPARED TO A USUAL CARE APPROACH IN THE MANAGEMENT OF LOW-BACK PAIN IN THE OUTPATIENT PHYSICAL THERAPY SETTING

by

John Christopher McGee

B.S. Physical Education, Salisbury University, 1990

Master's Physical Therapy, University of Maryland Eastern Shore, 1998

Master's of Science, Financial Management, University of Maryland University College, 2008

Master's Business Administration, University of Maryland University College, 2009

Submitted to the Graduate Faculty of

the School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

John Christopher McGee

It was defended on

September 21st, 2009

and approved by

G. Kelley Fitzgerald PhD, PT, FAPTA, Associate Professor, Department of Physical Therapy

Pamela B. Peele, PhD, Associate Professor, GSPH; Department of Health Policy & Management

John D. Childs, PhD, PT, MBA, OCS, CSCS, FAAOMPT, Associate Professor, U.S. Army-Baylor University, Department of Physical Therapy

Michel Landry, PhD, MScPT, Assistant Professor, University of Toronto, Department of Physical Therapy

John W. Wilson, PhD, Assistant Professor, GSPH; Department of Biostatistics

Dissertation Advisor: Anthony Delitto, PhD, PT, FAPTA, Professor & Chair, Department of Physical Therapy

THE COST-EFFECTIVENESS OF A TREATMENT-BASED CLASSIFICATION (TBC) APPROACH COMPARED TO A USUAL CARE APPROACH IN THE MANAGEMENT OF LOW-BACK PAIN IN THE OUTPATIENT PHYSICAL THERAPY SETTING

John Christopher McGee, MPT, MBA, OCS, ATC

University of Pittsburgh, 2009

Purpose and Study Design: Retrospective cohort cost-minimization analysis (payer perspective) with decision analysis model to access cost-effectiveness of a treatment-based algorithm (TBC) for low back pain (LBP) compared to a "usual" care strategy in the outpatient setting

Methods: charge data was examined on 750 subjects with LBP from 42 regional clinics over 1 year period. Subjects were determined to be on or off protocol for the classification algorithms based on provider responses to minimum required initial exam and history intake data and subsequent interventions provided. Primary outcome measures were total net direct health care and physical therapy costs, along with total member and physical therapy member burden costs. In addition, protocol status was examined as a predictor variable for the following: top quartile of total direct health care and physical therapy expenditures, as well as total direct health care and physical therapy member burden. A 4% / yearly discounting rate was applied.

Results: Baseline characteristics of the combined sample demonstrated a significant proportion of Medical Assistance patients were given non-adherent care. In addition, a significant but not clinical difference was found in fear-avoidance behavioral questionnaire physical activity (FABQ_PA) scores. Incremental cost-savings were demonstrated in all primary outcome measures for the combined sample. The specific exercise and flexion off-

protocol subgroups demonstrated member burden savings but this was explained exclusively after adjustment by having Medical Assistance as an insurance type. Off-protocol status accounted for significant variation in explaining differences in the statistically different outcomes, as well as demonstrating predictive ability for attaining the top quartile of total direct health care expenditures. The decision analysis model demonstrated the dominance of classification approach to usual care across a variety of associated variable ranges and distributions.

Conclusions: This evidence supports the TBC as a cost-effective alternative for LBP treatment compared to usual physical therapy care. It appears beneficial for a payer to adopt strategies to improve compliance with the TBC. Further recommendations are suggested to either validate or cross-validate these findings and to improve outcomes reporting. The TBC should also be compared as a cost-effective alternative to treating LBP against primary-care (non-rehabilitative) and chiropractic.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Foremost, I must thank my wife Traci, and my four children; Abigail (9), Gavin (7), Zachary (5), and Jacob (2). This endeavor would have never been possible without your sacrifice and loving support. I also want to thank my mother Doris for being a true example of perseverance. You are all the greatest blessing in the world to me. Next, I would like to express my gratitude to both my advisor and my committee members for your tolerance, guidance, and insight. Any future professional success that befalls me ultimately stems from you. Lastly, I would like to take this opportunity to extend my appreciation to the U.S. Air Force Physical Therapy Branch for granting me this awesome opportunity to pursue knowledge. It is my true privilege and honor to serve with you in the greatest organization on the face of this earth. God Bless, and Blue Skys Forever!

John aka (Daddy) aka Major McGee

1.0 INTRODUCTION

Low back pain (LBP) occurs at epidemic proportions in the United States and associated annual health care expenditures pose an enormous burden to our society. Most clinical practice guidelines fail to recognize non-specific low back pain as a heterogeneous problem thereby recommending only education for all patients during the initial weeks of management, and referral to physical therapy is recommended only when recovery is delayed. Traditionally, a lack of conclusive research has provided physical therapists with sparse information to guide clinical decision making resulting in suboptimal outcomes and wide variations in practice patterns. Increasingly however, the latest evidence has demonstrated significantly improved effectiveness of using a physical therapy treatment classification approach to LBP based on selecting a particular early intervention strategy for clusters of patient signs and symptoms. Nonetheless, despite this evidence, wide variability in physical therapy practice patterns still exists. As such, the limited cost-effectiveness research involving physical therapy as a treatment option for LBP has also been inconclusive at best. Previously all studies have compared physical therapy as a single entity without consideration of any treatment-based classification schema. As a result, this homogenous grouping has provided limited effect sizes for physical therapy treatment when compared to any other form of intervention.

In order for physical therapy to be compared as a treatment arm in any economic analysis, it has to be given an opportunity to "put its best foot forward". A treatment-based classification approach to LBP has been demonstrated to be clinically effective but its cost-savings and subsequent cost-effectiveness has not been properly assessed. Improved value and decreased variability in the management of LBP may only be realized if the treatment-based classification approach is compared first to usual physical therapy care and then eventually to other medical alternatives.

1.1 STATEMENT OF PURPOSE

The purpose of this study was to determine the cost-effectiveness of the treatment-based classification (TBC) approach versus a usual care approach for the management of low-back pain (LBP) in the outpatient physical therapy setting. This project combined a cost-minimization analysis from retrospective data with a decision analysis model to make a cost-effectiveness inference of the TBC strategy compared to a usual care approach. This inference was based upon clinician treatent decisions made upon initial physical thearpy evaluation for this sample. In addition, this investigation was conducted over a time horizon of 20 months, and a payer perspective was applied in the analysis comparing total direct healthcare and physical therapy costs.

So far, only cost-minimization studies have been performed in the rehabilition literature examining physical therapist treatment decisions. The purpose of these studies was to exam the cost-savings associated with utiizing an active therapeutic versus a passive therapeutic approach to LBP. Although, these studies provided some general insight to the benefits of incorporating the literature and current evidence-based guidelines into practice, their analysis was limited to the exmination of Current Procedural Terminology (CPT) codes. Because these codes are often altered and biased to reflect a higher compensation schema, the infernces from these studies somewhat fall short of actually measuring the results associated with actual clinician treatment behavior. Therefore, this is the first study to attempt examaing the cost effectiveness of actual clinican responses to the reccommended treatment algorithms for LBP. The results from this investigation should provide a reasonable preliminary interpretation of the overall cost-effectiveness of the TBC. In fact, the data contained in this study should provide justification for a larger National Institutes of Health (NIH) federally funded grant in comparative analysis of evidence-based therapeutic interventions with an expanded time horizon of up to five years for better long-term generalizability of the TBC. In the meantime, these results should also be helpful in establishing compliance goals for seemingly underutilized algorithms and guidelines (e.g. manipulation and fear-avoidance behavioral interventions) given that the cost effectiveness of the TBC is demonstrated.

1.2 SPECIFIC AIMS AND HYPOTHESES

1.2.1 Specific Aim 1

Perform a cost minimization analysis comparing both direct health care and physical therapy costs for subjects classified as on-protocol according to the three major subsets (manipulation, specific exercise, and stabilization exercise) of a TBC approach to those who are considered off-protocol (receiving a usual physical therapy care approach) in the management of patients with LBP in the primary outpatient physical therapy setting. In addition, member total cost burden and physical therapy cost burden will be compared by protocol status to determine if treatment status influences a subject's actual out-of-pocket direct health care expenses.

1.2.2 Hypothesis 1.1

Overall direct health care costs, total physical therapy costs, member cost burden and member physical therapy cost burden for subjects properly classified and treated according to a TBC approach (on-protocol) will be substantially less than those costs associated with a usual physical therapy care approach (off-protocol) in the management of patients with LBP.

1.2.3 Hypothesis 1.2

Off-protocol status will be a strong predictor for subjects incurring the following; unusually high direct heath care costs, unusually high physical therapy costs, unusually high member cost burden, and unusually high physical therapy member cost burden.

1.2.4 Specific Aim 2

Develop a decision analysis model containing cost data from the retrospective Specific Aim 1 results and outcome / terminal end state data inferred indirectly from the literature in order to obtain the Incremental Cost Effectiveness Ratios (ICER's) for both treatment arms, thus gaining a better assessment of the overall cost-effectiveness of the TBC approach compared to a usual care approach.

1.2.5 Hypothesis 2

The combined treatment based classification approach will be more cost-effective then the usual care approach across a greater range of both cost and outcome variable distributions.

2.0 REVIEW OF THE LITERATURE

2.1 IMPACT OF LOW BACK PAIN (LBP)

Low back pain (LBP) has consistently placed an incredible societal and economic burden on the United States. In general , it is reported to affect at least 50%- 80% of individuals during their lifetime.(Rubin DI, 2007) Moreover, within a single year, approximately 15% - 20% of the population is expected to have an episode of LBP.(Rubin DI, 2007) The 3-month point prevalence for LBP in the U.S. reportedly ranges from 26% - 31% of the population or about 34 million individuals.(Deyo RA, Mirza SK, & Martin BI, 2006; Strine TW & Hootman JM, 2007) Furthermore, total healthcare expenditures in the U.S. incurred by individuals with LBP exceeds 90 billion dollars annually, with overall healthcare expenditures reportedly 60% higher for those with LBP than in individuals without LBP.(Luo X, Pietrobin R, & SX, 2004) It has been further documented that aside from the common cold, LBP is the most common reason why individuals visit a primary care physician's office.(Deyo RA & Phillips WR, 1996) Similarly, labor and production loss attributed to LBP ranges in the billions of dollars annually.(Luo X et al., 2004) In addition, LBP accounts for 50% of all patients seeking outpatient physical therapy care.(Di Fabio RP & Boissonnault W, 1998)

Currently, the total proportion of all physician visits attributed to LBP is 2.3% and has changed very little since the early 1990's, thus demonstrating that the problem is not improving.(Weiner DK, Kim YS, Bonino P, & Wang T, 2006) In fact, most evidence has indicated that the problem is worsening. For example, the number of Medicare LBP patients increased by 131.7% and associated charges for non-specific LBP has increased by 387.2% since 1991. Astonishingly, the latter figure has outpaced inflationary growth for the same period by over 300%.(Weiner DK et al., 2006) Similarly, in just a two-year period from 2000 – 2002, the state of Pennsylvania demonstrated a 5.5% increase in the number of LBP patients and a corresponding 33% overall increase in related charges. Owing largely to that figure was a 59% increase in injection charges, and a 20% - 42% increase in x-ray and diagnostic imaging charges. Conversely, physical therapy related charges increased by only 2% over that same time period.(Weiner DK et al., 2006) Comparably over a 6 month period in 2006, LBP was the third most expensive disorder ($6M) billed under the University of Pittsburgh Medical Center's (UPMC) Health Plan. Only cancer ($22M) and heart disease ($9M) were more costly under total expenses paid. Aside from total cost, what makes this figure ($6M) more extraordinary is that it accounted for only 949 patients, which is only a third of the other two leading disorders. Imaging services and pharmacology utilization accounted for greater then 50% of those charges while repeat rehabilitation (largely chiropractic care) expenses essentially accounted for another 1/3rd of that total expenditure.

2.2 USUAL CARE FOR LOW BACK PAIN IN PRIMARY CARE

In order to minimize clinical variation and decrease the economic burden of LBP, clinical practice guidelines have been published in the United States and throughout the world.(Chou R et al., 2007; *Clinical Practice Guideline: Acute Low Back Problems in Adults*, December 1994; Koes BW, van Tulder MW, Ostelo R, Burton KA, & Waddell G, 2001) Generally most of these guidelines share consistent features as would be expected since most developed countries share similar evidence. In fact, most all major recognized guidelines are in agreement over recommending against bed rest and early implementation of diagnostic imaging studies unless medical red flags are present during patient examination.(Koes BW et al., 2001) Most guidelines also emphasize the early and gradual activation of patients, the discouragement of prescribed bed rest, and the recognition of psychosocial factors as risk factors for chronicity.(Chou R et al., 2007; *Clinical Practice Guideline: Acute Low Back Problems in Adults*, December 1994; Koes BW et al., 2001) Inconsistencies do exist however for recommendations regarding exercise therapy, spinal manipulation, muscle relaxants, and patient information.(Koes BW et al., 2001) It is also reported that LBP is an ailment with a relatively benign course in about 90% of patients.(Patel AT & Ogle AA, 2000) As a result, primary care physicians tend to reserve specific interventions such as physical therapy after LBP patients fail to recover from graduated or stepped activity and education.(Patel AT & Ogle AA, 2000) Based on these recommendations, a stepped care regimen is the current best practice standard in the primary care management of LBP, and represents “usual care” in the United States and throughout the world.(Chou R et al., 2007; *Clinical Practice Guideline: Acute Low Back Problems in Adults*, December 1994; Koes BW et al., 2001)

As it turns out, a stepped care approach may be sufficient for some patients with LBP, permitting them to recover without incurring significant healthcare expenditures and time off work, but certainly this is not true for the majority.(Patel AT & Ogle AA, 2000) In fact, typically only about a third of LBP cases completely resolve within a year.(Cassidy JD, Cote P, Carroll LJ, & Kristman V, 2005; Croft PR, Macfarlane GJ, Papageorgiou AC, Thomas E, & Silman AJ, 1998) This starkly contrasts with the accepted notion that LBP is a relatively benign ailment. As a result, a sizeable subgroup of patients continues on to experience chronic, recurrent episodes of LBP accounting for a disproportionately large share of overall healthcare expenditures.(Cassidy JD et al., 2005; Croft PR et al., 1998; Wasiak R, Kim J, & Pransky G, 2006) Additionally, recurrence rates for LBP within the first three to six months after onset range from 20% - 33%.(Cassidy JD et al., 2005; Croft PR et al., 1998) To demonstrate the magnitude of this problem further, data compiled during a three year period in the state of New Hampshire showed that individuals with LBP recurrence contributed significantly to overall workers compensation claims through both additional care seeking and work disability.(Wasiak R et al., 2006) Those with non-specific recurrent LBP accounted for the largest portion of total length of work disability (69%), higher medical costs (84%) and indemnity costs (71%). Similarly, U.S. Army discharge data compiled from more then 15,000 soldiers demonstrated that back conditions accounted for the greatest 5-year cumulative risk of disability (21%, 19%, and 17% for intervertebral disc displacement, intervertebral disc degeneration, and nonspecific low back pain, respectively).(Lincoln AE, Smith GS, Amoroso PJ, & Bell NS, 2002) Marine Corps data tracked through the Defense Medical Epidemiology and Defense Medical Surveillance System revealed comparable results during a 2-year period.(Huang GD, Feuerstein M, & Arroyo F, 2001) Over that time back and upper extremity diagnostic categories were among the top four sources of outpatient visits and duty limitation among enlisted Marines. Back disorders were also found to be the fifth most common cause for lost duty time.

The proposed advantage of a stepped care approach which is supported by most clinical guidelines is that it avoids undue expenditure of healthcare resources on patients with a favorable natural history.(Patel AT & Ogle AA, 2000) It is theorized that intervening with more specific therapeutic interventions early in the course of care would be less cost-effective. In addition some believe it may actually impede recovery in some patients by excessively "medicalizing" the condition. Certainly this argument has merit when one considers the seemingly immeasurable resources that are continually allocated to such services as pharmacology, imaging, and surgery. Yet, it also fails to realize the potential benefit of recommending less costly and possibly more effective physical therapy services early in the course of LBP treatment.

As such, the "stepped care" approach to LBP may benefit a portion of the population, but it is quite obvious from the aforementioned discussion that this is not the case in at least a third of all LBP patients who go on to develop symptom chronicity, and further burden our health care system.(Cassidy JD et al., 2005; Croft PR et al., 1998) Part of the reason for this may be attributed to a general lack of adherence by primary care managers to the guidelines.(Fullen BM et al., 2007; Gonzalez UV, Palacio EL, & Lopez MJ, 2003) For example, providers report that they are well aware of published guidelines and for the most part agree on the proposed recommendations. Yet they also admit the main barrier to guideline compliance is their interpretations of patient's expectations and subsequent willingness to give in to patient demands for imaging, medications or other specific interventions.(Schers H, Wensing M, Huijsmans Z, van Tulder M, & Grol R, 2001) A more plausible explanation for the limited effect sizes in the usual care approach to LBP however may be the application of a single minded strategy to a multi-faceted clinical problem. In fact it would seem unreasonable for everyone with non-specific LBP to benefit from any single intervention approach.(Kent P & Keating J, 2004) The challenge therein lies with the ability to determine the best set of interventions to match with appropriate candidates who present with a similar subset of symptoms or clinical factors.(Borkan JM & Cherkin DC, 1996; Borkan JM, Koes B, Reis S, & Cherkin DC, 1998; Bouter LM, van Tulder MW, & Koes BW, 1998; Koes BW, van Tulder MW, & Thomas S, 2006) Otherwise broad implementation of the current usual treatment regimen will do little to contain the growing epidemic of LBP.

2.3 VARIATION OF CLINICAL PRACTICE

Despite the proliferation of published guidelines and the abundance of medical evidence, variation in clinical practice is still the observed norm. This trend is argued by many to be a major contributor to why health care spending now exceeds 35% of the U.S. gross domestic product.(Bodenheimer T, 2005a, 2005b, 2005c; McCarthy M, 2003; The Commonwealth Fund, 2006) For example, large variations in the quantity of care delivered to Medicare patients has been demonstrated across different geographic areas after adjusting for age, sex, race, and locale.(Fisher ES et al., 2003a, 2003b) Differences in physician practice patterns were primarily accountable for these results which indicate that residents in the higher spending regions received 60% more care but did not have better quality or outcomes of care. Increased utilization was explained by more frequent physician visits, more frequent inpatient visits , more frequent tests and minor (but not major) procedures, and increased use of specialists and hospitals. (Fisher ES et al., 2003b) Mortality outcome results for several different conditions also demonstrated that for each incremental 10% increase in end-of-life spending, no improvement effect on the relative risk of death was observed.(Fisher ES et al., 2003a) Physician fees also reflect dramatic changes over time. Since the managed care era every 1% reduction in Medicare physician fees corresponded with a 56% increase in the volume of their services. For example, if fees were cut for coronary artery bypass surgery, thoracic surgeons recouped about 70% of their revenue loss by increasing the volume of surgeries for both single payer and private plan patients.(Bodenheimer T, 2005c) Aside from variation in quantity of services, inappropriate utilization (non-evidence based) of clinical services is also widespread. Between 2000 and 2002 almost 61% of Pennsylvania Medicare cases reviewed had an MRI for nonspecific LBP when none of the 111 study participants had notable red flags as specified by the U.S. Department of Health and Human Services Agency for Health Care Policy and Research guidelines for low-back pain.(*Clinical Practice Guideline: Acute Low Back Problems in Adults*, December 1994; Weiner DK et al., 2006) Similarly, Pearlman et al. implicated primary care physicians for being responsible for the majority of echocardiography referrals (56% versus 29% cardiology referrals) and subsequently contributing to the per capita rise (7.7%) above the Medicare sustainable growth rate (5%) for those procedures.(Pearlman AS, Ryan T, Picard MH, & Douglas PS, 2007) Complicating matters is the relationship between providers and industry. Campbell and colleagues that concluded at least 94% of all practitioners have some sort of relationship with the pharmaceutical industry.(Campbell EG et al., 2007) They report that industry is engaged in a pattern of targeting specialists, whose role as opinion leaders may be used to influence non-specialists. Therefore, inappropriate and over utilization of services is evident across all levels of health care which demonstrates a need for fair and evidenced based cost-containment strategies across the board.

Physical Therapists also contribute to this pervasive quality of care chasm through treatment inconsistency for LBP. Granted, therapists do not perform radiographic testing or prescribe medications but they do target LBP with a plethora of standardized interventions over a repeated number of visits.("American Physical Therapy Association, Guide to Physical Therapy Practice: 2nd edition," 2001) Given the current fee for service structure, it is easy for practitioners and administrators to code to for the maximum expected return per patient visit.(Lingard EA, Berven S, Katz JN, & Kinemax Outcomes Group, 2000) Since LBP is the reason that care is sought in 50% of all patients presenting to outpatient clinics, this potentially translates into a great deal of spending waste.(Di Fabio RP & Boissonnault W, 1998; Jette AM & Delitto A, 1997) Lack of adherence to current evidence is also another problem which has lead to extensive physical therapy treatment variation for LBP.(Bridges PH, Bierema LL, & Valentine T, 2007) This situation appears to be a multi-factorial problem that was most positively correlated with therapist desire for learning, their education level in terms of highest degree held, and practicality.

As indicated therapists employ a wide range of standardized treatment interventions according to the Guide to Physical Therapy Practice which includes manual therapy, therapeutic exercise, modalities, traction, and functional training.("American Physical Therapy Association, Guide to

Physical Therapy Practice: 2nd edition," 2001) Most of these interventions are accepted as standard of care however evidence from randomized controlled trials has failed to offer definitive support for either strategy.("Philadelphia Panel, Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for low back pain," 2001) In fact despite research from over 1000 randomized controlled trials, the evidence remains contradictory and inconclusive at best for many of these interventions.(Hayden JA, van Tulder MW, & Tomlinson G, 2005; Koes BW et al., 2006) A reason offered for this lack of definitive evidence is that most study designs in the literature have utilized broad inclusion criteria resulting in heterogeneous samples.(Delitto A, 2005) In this respect it is quite obvious that many of these trials are modeled in the same "one size fits all" fashion which is reflective of many of the current clinical practice guidelines for LBP. Such inconclusive information has given clinicians very little to aid them in the selection of their intervention strategies which ultimately has resulted in widespread treatment variation and less then favorable outcomes.(Li LC & Bombardier C, 2001; Mikhail C, Korner-Bitensky N, Rossignol M, & Dumas JP, 2005) Thus it is not surprising why physical therapy is not recommended by most guidelines until at least 4 weeks after onset of LBP and implementation of a stepped care approach.

2.4 ALTERNATIVES TO USUSAL CARE – THE IMPORTANCE OF PATIENT SUB-GROUPS

Attempts to identify effective interventions for individuals with LBP have been largely unsuccessful.(Hayden JA et al., 2005) Traditionally the medical model has been based on classifying individuals by a pathoanatomical source of symptoms, yet the relevant pathology is reported to be identified accurately in only 10% of all LBP cases.(Abenhaim L et al., 1995) Therefore using a traditional approach and attempting to identify a pathoanatomic source will rarely be useful in guiding effective clinical decisions.(Fritz JM, Cleland JA, & Childs JD, 2007) Complicating this matter even further is that practice guidelines generally view LBP as a homogenous entity once medical red flags and nerve root compression are excluded.(Chou R et al., 2007; *Clinical Practice Guideline: Acute Low Back Problems in Adults*, December 1994; Koes BW et al., 2001) The Guide to Physical Therapist Practice advocates that a primary goal of the diagnostic process is to classify patients based on clusters of signs and symptoms, and not suspected pathoanatomical causes.("American Physical Therapy Association, Guide to Physical Therapy Practice: 2nd edition," 2001) The Guide and other evidence-based practice entities also advocate effective subgrouping methods which ultimately direct decision making to the most effective management strategies.(Rose SJ, 1989; Sackett DL, Haynes RB, Guyatt GH, & Tugwell P, 1991) In lieu of this, a great deal of constructive work has been performed by physical therapists describing sub-groups of patients based on clusters of signs and symptoms while matching them to a particular intervention strategy which is most effective.(Binkley J, Finch E, Hall J, Black T, & Gowland C, 1993; Riddle DL, 1998; Van Dillen LR et al., 1998) The Treatment Based Classification (TBC) approach originally proposed by Delitto and colleagues in 1995 is an example of such a method.(Delitto A, Erhard RE, & Bowling RW, 1995) In fact, research has demonstrated that clinical decision making based on this classification paradigm ultimately results in better physical therapy outcomes then decisions based on alternative strategies.(Brennan GP et al., 2006; Fritz JM, Delitto A, & Erhard RE, 2003) The TBC approach described by Delitto and colleagues was intended for patients with acute exacerbations of LBP causing substantial pain and limitations in daily activities.(Delitto A et al., 1995) After screening for serious medical red flags, the system proposes using the information gathered by the patient's history and physical examination to place them into 1 of 4 basic classification categories: manipulation, specific-exercise, stabilization, and traction. The major signs and symptoms now proposed as the criteria for placing a patient into one of these categories are listed in Table 1.(Fritz JM, Cleland JA, & Childs JD, 2007) Ultimately, one should reason that it is much more probable to minimize treatment variability by accurately identifying the sub-groups of patients who are likely to benefit from the specific interventions within the TBC, and to subsequently reduce associated treatment costs with more appropriate targeting of physical therapy resources.

Table 1. Treatment based classification (TBC) with proposed criteria

Classification	***Factors Favoring***
Manipulation	Does the patient meet at least the first 2 or any three of these criteria: • Recent onset of symptoms (<16 days) • LBP with no symptoms distal to knee • Hypomobility with spring testing of spine • FABQW score < 19
Stabilization	Does the patient meet at least three of the

	following: • Positive prone instability test • Positive aberrant movements during saggital plane motion • Average SLR > 91◦ • Age < 40 years
Specific Exercise	Does the patient have: • LBP with distal symptoms • Centralization of symptoms with 2 or more movements in the same direction (ie, flexion or extension) OR • Centralize with movement in 1 direction and peripheralization with movement in the opposite direction
Traction	Does the patient have: • Presence of lower extremity symptoms and signs of nerve root compression • Absence of centralization with movement testing.
Abbreviations: LBP, Low Back Pain; FABQW, Fear-Avoidance Beliefs Questionnaire Work Subscale; SLR, Straight Leg Raise.	

2.5 TREATMENT BASED CLASSIFICATION APPROACH (TBC)

2.5.1 Manipulation category

Study results supporting spinal manipulative therapy for the treatment of LBP has been mixed over the years. (Assendelft WJJ, Morton SC, Yu EI, Suttorp MJ, & Shekelle PG, 2003;

Aure OF, Nilsen JH, & Vasselien O, 2003; Cherkin DC, Deyo RA, Battie M, Street J, & Barlow WA, 1998; Childs JD et al., 2004; Godfrey CM, Morgan PP, & Schatzker J, 1984; Goldby LJ, Moore AP, Doust J, & Trew ME, 2006; "United Kingdom Back Pain Exercise and Manipulation (UK BEAM) Randomised Trial: Cost Effectiveness of Physical Treatments for Back Pain in Primary Care," 2004) Again, the use of heterogeneous samples within studies may be a reason for the limited findings regarding its efficacy and the subsequent lack of treatment generalizability. Traditionally, classifying a patient as needing manipulation relied heavily on mobility assessments and special tests rooted in biomechanical theory which later have been shown to have poor reliability and questionable validity.(Fritz JM, Cleland JA, & Childs JD, 2007) More recent research has omitted these biomechanical assumptions and has instead focused on identifying baseline examination factors that are associated with benefiting from a manipulation intervention.(Fritz JM, Cleland JA, & Childs JD, 2007) For example, prediction studies using chiropractic manipulation interventions have demonstrated that patients with shorter duration of symptoms and the absence of leg pain are more likely to benefit.(Axen I et al., 2005; Skargren EI, Carlsson PG, & Oberg BE, 1998) Since that time, a multivariate clinical prediction rule (CPR) has been developed and validated across multi-center clinical settings to accurately identify patients who best fit a manipulation classification.(Childs JD et al., 2004; Flynn T et al., 2002)

The CPR was first developed by examining predictors of improvement defined as 50% or greater reduction in disability occurring over 2 treatment sessions in 71 patients with nonradicular LBP.(Flynn T et al., 2002) The final five predictors which constitute the rule are: current symptom duration of less then 16 days, symptoms not extending distal to the knee, a score on the work subscale of the Fear-Avoidance Beliefs Questionnaire (FABQ) of less the 19, hypomobility of the lumbar spine as assessed with posterior-to-anterior pressure, and internal rotation range of motion of at least one hip greater than 35◦. When 4 out of 5 of these factors were present, patients were highly likely to improve (+LR, 24.0) while the presence of 2 or fewer were almost always associated with a failure to improve (- LR, 0.09). What this means is that if we assume by chance that 50% of patients with nonradicular LBP will improve with manipulation, the likelihood of improvement would increase to 97% when at least 4 of the factors were present and would drop to only 9% when 2 or fewer factors were present.

The same prediction rule was later validated in a second study recently published in the *Annals of Internal Medicine*.(Childs JD et al., 2004) That study was a multicenter randomized clinical trial of 131 patients with LBP within the military health care system. Patients were categorized based on the prediction rule as +CPR (at least 4 of 5 criteria present) or -CPR (<4 criteria present), and then randomized to receive a standardized exercise program with or without manipulation. The design created 4 groups based on treatment (manipulation or exercise) and CPR status (+ or -). The validity of the prediction rule was supported because patients who were positive on the rule and received manipulation experienced better clinical outcomes than the other 3 groups. In fact, repeated measures analysis of variance on the Oswestry disability scores found a significant 3-way interaction between time, treatment, and CPR status.(Childs JD et al., 2004) At follow-up, patients who were positive on the rule and received manipulation demonstrated significant and clinically important reductions in disability and pain compared to any other group ($p<0.05$). These findings suggest that the subgroup of patients identified by the CPR is uniquely responsive to a manipulation intervention. Therefore, the value of a CPR does not only lie in its ability to accurately predict success but also in its ability to identify patients who would benefit from an alternative treatment approach.

The aforementioned prediction rule that has been developed and validated includes 5 factors, which may limit its applicability in primary care. Two factors, duration of symptoms and presence of symptoms distal to the knee, are easily obtained via medical history intake within primary care. A manuscript recently published in *BMC Family Practice* examined the accuracy of these 2 factors in identifying the sub-group of patients with LBP likely to rapidly respond to a manipulation intervention.(Fritz JM, Childs JD, & Flynn TW, 2005) A cohort of 141 patients with LBP all receiving the same standard baseline examination, 1-2 sessions of manipulation and range of motion (ROM) exercise, with post-treatment re-assessment, was studied. This restricted rule resulted in a +LR of 7.2. Although lower than the original +LR of 24, a +LR of 7.2 indicates the probability of success with manipulation increases from 45% to 86% when both factors are present. The 2-criteria rule sacrifices a small degree of accuracy for a substantial increase in ease of use in primary care and thereby demonstrates its efficacy in accurately predicting which patients should best respond to manipulative therapy.

Traditional biomechanical theories that identified patients for manipulation have also supported the need for precise treatment techniques to address specific spinal dysfunctions.(Greenman PE, 1991; Maitland G, Hengeveld E, Banks K, & English K, 2000) As the efficacy of these theories has been questioned, so have their proposed manipulation techniques.(Chiradejnant A, Maher CG, Latimer J, & Stepkovitch N, 2003; Cleland JA, Fritz JM, Whitman JM, Childs JD, & Palmer JA, 2006) There is some evidence however to support thrust manipulation techniques over non-thrust manipulation techniques.(Bronfort G, Haas M, Evans RL, & Bouter LM, 2004) Given this, there is still no preponderance of evidence for the superiority of one manipulation technique over the other possibly suggesting that the choice of a specific manipulation technique may not be as important as previously thought.(Cleland JA, Fritz JM, Childs JD, & Kulig K, 2006) In addition the effects of manipulation may not be as specific in terms of the spinal level to which it is directed.(Beffa R & Matthews R, 2004; Haas M et al., 2003) In the previously discussed CPR studies, all patients received a similar manipulation technique(Childs JD et al., 2004; Flynn T et al., 2002) performed by a trained physical therapists in a consistent manner. A recent systematic review also suggests that when given discretion to select manipulation techniques, clinicians do not necessarily demonstrate different outcomes to those studies with predefined manipulation protocols.(Kent P, Marks D, Pearson W, & Keating J, 2005) Thus it appears that the accurate identification of patients whom are most likely to respond to manipulation is far more important then the selection of a specific manipulation technique.(Fritz JM, Cleland JA, & Childs JD, 2007)

2.5.2 Specific exercise category

Treating subgroups of patients that preferentially respond to repeated end-range movements emerged several years ago with the McKenzie based approach for LBP.(McKenzie RA, 1989) Earlier evidence however demonstrated no support for this approach over non-specific approaches, but most trials were conducted with heterogeneous samples of patients.(Cherkin DC et al., 1998; Dettori JR, Bullock SH, Sutlive TG, Franklin RJ, & Patience T, 1995; Malmivaara A et al., 1995) Supporting evidence is now beginning to surface that demonstrates there is a particular subgroup of patients that respond best to specific exercise interventions.(Browder DA, Childs JD, Cleland JA, & Fritz JM, 2007)

Inclusion criteria for the specific exercise subgroup are currently based upon the presence of a centralization phenomenon and a directional preference for a specific movement upon examination.(McKenzie RA, 1989) Centralization is generally defined as occurring when a movement or position of the lumbar spine results in the abolishment of pain or parasthesia, or causes the migration of symptoms from an area more distal and lateral to the buttocks to an area more proximal or closer to the midline of the lower spine.(Aina A, May S, & Clare H, 2004) As the evidence comes in from the field and the TBC evolves over time, the inclusion criteria may broaden or even change. So far, one study has demonstrated the usefulness of the centralization finding as a classification criterion for the specific exercise subgroup.(Browder DA et al., 2007) A directional preference is defined when a movement of the lumbar spine in one direction improves pain and limitation in ROM.(Kilpikoski S et al., 2002) The second requirement for this definition to be met is that movement in the opposite direction causes those same signs and symptoms to worsen. Evidence has demonstrated that when specific exercises are matched with a person's directional preference for that movement, they have greater reductions in disability over a 2-week follow-up then those receiving unmatched or controlled treatments.(Long A, Donelson R, & Fung T, 2004) Additional research is needed to examine the full utility of using centralization and directional preference for identifying the patients most likely to respond to specific exercise interventions or to discover additional classification criteria that may exist.

The general management recommendation for patients in the specific exercise classification is to use repeated end-range movements in the direction that causes centralization of symptoms.(Fritz JM, Cleland JA, & Childs JD, 2007) At least two systematic reviews which have pooled data from 6 randomized or quasi-designed studies have examined the treatment effects the McKenzie based program, a large component of which is the application of repeated end-range movements in the direction that provides centralization of symptoms.(Clare HA, Adams R, & Maher CG, 2004; Machado LA, de Souza MS, Ferreira PH, & Ferreira ML, 2006) Results demonstrated greater reductions in pain and disability for McKenzie based treatments in the short term but not at long-term follow-up. The smaller effect sizes may be explained by the broad and heterogeneous inclusion criteria of the studies reviewed.

The treatment based classification approach has proposed three categories for the specific exercise subgroup based upon the centralization of movement (extension, flexion, and lateral shift). The most prevalent and commonly studied category is the extension subgroup of the specific exercise classification.(Fritz JM & George S, 2000) At least three different studies have demonstrated improved short-term outcomes with repeated exercise movements.(Browder DA et al., 2007; Long A et al., 2004; Petersen T, Kryger P, Ekdahl C, Olsen S, & Jacobsen S, 2002) Two of the three studies matched patients based on directional preference or centralization.(Browder DA et al., 2007; Long A et al., 2004) The third study failed to match patients with either but the greater prevalence of extension oriented patients in the general population may have led to the slightly more favorable results.(Petersen T et al., 2002) For this category, there is also some recent evidence to suggest coupling repeated end-range movements with graded mobilization to promote lumbar extension may be a more optimal intervention strategy.(Browder DA et al., 2007; Petersen T et al., 2002) The flexion subgroup proposed by the TBC approach appears to be much less prevalent and typically includes older patients with a medical diagnosis of lumbar spinal stenosis.(Fritz JM, Delitto A, Welch WC, & Erhard RE, 1998; Long A et al., 2004) As such, most research to date has focused on the effectiveness of intervention strategies for stenosis patients versus a more general flexion specific exercise category of patients.(Fritz JM, Cleland JA, & Childs JD, 2007) A recent study examined patients over age 50 with imaging evidence for stenosis and a directional preference for flexion.(Whitman JM et al., 2006) A comparison was made between a manual therapy group receiving mobilization or manipulation of the spine, exercises to address strength and flexibility impairments, and an unloaded treadmill-walking program to a flexion oriented exercise group that received repeated flexion exercises, a similar unweighted treadmill-walking program and subtherapeutic ultrasound. Better outcomes were observed in the manual therapy treatment arm perhaps suggesting that patients within the flexion subgroup may benefit more from a multi-modal intervention strategy that includes other components beyond repeated flexion exercises alone. The final and least prevalent subgroup (7%) of the specific exercise classification is composed of those with a preference for repeated end-range lateral shift movements coupled with traction (mechanical or autotraction).(Delitto A et al., 1995; Fritz JM & George S, 2000; Long A et al., 2004) When compared to nonspecific advice and massage, this strategy resulted in correction of frontal plan posture deviations know as a visible lateral shift, but failed to demonstrate any significant improvement in disability outcomes after 3 months.(Gillan MG, Ross JC, McLean IP, & Porter RW, 1998) Therefore, further research is required to identify the most effective intervention strategies in the lateral-shift specific exercise classification.

2.5.3 Stabilization category

In recent years, research has emphasized the importance of spinal muscles in restoring and maintaining spinal stability for the treatment of LBP.(Danneels LA et al., 2001; Hides JA, Jull GA, & Richardson CA, 2001; Hodges PW, Moseley GL, Gabrielsson A, & Gandevia SC, 2003; Hodges PW & Richardson CA, 1998) Several randomized controlled trials investigating the effectiveness of this claim have provided inconsistent results.(Cairns MC, Foster NE, & Wright C, 2006; Goldby LJ et al., 2006; Hides JA et al., 2001; Koumantakis GA, Watson PJ, & Oldham JA, 2005; O'Sullivan PB, Phyty GD, Twomey LT, & Allison GT, 1997) As previously suggested, the variability in inclusion criteria may seemingly imply that stabilization exercises are effective for some, but not all patients with LBP which is consistent with the TBC approach. Most of the scientific inquiry conducted to identify the stabilization classification criteria has examined the usefulness of previously proposed examination criteria in recognizing radiographic evidence of excessive segmental spinal motion.(Abbott JH et al., 2005; Dupuis PR, Yong-Hing K, Cassidy JD, & Kirkaldy-Willis WH, 1985; Fritz JM, Piva SR, & Childs JD, 2005) However, inconsistent imaging findings of spinal motions in asymptomatic subjects have discounted the validity of a radiographic approach as a gold standard for identifying a spine as unstable.(Boden SD et al., 1990; Hayes MA, Howard TC, Gruel CR, & Kopta JA, 1989) In addition, this method alone does not account for the contribution of the spinal muscles in providing core stability.(Fritz JM, Erhard RE, & Hagen BF, 1998) Furthermore, the inability of a traditional pathoanatomical approach to identify a specific structural fault is well documented in a vast majority of patients with LBP.(Abenhaim L et al., 1995) Therefore variation in traditional examination methods presented a demand for more accurate prediction factors to identify which subgroups of patient's best respond to stabilization based exercises.

In response to the aforementioned demand, research has recently been conducted attempting to identify useful classification criteria for the stabilization classification subgroup.(Hicks GE, Fritz JM, Delitto A, & McGill SM, 2005; Hicks GE, Fritz JM, Delitto A, & Mishock J, 2003; Stuge B, Bragelien Veierod M, Laerum E, & Vollestad N, ; Stuge B, Laerum E, Kirkesola G, & Vollestad N, 2004) Hicks et al analyzed a group of 54 nonradicular LBP patients given 8 weeks of stabilization exercises that specifically targeted the multifidus/erector spinae, transversus abdominus, and oblique abdominal muscles.(Hicks GE et al., 2005) They identified the following 4 factors that were predictive of improvement: age < 40 years, average straight leg raise (SLR) range of motion (ROM) greater then 91º, aberrant sagittal plane movements with lumbar ROM, and a positive prone instability test. As a result, the investigators developed a preliminary CPR for the stabilization classification subgroup that was defined as positive when 3 or more of these factors were positive. The resulting positive likelihood ration of this CPR was 4.0. This means that if a patient had a 50% chance of improving with a stabilization exercise program, a positive CPR increases this probability to 80%. Even greater usefulness was demonstrated in determining which patients who were not likely to receive minimal benefit (<= 5 points on the Oswestry Disability Index) with stabilization exercises. The four factors predictive of failure were as follows: a negative response on the prone instability test, the absence of aberrant movements during sagittal plane lumbar motion, the absence of lumbar hypermobility as assessed with posterior-to-anterior pressure, and an FABQ physical activity subscale score of 9 or greater. The negative CPR for stabilization includes the presence of 3 or more of these findings and is highly predictive of failure (+LR, 18.8). This means given an initial 25% chance of failing to respond to treatment, the presence of at least 3 of these factors upon examination coupled with stabilization exercises would increase that probability of failure to 86%. Other researchers have proposed additional criteria to identify women who are postpartum as likely to benefit from stabilization treatment.(Stuge B et al., ; Stuge B et al., 2004) It would appear that patients in this category are similar to those who would elicit a positive CPR response (younger, increased flexibility, hypermobility and with aberrant lumbar sagittal plane movements) for stabilization.

Since the proposal of the TBC, research has shifted the focus of treatment of patients in the stabilization classification from avoiding to controlling movement.(Delitto A et al., 1995; Frymoyer JW & Selby DK, 1985; Hodges PW, 2001; Hodges PW et al., 2003; Hodges PW & Richardson CA, 1996; Nachemson A, 1985; O'Sullivan PB et al., 1997; Richardson CA, 1995) This has led to some controversy in regard to the best type of exercises to provide for patients within this subgroup.(Hodges PW, 2001; Hodges PW et al., 2003; Hodges PW & Richardson CA, 1996; McGill SM, 2001, 2003; McGill SM & Cholewicki J, 2001; O'Sullivan PB et al., 1997; Richardson CA, 1995) Some evidence has stressed the importance of the deep intrasegmental spinal muscles for stabilization and specific retraining. (ie, transversus abdominus, multifidus).(Hodges PW, 2001; Hodges PW et al., 2003; Hodges PW & Richardson CA, 1996; O'Sullivan PB et al., 1997; Richardson CA, 1995) Support for this approach stems from randomized controlled trials that have found better outcomes through interventions emphasizing these specific spinal muscles compared to no treatment or multidimensional treatment programs that lack a trunk strengthening focus.(Goldby LJ et al., 2006; Hides JA et al., 2001; O'Sullivan PB et al., 1997; Stuge B et al.) Other researchers have focused their stabilization efforts on targeting the strength and endurance of the larger intersegmental muscles (ie, erector spinae, oblique abdominals, and quadratus lumborum).(McGill SM, 2001, 2003; McGill SM & Cholewicki J, 2001) Two studies have questioned the contention that specific muscle retraining is the most appropriate approach to providing stabilization.(Cairns MC et al., 2006; Koumantakis GA et al., 2005) One of those studies randomized 97 LBP patients to specific muscle retraining or conventional physical therapy.(Cairns MC et al., 2006) Both groups received tailored exercise but the difference between the two was that the specific exercise group received additional retraining for the multifidus and transversus abdominus. The researchers found no differences between groups at 12 weeks or 1 year post intervention. The other study randomized 67 patients with recurrent LBP to a specific intrasegmental spinal muscle retraining group versus a general strengthening group that targeting the larger intersegmental spinal muscles.(Koumantakis GA et al., 2005) Somewhat superior outcomes were observed for the general-strengthening group at the 8 week follow-up, but no differences were apparent at the 20 week follow-up. Therefore, although specific retraining may be superior to treatments without a well defined strengthening approach, the superiority of this intervention strategy over a program stressing general strengthening of the larger spinal muscles has not been supported by the literature.

2.5.4 Traction category

Generally there remains a lack of empirical support for traction in the treatment LBP patients.(Clarke HA et al., 2006; Koes BW et al., 2001; van Tulder MW, Koes B, & Malmivaara A, 2006) Practice guidelines and systematic reviews usually do not advocate its use.(*Clinical Practice Guideline: Acute Low Back Problems in Adults*, December 1994; "Philadelphia Panel, Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for low back pain," 2001) In 1995, authors of the original TBC approach hypothesized that there is potentially a subset of patients who would likely benefit by traction.(Delitto A et al., 1995) The defining criteria proposed for this subset was the presence of distal lower extremity symptoms, signs of nerve root compression, and the absence of centralization with repeated movement testing.

Thus far, most of the evidence on traction has involved the use of heterogeneous versus homogeneous samples whom are the most likely to benefit from it as an intervention.(Beurskens AJ et al., 1995, 1997; Harte AA, Baxter GD, & Gracey JH, 2003; Werners R, Pynsent PB, & Bulstrode CJ, 1999) In an attempt to identify a subgroup likely to benefit, one study compared mechanical traction to sham traction (< 20 of patient's body weight) and examined the following variables secondarily: age, sex, presence of radicular symptoms distal to knee, duration of episode, general health of patient, maximum traction force used, and physical therapist's belief that traction would be beneficial.(Beurskens AJ et al., 1995) Upon analysis, none of those specific subgroups demonstrated a significant improvement given the real treatment. In another recent study 64 subjects with LBP, radicular pain, and signs of nerve root compression were randomized to receive a 6 week extension oriented treatment with or without mechanical traction during the first 2 weeks.(Fritz JM, Lindsay W et al., 2007) Between-group comparisons were conducted for changes in pain, disability and fear and avoidance beliefs. In addition, baseline variables were explored in an attempt to recognize possible predictive factors in order to help define subgroup criteria for those who are more likely to benefit from traction. The group receiving traction had greater benefits in disability (adjusted mean Oswestry difference was 7.2 points) and fear and avoidance beliefs (adjusted mean difference in FABQPA was 2.6 points) but no difference was noted among the groups after 6 weeks. The two baseline variables that were associated with a greater improvement with traction treatment were peripherilization of symptoms with repeated lumbar extension movements or with a crossed straight leg raise. The findings of this study may suggest that the subset of patients who best respond to lumbar traction have radicular symptoms, signs of nerve root compression, and peripherilization with extension movements or a crossed straight leg raise.

At this time there is enough available research to suggest that the majority of LBP patients are not likely to benefit from lumbar traction.(Fritz JM, Cleland JA, & Childs JD, 2007) Therefore its widespread use in the clinic would seem entirely inappropriate. In addition, it is not reasonable to believe that current clinical decision making schema used by therapists and physicians is adequate for properly identifying which LBP should receive a traction modality.(Beurskens AJ et al., 1995) Further research is needed to determine if there a true subgroup of LBP patients that exists whom will best respond with a lumbar traction intervention and to define those clinical characteristics which will help clinicians identify those patients more accurately in practice. Until that time effect sizes in traction studies and long-term clinical outcomes are likely to be small.

2.6 THE ROLE OF FEAR AND AVOIDANCE BELIEFS IN PATIENTS WITH LBP

The Fear and Avoidance Model of Exaggerated Pain and Perception (FAMEPP) for musculoskeletal pain was initially proposed in 1983 and later modified in the mid-nineties.(Lethem J, Slade PD, Troup JD, & Bentley G, 1983; Vlaeyen JWS, Kole-Snijders AM, Boeren RG, & van Eek H, 1995; Vlaeyen JWS, Kole-Snijders AMJ, Rotteveel AM, Ruesink R, & Heuts PHTG, 1995) Applying this model to LBP, an individual's behavioral response is characterized along a continuum between two extremes. Thus a patient will either respond in a manner ranging from avoidance (maladaptive) to confrontational (adaptive). Coping is a form of adaptive behavior or response in which an individual is able to resume their normal activities over time. Others with avoidance or maladaptive behavior may tend to avoid activities out of fear that they may cause further harm or experience an increase in their symptoms. When such beliefs or fear is elevated in LBP patients they may be at risk for developing altered movement patterns, which in turn can contribute to muscle guarding and the persistence of pain and disability.(Linton SJ, 2000; Pincus T, Burton AK, Vogel S, & Field AP, 2002; Pincus T, Vlaeyen JW et al., 2002) In addition patients may focus their attention more on their pain, making it difficult to accomplish even routine daily activities.(Vlaeyen JWS, Kole-Snijders AM et al., 1995; Vlaeyen JWS, Kole-Snijders AMJ et al., 1995) Mitigating this process is then hampered by a vicious cycle of events initiated by decreased activity levels, resulting in decreased physical performance, further deconditioning, persistent pain and disability, adverse psychological consequences, and general reductions in physical health. Subsequently the recovery process is seemingly stalled or reversed and the patient is unable to break this cascade of events.

Evidence suggests that elevated fear and avoidance beliefs are one of the most important risk factors for developing chronicity.(Frank JW, Brooker AS, DeMaio SE, & et al, 1996; Hazard RG, Haugh LD, Reid S, Preble JB, & MacDonald L, 1996; Leeuw M et al., 2007; Linton SJ, 2000; Linton SJ & Hallden K, 1998; Pincus T, Burton AK et al., 2002; Pincus T, Vlaeyen JW et al., 2002) In addition, fear and avoidance is the psychosocial factor that is most predictive for return to work and disability in acute LBP patients even after controlling for various sociodemographic covariates and pain.(Fritz JM & George SZ, 2002; Fritz JM, George SZ, & Delitto A, 2001; Grotle M, Vollestad NK, & Brox JI, 2006; Staerkle R, Mannion AF, Elfering A, & et al, 2004; Turner JA, Franklin G, Fulton-Kehoe D, & et al, 2006) The correlation between fear and avoidance beliefs and disability is further supported throughout several cross-sectional studies of patients with LBP.(Crombez G, Vlaeyen JW, Heuts PH, & Lysens R, 1999; George SZ, Fritz JM, & Erhard RE, 2001; Waddell G, Newton M, Henderson I, Somerville D, & Main CJ, 1993) Further evidence suggest that fear and avoidance beliefs are present early in the course of LBP which presents a challenge to clinicians to recognize and respond with appropriate treatment.(Frank JW et al., 1996; Leeuw M et al., 2007) Otherwise these beliefs will continue to adversely impact outcomes of care. Even more compelling is that higher levels of fear and avoidance beliefs are present in an asymptomatic population suggesting the potential role for pain-related fear as a precursor to the development of LBP.(Linton SJ, 2000) Therefore, the evidence strongly suggests that elevated fear and avoidance beliefs are an important barrier to recovery even in the infancy of an acute LBP episode which emphasizes the importance of screening and targeted intervention early in the course of care.

Fear and avoidance beliefs are typically measured by the Fear-Avoidance Beliefs Questionnaire (FABQ) which is a validated self-report assessment tool developed by Waddell and associates.(Waddell G et al., 1993) The FABQ has two subscales. One is a 7 item questionnaire accessing fear and avoidance beliefs about a person's work (FABQ work scale; score range, 0 – 42). The other is a 4 item questionnaire that assesses fear and avoidance beliefs about physical activity (FABQ physical activity scale; score range, 0-24) Higher scores are more indicative of increased levels of fear and avoidance beliefs. Both scales have statistically significant correlations with disability in chronic LBP patients.(Crombez G et al., 1999) Fritz et al. demonstrated that the FABQ work scale could best be used to predict return to work in patients given acute work related LBP. (score <29; negative likelihood ratio, 0.08, score > 34; positive likelihood ratio, 3.33).(Fritz JM et al., 2001) They examined patients with average symptom duration of 5.5 days and assessed return to work at 4 weeks. Similarly Al-Obaidi et al. investigated FABQ physical activity scores in a group of Middle Eastern patients with LBP for more then 2 months who were not receiving workers compensation benefits.(Al-Obaidi SM, Beattie P, Al-Zoabi B, & Al-Wekeel S, 2005) They determined that the FABQ physical activity scores were predictive of patients not experiencing a clinically meaningful improvement after a 10 week lumbar extensor strengthening program. (FABQ physical activity score >=29; positive likelihood ratio, 3.78) However, interpreting these results is difficult because the investigators used 5 items to score the FABQ physical activity scale versus the 4 items validated by Waddell and associates.(Al-Obaidi SM et al., 2005; Waddell G et al., 1993) In a secondary analysis of patients participating in physical therapy clinical trials, George et al. further demonstrated that the FABQ work scale (cut off > 29) was the better predictor of long-term self reported disability then the FABQ physical activity scale (cut off > 14) up to a 6 month time frame. (positive likelihood ratio range 1.14 - 5.15; negative likelihood ratio range 0.30 – 0.83)(George SM, Fritz JM, & Childs JD, 2008)

Research has suggested that an intervention strategy based on the FAMEPP should educate the patient in a specific manner and appropriately address their fear and avoidance beliefs.(George SZ, Bialosky JE, & Fritz JM, 2004; George SZ, Fritz JM, Bialosky JE, & Donald DA, 2003) Patient education using the model differs from that of a traditional approach in that it de-emphasizes anatomic findings (diagnostic labels), encourages the patient to take an active role in their recovery process, and helps change their understanding of back pain from one of a serious disease state to that of a relatively common problem.(George SZ et al., 2003; Vlaeyen JWS, Kole-Snijders AM et al., 1995; Vlaeyen JWS, Kole-Snijders AMJ et al., 1995) In one recent randomized clinical trial, the effect of a fear and avoidance based physical therapy intervention was examined in patients with acute LBP.(George SZ et al., 2003) Sixty-six consecutive patients with LBP of less then 8 weeks duration were referred to physical therapy and then randomized to receive either a TBC treatment with fear and avoidance based intervention (n=34) or a standard TBC physical therapy approach without fear and avoidance intervention. (n=32). The intervention lasted for 4 weeks. Outcome measures for disability, pain intensity and fear and avoidance beliefs were collected at 4 week and 6 month follow-up periods. The fear and avoidance based intervention consisted of a validated *Back Book* educational pamphlet that covered the aforementioned principles, and a graded exercise program based on a predetermined and individually tailored quota of exercise parameters. As the patient met their exercise quota, they were given positive encouragement for behavioral reinforcement and a new quota of parameters was established. The prediction of disability at 4 weeks and 6 months was significantly improved after considering the interaction between the type of treatment and the baseline level of fear and avoidance beliefs upon entering treatment. The nature of this finding suggests that patients with elevated fear and avoidance beliefs benefit more from a fear and avoidance based intervention program. Finally, the sub-group of patients with elevated baseline levels of fear-avoidance beliefs showed greater improvements in clinical outcomes when treatment was based on the fear-avoidance model, while patients with lower baseline levels of fear avoidance beliefs benefited more from the TBC treatment alone. The results of this particular trial support the need to risk-stratify patients and individually tailor treatments according to validated baseline fear and avoidance measures. The compelling evidence discussed in this section made it imperative that the results in this study were at least controlled for a patient's baseline fear-avoidance levels.

2.7 COST EFFECTIVENESS

Public and private concern about the rising costs of medical care coupled with inadequate resources has stimulated debates over a variety of proposed solutions. Medical professional societies have developed and distributed guidelines to help clinicians provide the most appropriate services.(Eisenberg JM, 1989) In addition state and federal governments hoping to reduce their resource outlays have focused more of their attention on research explaining clinical practice variations.(Davis K et al., 2007; Eisenberg JM, 1989) Furthermore, private payers ardently examine medical service utilization patterns as a top priority for implementing cost control decisions.(Weiner DK et al., 2006) While it is well accepted that these institutions may often diametrically oppose one another, there appears to be some general consensus among them on the need to cut costs. Underlying this need is a shared optimism that medical care can be made cost-effective. To many, the term cost-effectiveness implies that less money should be spent.(Doubilet P, Weinstein MC, & McNeil BJ, 1986; Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, & Stoddart GL, 2005; Eisenberg JM, 1989) In reality, cost-effectiveness means that the use of such resources is made more efficient to obtain more value for the same expenditure. In order to achieve this, some have suggested that health policy experts must expand from broad systems-level considerations to include questions on specific clinical policy.(Eisenberg JM, 1989)

The crux of the aforementioned situation lies in how those policy decisions are best determined and the influence that has on clinical autonomy and subsequently patient care. However, almost all providers should agree that there exists a point at which the extra money spent for limited clinical outcomes is not worthwhile and therefore represents inappropriate practice. Thus it should be important to realize that resources which are inappropriately misspent could be devoted to other types of medical care which is more likely to achieve greater benefits, or to some other important societal purpose. This construct is better known as an "opportunity cost."(Detsky AS & Naglie IG, 1990; Heyne P, Boettke P, & Prychitko D, 2003) Anytime an individual or group decides to spend a dollar for a particular item or service, they have foregone the opportunity to use that dollar for something else. Accepted practice for addressing these issues relies on a thorough, legitimate and systematic process of economic analysis. Critical assumptions of this process are that choices must be made between alternative uses of limited resources, and that decisions are based on both costs and outcomes.(Detsky AS & Naglie IG, 1990; Drummond MF et al., 2005; Eisenberg JM, 1989; Gold MR, Siegel JE, Russell LB, & Weinstein MC, 1996; Weinstein MC & Stason WB, 1972) In the final stage one must understand that the purpose of economic analysis is not to decide on any one alternative, but to simply demonstrate the consequences of allocation decisions to best inform those who are responsible for policy and clinical choices.(Drummond MF et al., 2005)

After an intervention has demonstrated its efficacy under optimal protocol situations, and it effectiveness in real world contexts, its efficiency or cost effectiveness should be studied as compared to a standard alternative to determine the value of any added costs of treatment. This process considers both the effectiveness of the health care intervention and the resources required to deliver it.(Detsky AS & Naglie IG, 1990) It is important to note that the term cost effectiveness can explicitly be used in two different ways: one as a broad generic term to include "cost-utility" and "cost benefit" analysis; and as a specific type of health care economic analysis that measures outcomes in units such as "life years gained" or "number of diseases averted". An intervention is said to be efficient or cost effective when it achieves its maximal increment in health benefits for a target population for a fixed amount of resources.(Detsky AS & Naglie IG, 1990)

2.7.1 Taxonomy of economic analysis

There are five basic types of cost studies, which differ according to how they include and measure effectiveness.(Drummond MF et al., 2005; Eisenberg JM, 1989; Homik JE & Suarez-Almazor M, 2004) See Table 2 below:

Table 2. Types of cost effectiveness analyses

Type of Analysis	Cost Measure	Benefit Measure
Cost Minimizing	$	None
Cost Consequence	$	Multiple
Cost Benefit	$	$
Cost Effectiveness	$	Clinical State
Cost Utility	$	Utility (e.g. QALY)

2.7.1.1 Cost minimization (cost identification)

Cost minimization analysis otherwise known as cost identification analysis simply addresses the question of *what is the cost*.(Eisenberg JM, 1989) As the name implies, these studies are most useful when used to “identify” the lowest cost of alternative treatment or diagnostic strategies. Because of this, cost minimization analysis is frequently carried out to quantify the economic burden of a disease or its treatment. However when conducting cost minimizing studies, an implicit assumption is made that the efficacy of the different strategies under investigation is broadly equivalent, so that the difference between them reduces strictly to a comparison of costs.(Drummond MF et al., 2005; Eisenberg JM, 1989) This equivalency assumption between alternative strategies is often violated as it is typically based on inadequate previous research or professional opinion.(Drummond MF et al., 2005) Therefore the limitation of a cost minimization study is that it does not evaluate what expenditures mean in terms of overall health outcomes.(Eisenberg JM, 1989)

2.7.1.2 Cost consequence

Cost consequence analysis is a disaggregated type of study that makes few assumptions and places a relatively greater burden on the consumer of the analysis.(Gold MR et al., 1996) Cost consequence analysis methods simply revolve around enumerating both cost and consequences of competing medical strategies and listing them separately. As such, no attempt is made by the investigators to measure health outcomes in a similar metric, nor do they indicate the relative importance of those outcomes.(Gold MR et al., 1996) This judgment is left open for the user of the study. In essence cost consequence analysis is based on the premise that the intended audience integrates the final incongruent list of pros and cons to make their own value tradeoff decisions. Depending on the user of the information, it is quite clear how varied and biased interpretations of cost consequence research can easily be made which does not does provide a sound basis for making the most informed health economic decisions.

2.7.1.3 Cost effectiveness

Cost effectiveness analysis goes well beyond just comparing the costs of interventions. It is a method that attempts to evaluate those expenditures in relation to the health outcomes produced. In this case the effectiveness of the treatments is not assumed to be equal unlike cost-minimization analysis. Consequently costs are estimated per unit of improvement; e.g. cost per life saved, cost per reduction in pain, costs per stroke prevented.(Homik JE & Suarez-Almazor M, 2004) As such, results are expressed as incremental costs versus incremental outcomes. This means the incremental analysis is the net difference in costs divided by the net difference in outcomes. Therefore the formula for incremental cost effectiveness ratio (ICER) is as follows:

Cost A – Cost B / Effect A – Effect B.

After plotting individual intervention ICER's, a cost effectiveness plane is normally produced with a corresponding slope. The slope represents efficiency frontier with any point along it representing the additional costs along with the additional effects that can be purchased. Strategies first must be determined if they are simply dominated by other strategies having both lower costs and greater therapeutic effects. In the cost effectiveness plane in Figure 1 below, *strategy* ***B*** can be eliminated because it is both more costly and less effective (to the left of the efficiency frontier) then *strategy* ***A.*** The next step is to examine the graph for evidence of extended dominance. This means that a linear combination of other strategies can produce greater benefits at lower costs. In Figure 1, *strategy* ***C*** is more expensive and less therapeutic then any strategy in the red-bounded region. After excluding strategies based on simple and extended dominance, ICER's are recalculated for the non-dominated treatment options so that they may be compared. In the case depicted in Figure 1, an administrator would then have to consider the trade-offs between purchasing the more costly, yet more effective *strategy **D*** over *strategy **A***. Often, a budgetary constraint influences this decision. A benefit of cost effectiveness analysis is that it can compare different programs across various disciplines as long a consistent outcome measure is used.(Drummond MF et al., 2005) Studies that use different outcome measures may have to be compared via cost utility or cost benefit analysis whereby a consistent qualitative adjustment can be made to standardize those scores.(Drummond MF et al., 2005; Gold MR et al., 1996; Homik JE & Suarez-Almazor M, 2004)

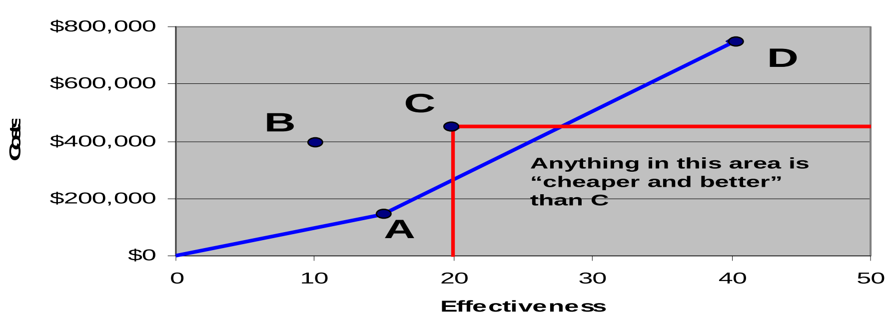

Figure 1. Cost effectiveness plane with efficiency frontier demonstrating simple and extended dominance

2.7.1.4 Cost utility analysis

Cost utility analyses are often viewed as extensions of cost effectiveness analyses.(Eisenberg JM, 1989; Gold MR et al., 1996) The purpose of this type of analysis is to compare the costs of two or more interventions in relation to the societal value of the resulting health outcomes.(Homik JE & Suarez-Almazor M, 2004) Outcomes computed in cost effectiveness studies can be converted to a utility measure which is a value score that society or individuals are willing to assign to a specific health status.(Homik JE & Suarez-Almazor M, 2004) The point of utility measurements is that a time period spent in perfect health is not equivalent to the same amount of time spent in limited health with mild, moderate or severe conditions. In other words the utility of an outcome is not the same as the outcome itself.(Drummond MF et al., 2005) Unlike cost effectiveness analysis, cost utility analysis allows researchers and economists to make comparisons across various diseases, health states, and interventions once outcomes are converted to a common denominator.

2.7.1.5 Cost benefit analysis

Cost benefit analysis compares two or more interventions, measuring both costs and outcomes in dollar or other commensurate currency units of measurement.(Eisenberg JM, 1989; Gold MR et al., 1996; Homik JE & Suarez-Almazor M, 2004) This valuing of the outcomes or consequences is a unique and distinct feature of the cost benefit analysis approach.(Table 3) The importance of expressing outcomes in a common currency unit is that it allows researchers and policymakers the opportunity to compare health programs economically across a wide variety of social needs such as: housing, defense, transportation, welfare, education, etc. While this seems logical, many would argue that the absurdity of affixing a cost value on human life makes this approach irrelevant.(Drummond MF et al., 2005; Gold MR et al., 1996) In some respects this is a very valid point, however societies and policy makers routinely perform similar valuations by trading off health care objectives against other benefits.(Drummond MF et al., 2005) For example when society votes to determine the amount of funding for key life saving programs, they are implicitly making this judgment.

2.7.2 Quality of economic evaluation in health care

Frequently, many who assert that a specified intervention is cost effective fail to support that contention with appropriate documentation of associated costs and benefits.(Weinstein MC & Stason WB, 1972) In fact even the criterion for term “cost effective” is often wrongly implied or not clearly identified in the medical literature.(Doubilet P et al., 1986) According to economists, the defining criterion for one strategy being more cost effective then another is if it meets any of the following conditions (a) less costly and at least as effective; (b) more effective and more costly, its additional benefit being worth its additional cost; or (c) less effective and less costly, the added benefit of the rival strategy not being worth its additional cost. Conditions (b) and (c) may also be further defined to a willingness-to-pay (WTP) level (i.e. strategy **X** is cost-effective as long as one is willing to spend at least $50,000 per life year gained). Often articles cite that an intervention is cost-effective by simply equating that statement with cost savings only and without any consideration to the benefits involved. Therefore selecting the cheapest strategy would likely be the interpretation made from these studies. Unfortunately, reliance on cost-minimization alone would ultimately lead to the exclusion of many widely accepted programs (such as treatment for severe hypertension) that do save money but provide a reasonable benefit at an accepted cost. Conversely, other trials claiming cost effectiveness have been routinely performed examining only effectiveness of interventions without any consideration towards monetary costs. Another way the term cost effectiveness is misused is when it is equated with cost savings with at least an equal (or better) health outcome. Based on this stringent criterion, any programs that do not save money would be excluded as well as programs that reduce costs significantly but lead to perhaps a smaller degree of effectiveness.

Williams described the essential elements of health care economic evaluations several decades ago.(Williams A, 1974) More recently Drummond and colleagues refined the methodological areas generally agreed upon by economists.(Drummond MF, Brandt A, Luce B, & Rovira J, 1993) Despite these recommendations, reviews of published studies have demonstrated significant gaps in the quality of work and have lead to misleading interpretations.(Adams ME, McCall NT, Gray DT, Orza MJ, & Chalmers TC, 1992; Balas AE et al., 1998; Jefferson T et al., 1998; Neumann PJ, Stone PW, Chapman RH, Sandberg EA, & Bell CM, 2000; Udvarhelyi S, Colditz GA, Rai A, & Epstein AM, 1992) Udvarhelyi et al.(Udvarhelyi S et al., 1992) reviewed 77 published articles from 1978 to 1987 and determined that only three articles (4%) adhered to all of the following six fundamentally agreed upon principles of economic reporting: an explicit statement of analysis perspective should be made, benefits of comparable strategies should be explicitly described, costs and component costs should be specified within the analysis, incorporation of discounting across both costs and benefits over extended time frames should be identified, sensitivity analysis should be performed to test important assumptions and uncertainty, and a proper summary measurement of efficiency must be included such as a cost-benefit ratio or incremental cost-effectiveness ration (ICRE) . In their review, only 18% of the papers stated the perspective from which the analysis was performed. Cost reporting varied considerably and 4% of the articles reviewed failed to report any costs at all. Discounting was appropriately applied in less then 50% of the articles while only 30% acknowledged performing sensitivity analysis to test the robustness of their conclusions. In addition, only 10 of 77 articles

(13%) applied an accurate summary of cost-effectiveness or cost-benefit measurement. Similarly Balas et al.(Balas AE et al., 1998) concluded that statements regarding costs without substantiating data are made habitually in clinical trials. They reviewed 181 articles from 1966 to 1995 under the MEDLINE search medical subject heading (MeSH) terms *costs and cost analysis, cost control, cost of illness, cost savings, or cost-benefit analysis.* Their results demonstrated that only 97 articles (53.6%) included the actual costs of the intervention in the analysis. In another systematic review, Adams et al.(Adams ME et al., 1992) analyzed 121 out of 50,000 randomized controlled trials that were published from 1966 to 1988. After assessing those 121 articles, they calculated a mean quality of research score of .32 (scale 0 – 1) and a mean completeness of assessment score of .52 (scale 0 – 1) further demonstrating a lack of quality in medical economic reporting.

Underscoring a need for more consistency and accuracy in reporting, some medical journals have adopted guidelines for publication for their authors and reviewers.(Drummond MF & Jefferson TO, 1996; Gandjour A, Jefferson T, Demicheli V, & Vale L, 2002) Despite these changes, only mild improvements have been made in guideline adherence and overall publication quality.(Jefferson T & Demicheli V, 2002; Jefferson T et al., 1998; Neumann PJ et al., 2000) This suggests the need to train reviewers in economic analysis and to enforce more aggressive third party auditing procedures. However, as progress is being made in the general medical literature, no standardization initiatives have yet been proposed in the rehabilitation science fields. More specifically, no guidelines for cost-effectiveness research exist in the physical therapy literature at this time.

2.7.3 Cost effectiveness analysis of physical therapy as an intervention strategy for low-back pain (LBP)

The efficiency of physical therapy as an intervention strategy for LBP has been examined in several studies, but little evidence exists to suggest it is a cost-effective treatment option over other alternative approaches. However interpretation is limited mainly because of the poor quality of research methodology. In many cases claims about cost-effectiveness are made in the absence of a full economic analysis and are simply not substantiated by the data. For example, Kominski et al(Kominski GF, Heslin KC, Morgenstern H, Hurwitz EL, & Harber PI, 2005) performed what they labeled as an economic analysis comparing 4 different treatment strategies for LBP: physician, physician with physical therapy, chiropractic, and chiropractic with physical medicine modalities. They concluded that physical therapy combined with physician treatment and chiropractic care combined with physical modalities does not appear all that cost-effective. However the researchers simply performed a cost-minimization study without any description of the analysis perspective or component costs involved. In addition, they failed to account for benefits because they stated no difference between the opposing strategies existed in the literature. Torstensen et al.(Torstensen TA et al., 1998) concluded that both graded activity and conventional physical therapy were more efficient then a program of self-exercise in chronic LBP patients. They examined self-reported patient satisfaction as benefits and lost productivity wages secondary to work absences as cost measures. Aside from indirect productivity costs, no other direct expenses were accounted for including the costs of the interventions. Furthermore, ICRE's were not explicitly stated in the results of the study. Skargren et al.(Skargren EI et al., 1998) compared costs and outcomes separately between physical therapy and chiropractic care. They explicitly computed directs costs associated with both types of treatments, and additional health care services over a 12-month period. They used the Oswestry measure for disability to assess patient outcomes. The physical therapy group sought fewer ancillary services and had fewer recurrences. However, no statistically significant differences were found for overall outcomes or costs. Again, aside from comparisons of groups, no ICRE's were computed in the results limiting the interpretability of cost-effectiveness from the findings. Other studies have done a reasonable job applying the principles of economic analysis but have deviated slightly from traditional statistical analysis methods. For example, one study performed a cost-utility analysis of routine physical therapy treatment to physiotherapy evaluation and advice only.(Rivero AO, Gray A, Frost H, Lamb SE, & Stewart SB, 2006) Instead of computing ICER's to determine the additional cost per QALY (quality adjusted life year) for the opposing strategies, they simply compared the costs and the utility levels thereby showing no differences between the groups. They then concluded that the lack of significance renders the higher out of pocket expenses for routine physical therapy less cost-effective. More recently, another study performed a cost-effectiveness analysis of general practitioner care (GP) and physical therapy.(Luijsterburg PA et al., 2007) The researches initially planned a cost-utility study a-priori as they collected QALY's. After determining the utility values were not significant between the two groups at comparison, they then switched to a cost-effectiveness analysis using global perceived effect (GPE) as the incremental outcome measure. They finally computed an ICER of 6,244€ per additional GPE gained for the physical therapy group suggesting that GP care alone was a more cost-effective strategy. These results are very misleading not only because the deviation from protocol but also because the physical therapy arm consisted primarily of exercise authorized by the GP's. In addition, therapists were not allowed to perform manipulative therapy.

Another reason for the limited evidence supporting the cost-effectiveness of rehab related interventions is the frequent incorporation of heterogeneous samples for the physical therapy treatment arm. As has been previously discussed, non-homogenous samples in trials examining interventions for LBP have often produced limited effect sizes. It would be natural to assume then that a "one-size fits all" treatment strategy would have limited efficiency when analyzed in economic research. Whitehurst et al.(Whitehurst DG et al., 2007) compared usual physical therapy care to a brief behavioral pain management program (BPM) in LBP patients with < 12 weeks duration. They found that the BPM was more cost-effective at a WTP threshold of 10,000 pounds per QALY gained. The non-specific physical therapy arm actually had higher effectiveness but it also had much higher costs. Taylor et. al.(Taylor RJ & Taylor RS, 2005) developed a decision analysis model examining spinal cord stimulation (SCS) to a general non-surgical management program for failed back surgery syndrome. The investigators determined that the SCS dominated the opposing generic strategy with both better costs and better effectiveness. Niemisto et al.(Niemisto L et al., 2005) performed a cost-effectiveness analysis comparing a combined program of non-specific manipulation, stabilization and a physician consult to a physician consult alone for chronic LBP patients. The combined group demonstrated a 1 point improvement in the visual analog scale (VAS) rating over the consult group. However, the incremental cost per unit change in VAS was $512 leading the investigators to suggest that a physician consult alone is more cost-effective in minimizing disability and improving quality of life. Finally, the United Kingdom back pain exercise and manipulation (UK BEAM) trial was a cost-utility study comparing four common approaches to LBP treatment within physical therapy.("United Kingdom Back Pain Exercise and Manipulation

(UK BEAM) Randomised Trial: Cost Effectiveness of Physical Treatments for Back Pain in Primary Care," 2004) The study compared a "best care" approach consisting of active management and the *Back Book* educational handout to three other approaches: an exercise program, a spinal manipulation program and a combined program of exercise and manipulation. The investigators concluded that spinal manipulation is a more cost effective addition to "best care" for back pain at a WTP threshold of 10,000 pounds. The BEAM Trial is one of the only complete cost-utility analyses to support manipulation as an intervention. It is important to understand that neither of the aforementioned studies made any effort to reduce variation in the physical therapy treatment strategies by matching specific evidence based interventions to homogenous samples.

Recently the literature has supported a cost-savings when treatment groups are matched to evidence-based guideline interventions. Fritz et al.(Fritz JM, Cleland JA, & Brennan GP, 2007) compared an active treatment guideline adherent group to a passive treatment (primarily modality driven) non-guideline adherent group. They demonstrated that patients receiving adherent care had fewer office visits and lower charges. In addition they showed a significant improvement in disability ratings when compared to the non-adherent group. Although a formal cost-effectiveness analysis was not undertaken, this is preliminary evidence to suggest that matching subgroups of patients to interventions may ultimately lead to improved efficiency through reduced clinical variation. No other cost-effectiveness analyses have been performed examining matched physical therapy treatment subgroups at this time. The effectiveness of the TBC approach has been documented but its efficiency as compared to alternative treatments within physical therapy has yet to be determined. If the cost-effectiveness of the TBC could be established, then appropriate steps may be taken to improve adherence and reduce clinical variability across a broad spectrum. This would allow physical therapy to "put its best foot forward" in future cost-effectiveness investigations comparing it to other medical alternatives.

3.0 METHODS

3.1 EXPERIMENTAL DESIGN

The experimental design of this study consisted of two phases: (1) a cost-minimization analysis to compare the associated direct health care costs for LBP patients treated under a TBC physical therapy approach (on-protocol) to patients receiving a usual physical therapy care approach (off-protocol) and (2) a decision analysis model based upon the results of the primary analysis in order to better access the overall cost-effectiveness of the TBC. The study was a retrospective cohort analysis conducted from a payer perspective over an 18 month time horizon. Data was extracted from the clinical outcome and financial databases maintained by the Centers for Rehabilitation Services (CRS), a non-profit multi-center outpatient orthopedic physical therapy subsidiary of the University of Pittsburgh Medical Center (UPMC), an integrated healthcare delivery system, and the UPMC Health Plan, a non-profit health insurance company, another integrated subsidiary of UPMC. The data collection period for this investigation ranged from October 15th, 2007 to April 15th, 2009. Two additional months were added to this time period to allow for all associated charge claims to be settled. Preliminary results from this study should contribute to a much larger Cost Effective / Cost Utility Analysis to access the overall long-term impact of the current TBC approach. This study was reviewed by the University of

Pittsburgh's Institutional Review Board (IRB) and designated as "exempt" under section 45 CFR 46.101 (b)(4) for medical retrospective chart review with an honest broker.

3.2 UPMC LOW-BACK INITIATIVE (LBI) BACKGROUND

In 2005, the University of Pittsburgh Medical Center (UPMC) Health Plan adopted a Low-Back Initiative program (LBI) to examine clinician treatment behavior, along with cost and disability outcomes for all back pain patients seeking care within their network. As part of the LBI, the plan mandated the use of a web-based reporting tool or minimum data set (MDS) by all of its' physical therapists within each of the 42 UPMC CRS clinics in Southwestern Pennsylvania. The MDS contains critical demographic, historical, objective and treatment information that is necessary to examine if therapists are treating LBP patients according to the TBC (on-protocol) or not (off-protocol). A paper based version of the MDS variables along with their definitions can be viewed in Appendix A. The MDS was institutionalized in 2006 and training was provided by the CRS executive and senior clinical staff to ensure the reporting clinicians knew how to interpret and fill in the instrument. Reporting of MDS data actually involves two separate and sequential processes: 1) therapist collection of required data points upon initial patient examination for paper record, and 2) entry of collected MDS information into the CRS web-portal which then transfers the data to a composite server. As of October 2007, only 17% of all MDS's were actually uploaded on the web portal. The remaining 83% of the web-portal MDS's were missing information from at least one important variable within the data set. Such a high level of incomplete data would have undoubtedly prevented any meaningful data analysis in this study.

Prior to this period, investigation by Landry et al.(Landry MD & Sibbald WJ, 2002) demonstrated five different educational based strategies to change provider practice behavior: (1) Academic Detailing (2) Audit and Feedback (3) Local Opinion Leaders (4) Reminder Systems and (5) Printed Material. Overall, those five strategies have varying degrees of evidence that supports their use. However, of those five strategies, auditing and feedback appear to be the most commonly used technique and one that yields the most statistically significant results. Based on this evidence, the primary investigator (PI) in this study and his advisor (CRS Director) determined that a weekly CRS Quality Assurance surveillance program along with internal incentives would be the most appropriate intervention to attempt to improve compliance with the web-based MDS reporting.

In 2007, the PI developed a software reporting program that identified specifically what critical variables were missing by patient "dummy" account number and therapist. The program was manually validated to ensure accuracy and later cross-checked with similar internal auditing reports developed by the UPMC health plan. Each week beginning Oct 15th, 2007 a report was generated to determine specifically what data was missing from all incomplete cases. These reports were immediately sent to the CRS Quality Director who proceeded to email the providers with incomplete cases in an effort to have them fill in the missing data. CRS internal policy provided for direct email or telephone contact from the Director if a therapist failed to complete a missing case after a period of 4 weeks. This audit / feedback surveillance program has continued to operate in this manner since its inception. No monetary incentives or penalties were ever used during this period. However, part of the Directors correspondence conveyed the desire from CRS to switch to a new operating web-based interface which was perceivably much more cumbersome and time-consuming. This provided somewhat of an internal incentive for the therapists to upload the missing data from the patient records to the web-based portal. Additionally it is possible that the providers were also partly motivated by the fact that upper level management was observing their actions.

By January 11th, 2008 the compliance rate for completion of the web-portal MDS's was at 90.8% and by mid-February, 2008, the compliance rate for completion exceeded 97%. Never has this completion rate dipped below 90.0% as a whole, demonstrating the effectiveness of the surveillance program. It should be noted again that the clinicians did not arbitrarily assign missing MDS data post-hoc as a result of surveillance reporting, but indeed had obtained all missing data points at the time of the initial exam for the paper-based records. The process issues of inputting the data actually revolved around which data points specifically were required to be uploaded into the web-portal, and whose responsibility (therapist or administrative staff) it was to input those variables. This means that the majority of the MDS data (> 90%) was actually available when this study's proposed cohort collection period began (Oct 15th, 2007), but was just not uploaded into the web-server which is an entirely separate and exclusive process as stated above. This problem was determined by the CRS Quality Director and subsequently verified by the CRS Director of Rehabilitation Services. Surveillance reporting now occurs every 2-3 weeks but no reverse trends have been noted with the less frequent monitoring. It is quite likely that the majority of the therapists now realize what additional computer based documentation requirements are expected of them and understand the priority those have in helping the LBI succeed. As a result, a much more complete electronic data set should contribute to improved accuracy in this study's analysis and in subsequent investigations.

3.3 SUBJECTS

The study sample was identified from the clinical outcomes database maintained by CRS. All patients presenting with a new episode of care and assigned 1 of the 27 ICD-9 diagnostic codes defined under UPMC's Low Back Initiative (LBI) was screened for study inclusion. (Table 3) In addition, all 42 outpatient UPMC CRS clinics in Southwestern PA participated in the LBI and their records were subsequently extracted and reviewed. A new episode of care was defined as the time from the date of the patient's initial physical therapy evaluation to their last physical therapy visit. If no visits occurred for more then 60 days from the subject's last physical therapy visit, the episode of care was considered complete. All new episodes of care in this study occurred from October 15th, 2007 to October 14th, 2008.

From October 15th, 2007 to October 14th, 2008 the CRS clinical outcomes database identified 942 new LBI episodes of care eligible for inclusion in this study. Aside from critical Minimum Data Set (MDS) variables (Appendix A) used to determine the on/off protocol status for the TBC, the database also includes the patient's age, gender, Oswestry Disability Index scores, Numeric pain scale scores, Fear-avoidance behavioral scores (both work and physical activity subscales), types of interventions utilized, patients history of comorbidities, the dates of physical therapy service and whether the patient had sought medical care in the past for their LBP.

Table 3. UPMC LBI ICD-9 codes with description

ICD_ID Code	**ICD Diagnosis Description**
721.3	Lumbosacral Spondylosis without Myelopathy
722.10	Displacement of Lumbar Intervertebral Disc without Myelopathy
722.32	Schmorl's Nodes of Lumbar Region
722.52	Degeneration of Lumbar or Lumbosacral Intervertebral Disc
722.93	Other and Unspecified Disc Disorder of Lumbar Region
724.02	Spinal Stenosis of Lumbar Region
724.2	Lumbago
724.3	Sciatica
724.5	Backache, Unspecified
724.6	Disorders of Sacrum
724.70	Unspecified Disorder of Coccyx
724.71	Hypermobility of Coccyx
724.79	Other Disorders of Coccyx
738.5	Other Acquired Deformity of Back or Spine
739.3	Nonallopathic Lesions of Lumbar Region, not Elsewhere Classified
739.4	Nonallopathic Lesions of Sacral Region, not Elsewhere Classified
846.0	Lumbosacral (joint) (ligament) Sprain
846.1	Sacroiliac (ligament) Sprain

846.2	Sacrospinatus (ligament) Sprain
846.3	Sacrotuberous (ligament) Sprain
846.8	Other Specified Sites of Sacroiliac Region Sprain
846.9	Unspecified Site of Sacroiliac Region Sprain
847.2	Lumbar Sprain
847.3	Sprains and Strains of Sacrum
847.4	Sprains and Strains of Coccyx
756.12	Spondylolisthesis
756.11	Spondylolysis, Lumbosacral Region

3.3.1 Inclusion criteria

The following inclusion criteria were used to determine eligibility for the study. Potential candidates must have satisfied all criteria, but consent to participate was not be required due to the type of exemption that was approved through the IRB review process.

1. All patients newly referred to physical therapy at CRS with any of the diagnostic codes related to the lumbosacral spine listed in Table 3.
2. 18 – 65 years of age

3.3.2 Exclusion criteria

Patients were excluded from participation if they met any one of the following exclusion criteria:

1. Presence of any medical “red flag” for a serious spinal condition (e.g., cancer, compression fracture, osteoporosis, infection, etc. as determined by the initial therapist evaluation and their subsequent decision not to collect critical MDS data on those subjects)
2. Current pregnancy
3. Prior lumbar spine surgery
4. Non-English speaking patients were already excluded from the UPMC LBI

The reason(s) for a patient’s ineligibility was monitored and recorded and eligibility rates were determined.

3.4 DATA EXTRACTION

For this study, MDS data from October 15^{tth}, 2007 to October 15th, 2008 was extracted by an honest broker in order to encrypt patient sensitive information. Afterwards, the honest broker provided two separate databases to the PI with one containing all patient exam / intervention information and the other one containing important patient history information. The PI then integrated both of these data files with the validated SAS programming as previously discussed regarding the CRS MDS compliance above. In this manner, the data file integration programming matched a subject’s initial exam data with the historical data so that all information analyzed was from a subject's first visit or initial evaluation only. Given the degree of effectiveness in previous trials, it was reasonable to assume that significant cost-savings could alone be generated based upon appropriate classification and treatment upon entry to physical therapy. Afterwards, separate programming files were then used to determine and report which subjects were on-protocol (TBC adherent) or off-protocol (non-TBC adherent or usual care) based on the cohort definitions listed in Appendix B.

3.4.1 Programming to determine TBC protocol status

As indicated above, the PI developed the programming code for determining which subjects were treated on versus off protocol based upon their initial presentation to physical therapy. This programming was based upon the cohort definitions listed in Appendix B for 3 of the 4 major TBC classifications. Furthermore, this programming was dependent upon the variables contained within the MDS.(Appendix A) It should be noted that the traction classification was excluded from this study due to a lack of definitive evidence in order to define this subgroup. The Specific Exercise programming was further broken down to differentiate Flexion Specific versus Extension Specific candidates. Due to a general lack of prevalence and scientific understanding, the MDS does not contain a specific variable to identify lateral shift candidates at this time. Therefore, this subgroup was not defined in this study. As discussed above, the custom programming determined which cases were potential TBC candidates and who was treated on versus off-protocol within those three major classifications. To ensure accuracy, these programs were manually validated by the PI. In addition, they have served as the benchmark for the UPMC Information Technology (IT) department to develop their programming as reporting tools for future LBI policy decisions. In that respect, both the programs from this study and the UPMC IT department have been cross-validated with each other and have yielded identical results over multiple trials.

Since March 24th, 2008, the cohort analysis programs have also been used to assist the CRS QA office through biweekly reporting. The goal of this was to enable the QA director to identify therapists who were failing to treat patients within the LBI according to the TBC. The director's goal was provide constructive but non-punitive feedback to those clinicians via email as an indicator that the UPMC Health Plan was monitoring their treatment behavior. In the long-term this will be a method for the Health Plan to reduce treatment variation for LBP by improving compliance with the current evidence in the literature. The scope of this however is beyond the aims of this particular study, but it does attest to the accuracy of the data extraction and reporting instrument which was used in this investigation.

3.4.2 Direct health care and physical therapy costs extraction

Since this study was performed from a payer perspective, the focus of the analysis was primarily on direct health care costs. In traditional cost analysis, the term "direct" generally refers to those changes in resources attributable to the intervention or treatment regimen.(Gold MR et al., 1996) Thus direct costs include the value of all the goods, services, and other resources that are consumed in the provision of an intervention or in dealing with the side effects or other current and future consequences linked to it.(Gold MR et al., 1996) Although these costs are often thought of as involving a monetary transaction, it is truly the use of the resource that defines the direct cost. Direct costs encompass then all types of resource use, to include the consumption of professional, family, volunteer or patient time.(Gold MR et al., 1996) These can subsequently be subdivided into direct non-health care costs and direct health care costs. Direct non-health care costs can include such items as patient child care costs associated with appointments, patient travel expenses associated with appointments, and the time spent for family members or volunteers to assist the patient with their health care. Direct health care costs are therefore limited to the costs of tests, drugs, supplies, health care personnel, and the medical facilities associated with the patients care. In many respects, computing costs such as those associated with the medical facilities involves some sort of consistent accounting or allocation method. In this study, UPMC served as the provider or facility and its subsidiary (UPMC Health Plan) served as the payer. Since the provider ultimately served as its own payer, this negated the need for cost allocation accounting on the PI's behalf because all concerned direct health care costs were ultimately reflected in the Health Plan's charge data.

As indicated above, the initial data collection period to establish the TBC cohorts spanned a 12 month period from October 15th, 2007 to October 14th, 2008. Downstream cost data for this sample was collected over an additional 6 month period until April 15th, 2009. Furthermore, an allowance of 2 more months was supplemented to this period in order to ensure that all charge claims had been settled by the UPMC Health Plan. Therefore, the effective time range for cost collection in this study ranged from October 15th, 2007 through June 15th, 2009 or 20 full months. At the end of this data collection period, the PI submitted the entire study sample in spreadsheet format to the honest broker. The subjects coded "dummy" account variables were then matched to their appropriate UPMC Health Plan ID number by the honest broker in order for her to obtain relevant cost data. All charges associated with any of the 27 diagnostic codes listed under Table 3 were extracted for each Health Plan Member ID in the sample. This was done primarily because a single episode of LBP may have been labeled with multiple ICD-9 codes by different providers. This is a relatively common practice in health care in order to obtain authorization for payment of specific services. In addition, this step accounted for the potential of follow-up episodes of care being assigned a different ICD-9 then was initially provided upon entry into the LBI. The ICD-9 code list in Table 3 was previously agreed upon as the most comprehensive by the Health Plan administrators, as it contains the most prevalently used diagnoses in the treatment of LBP. Data for pharmacology services were provided by the Health Plan in a separate spreadsheet and were not associated with any specific diagnostic codes as were the other charges that were captured in this study. The decision was made to error on the side of inclusion when capturing a particular class of drugs for this analysis. Aside from focusing only on specific drugs that might be indicated in common practice for LBP, the net was expanded to include any class of drugs that may have been prescribed with little if any supportive evidence. As such, any prescription drug charge associated with any of the following major drug classes were included in this analysis:

1: Analgesics / Narcotic Combinations (e.g. Hydracodone / Ibuprofin)

2. Anti-Anxiety Drugs

3. Glutocorticoids (e.g. Prednisone, Dexamethasone)

4. Narcotic Analgesics and Non-Salicylate Analgesic Combinations (e.g. Tylenol with Codeine)

5. NSAID's

6. Skeletal Muscle Relaxants

7. Anti-Convulsants (e.g. Neurontin)

8. Anticholinergics / Antispasmodics (e.g. Dicyclomine HCL)

9. Anti-inflamatory / Antiarthritic Agents (e.g. Hyaluranate Sodium)

10. Selective Seretonin Receptor Inhibitors (e.g. Cymbalta)

11. Alpha-2 Receptor Antagonist Anti-depressants

Finally, any charges made prior to October 15th, 2007 or after April 15th, 2009 were not reflected in the analysis.

Once charge data was acquired, the honest broker reconverted the sensitive Health Plan ID numbers back to the original "dummy" account variables and then resubmitted the report back to the PI. Data from the report was then gleaned and interpreted by the PI to determine the total direct health care charges and physical therapy charges by year for each subject. In addition, the report also reflected claims repricing to the Medicare fee schedule so that the results of this study would be generalizable across a public payer system, and to further account for contracting bias in the analysis. Lastly a standard 4% discount rate per year was applied to all charges to account for inflationary increases.

3.4.3 Additional cost extraction and computation

Since the payer also depends on a steady flow or demand from their enrollees in order to provide revenue, health plan administrators are often concerned with reducing obstacles that prevent individuals from purchasing or renewing their plans. A member's fiscal responsibility or cost burden for services would be one thing that potential enrollee's may consider when weighing their healthcare options. Certainly, higher member burdens would be an obstacle that may deter an enrollee from attending repeat rehabilitation visits or other highly deductible services within the plan. In one sense, higher burdens may prevent individuals from attaining the comprehensive care they need in order to improve upon their health. By contrast, lower member burdens may contribute to patterns of abuse by both providers and patients through over prescription of unneeded and possibly more costly services. In this respect, an understanding of member cost burden helps health plan administrators to establish a payment structure that is balance neutral and is fair to both the provider and the member while at the same time maximizing their revenue. In this study, individual member cost burden was defined as the net difference between the amount allowed to be paid for a particular service by the Health Plan (regardless of the providers actual charge (amount billed)) and the amount the Health Plan actually paid (reimbursed the provider) for those services. This difference represents the actual "out of pocket" expenses for each member and is reflected in the charge structure for their regular co-payments. In this study, the actual provider charges, the amounts allowed by the Health Plan and subsequently the amount paid for those services was extracted. As indicated previously, the Medicare Fee Schedule (MFS) was applied to those variables in order to provide a repricing schema that eliminated contractor bias and to improve generalizability. Individual physical therapy member cost burden was computed in the exact same manner as total member cost burden but was only inclusive of rehabilitation related services. A standard 4% discount rate per year was also applied to all charges to account for inflationary increases. Aside from simply comparing total costs and total physical therapy costs, this study also compared total member cost burden and total member physical therapy cost burden between on and off protocol groups. In addition, the top 25% of total costs, total physical therapy costs, total member burden and total physical therapy burden were computed and used as outcome variables as described under the Specific Aim 1 analysis below in order to determine if there was a strong relationship with protocol status as a predictor variable.

3.5 STATISTICAL DESIGN

3.5.1 Baseline characteristics

Data for the following baseline characteristics were collected on all subjects from the UPMC clinical database: age, gender, initial Oswestry (OSW) score, Initial Numeric Pain Score, Fear-Avoidance Behavioural Questionnaire (FABQ) scores (both work and physical activity subscale raw scores), the Chalrson Comorbidity Weighted Index Composite Score (CCI), the subjects insurance type-otherwise known as line of business (LOB) which reflected whether they had commercial insurance (CM) or received state sponsored medical assistance (MA), if they sought previous medical care (SMC) for their LBP, and finally the types clinical interventions to define if subject was initially treated according to an active physical therapy approach. In addition, information was provided on a subject's enrolment status and the exact time of their enrolment period with the Health Plan.

Initial symptom severity was assessed using the OSW and the Initial Pain Scores. The OSW is a self administered questionnaire that includes 10 items, each scored 0 to 5, and are related to a patient's LBP and their tolerance of daily activities. The total score is then summed and expressed as a percentage of disability. The OSW has high test-retest reliability and is responsive to changes in patients with LBP.(Fritz JM & Irrgang JJ, 2001) The minimum clinical important difference (MCID) for significant disability improvement on the OSW is a 6 point reduction from the baseline measure. The numeric pain rating scale requires that the subject rate their pain from 0 (none at all) to 10 (worst imaginable). Pain scales have been demonstrated to have both concurrent and predictive ability as measures of pain intensity and are responsive to change among patients with LBP.(Jensen MP, Turner JA, & Romano JM, 1994)

As stated previously, a subject's associated cormorbid illness was assessed using the Charlson Cormidity Index (CCI) Comorbid illness plays an essential and poorly defined role in the management of most health conditions including LBP. In this respect the presence of comobidities may confound potential clinical trial results and limit research findings to older and sicker patients. The CCI developed in 1987, comprises 19 medical conditions weighted 1-6 with total scores ranging from 0-37. During development of the index, mortality for each the 19 conditions was converted to a relative risk RR for death within a 12 month period. A weight was then assigned to each condition based on the associated RR for death. A weight of 1 is assigned to the following comorbidities: myocardial infarct, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, connective tissue disease, ulcer disease, mild liver disease, and diabetes. A weight of 2 was assigned to the following: hemeplegia, moderate to severe renal disease, diabetes with end organ disease, any tumor , leukaemia, and lymphoma. Moderate or severe liver disease is assigned a weight of 3. Lastly, a weight of 6 is assigned to the following conditions: metastatic or solid tumor, and AIDS. The CCI has been demonstrated to have good test-retest reliability, and moderate to good inter-rater reliability. In addition, it has been demonstrated to have good predictive validity for mortality, disability, readmission and hospital length of stay outcomes. (de Groot V., Beckerman H., Lankhorst G., & Bouter C, 2003)

Next, active physical therapy intervention strategies are supported over passive strategies by current practice guidelines and furthermore have been demonstrated to be associated with improved disability outcomes, decreased health service costs, and lower medical utilization rates.(Fritz JM, Cleland JA, & Brennan GP, 2007; Fritz JM, Cleland JA, Speckman M, Brennan GP, & Hunter SJ, 2008) Simply by knowing the types of interventions given to a subject through MDS reporting (Appendix A), a determination was then be made whether an active or passive therapy approach was given on the initial physical therapy visit. For this study, a similar method was employed that was reported by Fritz et al. to determine a specific treatment approach type. In this manner, adherence to an active approach was defined as one that had at least 75% active interventions and was computed in the following manner: (# active interventions) / (# active interventions + # passive interventions) * 100%. Active interventions are those that would be more consistent with active guideline recommendations and prior CPT code classification in the literature. Table 4 below specifies whether the interventions collected in this study from the MDS (Appendix A) were considered as active versus passive.

Table 4. Active versus passive interventions

Active Interventions	**Passive Interventions**
Patient Education	Mobilization Grade I-IV
Flexion Exercises	Myofascial Release
Extension Exercises	Soft Tissue Massage
Flexibility Exercises	NMES (Pain Control)

Stabilization Exercises	Heat Modalities
General Conditioning Exercises	Cold Modalities
Aerobic Exercises	Traction - Mechanical
Functional Training	Traction - Auto
Mobilization Grade V	De-weighting
Behavioural Exercises	Cranio-sacral therapy
	NMES (Strengthening)

It would be reasonable to assume that the type of an individual's insurance could also influence the way they seek or the manner in which health care is delivered to them. Plans with minimal to no co-payments or very low cost burdens do little to de-incentive individuals from pursuing a wider range and greater volume of medical services. In addition, providers may seek to increase or decrease their volume of services or procedures to individuals based upon how well they are reimbursed from the insurance carrier. It is for these reasons that different cost strata for health care may exist among carrier types and why this variable was adjusted for during this study. In this respect, an individual's line of business (LOB) was classified categorically as either CM (commercial insurance) or MA (medical assistance). None of the subjects in this study were Medicare patients as would be expected since anyone over 65 was excluded from the analysis. Likewise subjects who have sought medical care in the past for LBP are beginning to demonstrate signs of symptom chronicity thereby increasing their risk of accruing higher health care costs. Therefore this variable (SMC-sought medical care) was also collected in order to compare groups at baseline and to be adjusted for in subsequent analysis.

Lastly, it would also be reasonable to assume that subjects with a longer data collection period had a greater opportunity to accrue more costs. Therefore, a variable called study-months was created which was calculated as the time from the initial physical therapy visit until the end of the study or the subject's enrolment period if this date preceded April 15th, 2009. This variable was also compared at baseline and adjusted for in subsequent analysis as described below.

As indicated above, the following variables were compared descriptively between subjects who were on versus off protocol according to the TBC: age, gender, initial OSW, initial numeric pain scores, initial raw FABQ_WK and FABQ_PA scores, initial CCI scores, LOB, SMC, the total time enrolled in the study (study-months), and lastly, their treatment status in regards to either receiving an active or passive therapy approach. Gender data was categorically classified as either male or female. The following three variables were dichotomously classified as well: LOB (CM / MA), SMC (Y / N) and active treatment (Y / N). The other seven aforementioned characteristics were accessed as continuous variables. The data for each continuous variable was plotted as a histogram with accompanying Q-Q plots and Shaprio-Wilk tests performed to access normality. In addition skewness was accessed by skewness-kurtosis tests. All continuous variables were non-normal and non-skewed. As a result, data transformation was not necessary; however the violation of normality assumptions required the use of non-parametric Wilcoxon Rank Sum test for baseline comparison. An alpha level of $p<0.05$ was used for all comparisons. Next, chi-Square ($\aleph^2$) tests ($p<0.05$) were used to compare all categorical variables between the on and off protocol combined groups and classification subgroups. Descriptive statistics were also computed, including frequency counts for all categorical variables and measures of central tendency (mean, median) and dispersion (variance, standard deviation) for continuous variables to summarize the data and describe the characteristics of patients within each classification group.

Since each sub-group was analyzed independently, and since successive tests were not performed, Bonferonni or other correction procedures to minimize Type I error were not used during these comparisons. However, because there were greater then 3-4 significant differences (11 variables * 6 groups * alpha (0.05) = 3.3 expected significant findings) at baseline, it is important to point out that the lack of control for Type I error may have resulted in some spurious findings. In order to make the analysis more robust, hierarchical or stepwise linear regression was applied to adjust for every aforementioned baseline characteristic for each outcome variable.

Lastly, as previously indicated, patients were classified as TBC on / off protocol based upon their presentation on the initial (first exam) physical therapy visit. All excluded subjects were documented as well as the specific reason for their exclusion. Therapist compliance with MDS criteria throughout the duration of this observational analysis was considered excellent (>95%). This means, of the 750 classified cases thus far, only 36 (4.8%) MDS's were considered incomplete (missing at least one variable). These cases were statistically compared to the completed cases {714 (95.2%)} to ensure there were no baseline differences in terms of the 11 variables that were analyzed as discussed above. Wilcoxon Rank Sum tests were used to compare the seven continuous variables and Chi-square (א^{2}) tests were used to compare the four categorical collected at baseline. An alpha level of $p<0.05$ was used for these comparisons. Since no statistical differences were observed during these comparisons, SAS assigned a missing code to these incomplete data points for subsequent regression analysis. It should be noted that all stepwise and logistical regressions in SAS procedurally eliminated the subjects with any missing variables for the analysis, but at least a determination was made from baseline comparisons that this group was not any different then the group with completed data sets, and that these subjects did not account for any variation in explaining the results in this study.

3.5.2 Cost minimization analysis

Total net direct health care charges were computed for the sample and broken down into total net direct health care charges for on protocol subjects and total net direct health care charges for off protocol subjects. In addition, total physical therapy costs were computed in the same fashion. This provided both the total direct health care charges and total physical therapy charges for each combined sample in order to determine if there was any cost savings associated with using the TBC approach over the usual care approach for treatment of LBP. Descriptive statistics were also computed such as measures of central tendency (mean, median) and dispersion (variance, standard deviation) in order to summarize the continuous charge data. Next, Shapiro-Wilk tests determined that the data was non-normal so non-parametric Wilcoxon rank-sum tests were used to compare the medians for costs between the on and off protocol groups. An alpha level of $p<0.05$ was used to determine statistical significance for all comparisons. As indicated before, a standard discount rate of 4% per year was applied to all charges prior to analysis to offset inflationary increases. After the sample was examined as a whole, the separate TBC cohorts (e.g. manipulation, specific exercise, and stabilization) were parceled out from the combined sample and analyzed singly to determine if individual cohort cost savings existed. Once again, the total direct health care costs and total physical therapy costs associated with being on protocol versus off protocol were compared for each category similar to the way the combined sample was analyzed. Alpha levels of $p<0.05$ were used to determine statistical significance.

Lastly, total member cost burden was also computed by first obtaining the difference between the Health Plan allowable charge and the amount they actually reimbursed or paid for each line item of charge data. These amounts were then summed by individual Health Plan ID "dummy" account numbers in order to determine the total cost burden for each subject. The member cost-burden for all physical therapy related charges were also extracted by procedural code from this same data set and then summed for each subject. Both total member cost burden and total member physical therapy cost burden was compared between the on and off-protocol combined groups using non-parametric Wilcoxon Rank Sum tests ($\alpha = 0.05$) in order to determine if there was a difference in member "out of pocket" expenses. MFS repricing and the 4% discounting rate was also applied for this procedure. Similar to the previous total cost variable comparisons, the TBC cohorts (e.g. manipulation, specific exercise, and stabilization) were parceled out from the combined sample and analyzed singly using the same statistical procedures to determine if individual member "out of pocket" were different within the major classifications of the TBC.

Because of the lack of control for Type I error with multiple baseline comparisons, the decision was made to perform exploratory hierarchical linear regression analysis to determine what effect if any the independent cohort variable (protocol status) had in accounting for the variation in the outcome variables (cost and member burden) after adjustment. Therefore the cohort (protocol) status and each of the 11 baseline characteristic variables previously discussed were entered univariately into a linear model corresponding with each of the 4 identified outcomes (total direct net health care costs, total direct net physical therapy costs, total member burden and total physical therapy member burden). Variables were included in the second step if their Pr. > F $\leq$ 0.15. The cutoff threshold to be included in subsequent steps was $p < 0.05$. Variables were entered into the final model by smallest p-value at each step. Once all variables were either eliminated or included in the model, all possible interaction effects between significant variables were explored in a similar stepwise procedure. The cohort or protocol status variable was then added back to the model to determine multi-viarately if this variable accounted for any additional variation beyond adjustment. The threshold for inclusion was again $p < 0.05$. Possible interaction effects with protocol status were also explored. Each outcome variable was accounted for in this manner for the combined sample, and for each TBC subgroup classification. A total of 24 linear regression models were developed for final analysis (6 group or subgroup comparisons * 4 outcome variables). All cost data was transformed by [log (10)] in order to fit the normal distribution assumptions for linear regression analysis with continuous outcome variables.

By extending the primary hypothesis; instead of just being simply more costly, it was also reasonable to believe that being off-protocol according to the TBC would be a strong predictor of developing unusually high total direct health care costs for LBP treatment within the health plan. For this study, subjects with unusually high costs were defined as those in the 75^{th} percentile of the distribution of total direct health care charges. The relationship between being on or off-protocol and incurring high costs (dichotomous Y / N outcome) was examined using a hierarchical logistic regression model. As with the hierarchical liner regression analysis, protocol status and each of the 11 baseline characteristic variables were first entered univariately into a logistic model corresponding with each of the 4 identified outcomes (top 25% (Q_3) of total direct net health care costs, Q_3 total direct net physical therapy costs, Q_3 total member burden and Q_3 total physical therapy member burden). Variables were included in the second step if their Pr. > Chi-square ≤ 0.15. The cutoff threshold to be included in subsequent steps was $p < 0.05$. Once all variables were either eliminated or included in the model, all possible interaction effects between significant variables were then explored in a similar stepwise procedure. The cohort or protocol status variable was then added back to the model to determine multi-variately if this variable accounted for any additional variation beyond adjustment that would explain predictive ability. The threshold for inclusion was $p < 0.05$. Possible interaction effects with protocol status were also explored. Since the outcome variables in this case are dichotomous, the assumptions for binomial distributions were followed. In this respect there was no need for data transformation because normality is not an assumption. Adjusted odds ratio's (OR) with 95% confidence intervals were also computed. The relationship between being on or off-protocol and incurring high costs or member burdens were be analyzed in the same manner for each of the three TBC subgroups. As a result, a total of 24 logistic regression models were ultimately produced in this analysis (6 group or subgroup comparisons* 4 outcome variables).

3.5.3 Decision analysis model to determine cost effectiveness

A major assumption that is often made with cost minimization analysis is that the outcomes among the comparators are equal or that their effectiveness is the same. As indicated in the background review, there have been no real effectiveness studies demonstrating the collective superiority of the TBC over usual physical therapy care in treating LBP. What has been shown is that individually, each of the different algorithms has demonstrated effectiveness over a standard alternative treatment strategy for a defined subset of subjects. In some cases the difference was very large. It was reasonable to assume then that this would translate into a collective effectiveness of the TBC approach that is far superior to a usual care approach. At this time, it is not feasible to assess clinical outcomes from the UPMC database because follow-up outcome measures (such as OSW) were not being consistently recorded by the clinicians. CRS policy is changing to improve this limitation and a separate quality initiative will likely be needed to see that this effort succeeds over time. In any case, this will be a focus of any cost effectiveness analysis beyond this study, but uncertainty remains over the collective TBC effectiveness at this time.

Because uncertainty existed primarily over the level of collective treatment effectiveness in this investigation, a simple (non-Markov) decision analysis model (Figure 2) was composed to compare the two treatment strategies (On-protocol versus Off-protocol) once the cost data was analyzed for specific aim 1. TreeAge Pro® 2009 software was used to build and analyze the model.

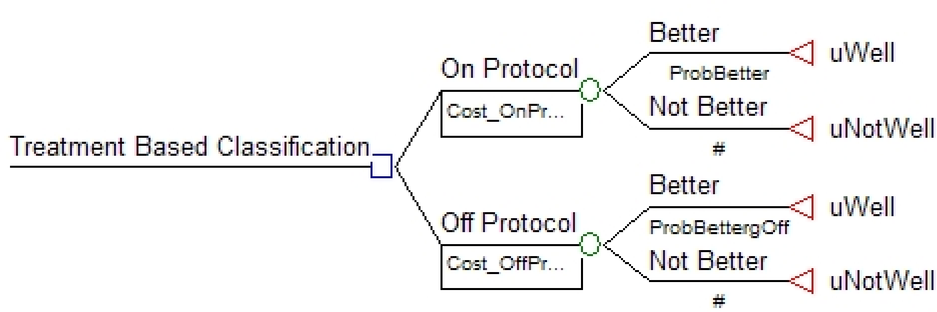

Figure 2. Decision analysis model

The model (Figure 2) that was developed above essentially describes two treatment strategies or choices at the main decision node (square): 1) On-protocol according to the TBC or 2) Off-protocol according to the TBC. As the model progresses upstream (to the right), there are basically two outcomes that could have occurred at each chance node (circles). This means that subjects either got better or not based upon if they were on or off protocol according to the TBC. Since these outcomes arose as a result of either main choice, they were associated with a conditional probability of occurring which will be discussed more below. The triangles demarcate the terminal node or end state for the subjects. In this study, the primary investigator was simply concerned whether the patients got better or not so there was no need to extend the model further. To the right of the terminal nodes are the utilities or values which represented those end states. Since, the outcome information obtained from the UMPC clinical database was deemed unreliable due to a lack of reporting and limited time horizon follow-up; these values were obtained from the literature. The rational for these utility variable assignments is also explained below.

The mean cost data was applied to the model directly from the specific aim 1 combined sample analysis in order to assign real numerical quantities for the two cost variables; Cost_OnProtocol and Cost_OffProtocol. Since the primary focus for this study was to obtain the total costs per cohort (on versus off protocol), and since the associated outcome data from this sample was considered unreliable, separate cost variables were not assigned to groups of subjects that got better or not given their TBC classification status. Instead, the default costs were set for each treatment arm based upon the subject's classification as being on versus off protocol. Distributions were also assigned in TreeAge®, as well as measures of central tendency and variability for the two cost variables based upon the results from the specific aim 1 analysis. This allowed the model to perform further sensitivity analysis across a range of known values.

Conditional probabilities are always associated with chance node outcomes in decision analysis models. As indicated above, it was not possible to infer directly from this study what proportion of patients got better or not based upon their classification status. In addition, there was no collective figure available in the research to demonstrate how much more effective the TBC is versus usual care. However, data was available from studies examining the effectiveness of each main TBC subgroup against a comparative standard treatment. Results from manipulation and stabilization studies have demonstrated a range of improvement in the post-treatment probability of success from 80% to 97% given a pre-treatment probability of success around 45% to 50%.(Childs JD et al., 2004; Flynn T et al., 2002; Fritz JM, Erhard RE et al., 1998; Hicks GE et al., 2005) Lastly, the specific exercise category has demonstrated significant improvement in the literature compared to a standard treatment.(Browder DA et al., 2007; Clare HA et al., 2004; George SZ et al., 2003; Petersen T et al., 2002) The literature also indicates that typically about a third (0.33) of LBP cases treated under usual care completely resolve within a year.(Cassidy JD et al., 2005; Croft PR et al., 1998) Based upon this information, an assumption was made regarding the conditional probabilities that a subject got better or not given they were on or off protocol. Since it was understood that the separate subgroups of the TBC were individually effective against common usual care strategies, it was reasonable to assume for this model a higher conditional probability of improvement for the on protocol group then 0.33 which was assigned to the probability of getting better for the usual care group. In this case, a conservative figure of 50% was chosen even though this figure was believed to be much higher. By demonstrating a significant cost-effectiveness given this small (possibly worse case scenario) figure, it was felt that this would lend considerably more credibility to the TBC algorithms as an associated cost-savings from specific aim 1 was hypothesized. The probability variables are designated below the "Better" labels in both protocol arms. They were defined as the probability for getting better if on-protocol (ProbBetter) and the probability of getting better if off protocol (ProbBettergOff). The # symbol therefore denotes the compliment probability of (1 - ProbBetter) for the "Better" condition in the on-protocol arm and the probability of (1-ProbBettergOff) for the "Not Better" condition in the off-protocol arm. Means and standard deviations were also computed from the final Oswestry score differences in the current data set and compared using non-parametric Rank Sum Wilcoxon procedures due to rejection of distribution normality. This was done simply to estimate if the outcome probability figures in this sample reasonably approximated the results which were indirectly obtained from the literature as described.

Utility scores are usually derived from validated scales that access a person's overall quality of life. These scores normally range from 0 (death) to 1.00 (perfect health). Aside from global quality of life scales, there are also specific scales for certain conditions such as back pain and associated disability. Throughout the literature, the range associated with good back pain outcomes is around (0.83 to 0.95).(Malter AD, Larson EB, Urban N, & Deyo RA, 1996; Taylor RJ & Taylor RS, 2005) Severe chronic back pain has been associated with baseline utility values around 0.35.(Soegaard R et al., 2007) Other moderate LBP utilities range from 0.59 to 0.79.(Fryback DG et al., 1993; Gold MR, Franks P, McCoy KI, & Fryback DG, 1998; Malter

AD et al., 1996; Soegaard R et al., 2007; Whitehurst DG et al., 2007) For this study, a utility of 0.89 was chosen for the “Better” end state, and 0.69 for the “Not Better” end state.

Since the distributions for the conditional probabilities of improvement or the terminal end state utilities were unable to be computed from the literature, the recommended guidelines for probabilistic analysis were followed in this model.(Briggs AH, Goeree R, Blackhouse G, & O'Brien BJ, 2002) Beta distributions were selected for all probabilities and uniform distributions for all utilities. Since the cost distributions were known, these were properly designated versus arbitrarily selecting a gamma or lognormal distribution as recommended.

Once all pertinent variable information was applied to the model, the software program performed the cost-effectiveness analysis. It accomplished this through a method of “averaging out and folding back” the decision tree.

After the cost-effectiveness analysis was accomplished, one-way sensitivity analysis was performed varying the probability of getting better given either strategy, the total direct health care costs, and the utility values associated with the end states in order to produce a tornado diagram demonstrating which individual variables had the greatest effect on the ICER’s. Two-way sensitivity analysis (varying two of these variables at a time) was then performed to better access the stability of the model under change. In addition, a probabilistic sensitivity or Monte Carlo analysis (similar to a bootstrapping method) was performed varying all parameters simultaneously across their distributions to test the robustness of the results over several thousand iterations.

4.0 RESULTS

4.1 STUDY SAMPLE WITH INCLUSION / EXCLUSION BREAKDOWN

A total of 1, 237 subjects were enrolled in the LBI from Oct 15th, 2007 to Oct 15th, 2008. Out of those, 295 (23.85%) subjects were excluded leaving a sample of 942 subjects to be analyzed. The breakdown for subject exclusions can be viewed in Table 5 below.

Table 5. Reasons for subject exclusion

Reason for Exclusion	**Frequency**
Past Surgical	169
Age < 18 years	61
Age > 65 years	43
Wrong Diagnosis (Not lumbar spine related)	12
Pregnancy	5
Subject Refusal	4
Medical Red Flag Present	1
Total	**295**

4.2 TREATMENT BASED CLASSIFICATION (TBC) COHORT BREAKDOWN

Out of the remaining 942 subjects eligible for this study, a total of 750 (79.62%) were classified into at least 1 of the 3 main TBC cohorts. 103 (13.73 %) subjects fit the criteria as manipulation candidates and 149 (19.87%) were classified as specific exercise candidates [74 (9.87%) flexion specific & 75 (10.0%) extension specific]. In addition, there were 498 (66.4%) total stabilization candidates. The remaining 192 (20.38%) subjects did not fit the definitions in this study to be classified into either TBC category.

A total of 14 out of 103 (13.59%) manipulation candidates met the criteria for being on-protocol under the criteria listed in Appendix B. The remaining 89 (86.41%) were considered to be off-protocol based upon those same definitions. In contrast the specific exercise subgroup on-protocol proportion comprised a much larger percentage (81.21%, n=121) then those considered off-protocol (18.79%, n=28). The breakdown of flexion exercise specific candidates and extension exercise specific was roughly equivalent [Flexion exercise specific (on-protocol; 81.08%, n=60) (off-protocol; 18.92%, n=14), Extension exercise specific (on-protocol; 81.33%, n=61) (off-protocol; 18.67%, n=14). The stabilization cohort was more equally distributed with regard to on and off-protocol with 253 (50.80%) off-protocol candidates and 245 (49.20%) on-protocol candidates. The stabilization cohort included 363 (72.89%) subjects who had signs and symptoms consistent with non-success with stabilization exercises (negative prediction rule candidates) and 143 (28.71%) subjects who were positive to the CPR for a positive response to stabilization exercises (prediction rule candidates). Out of those, 8 (1.6%) subjects were both negative and positive prediction rule candidates. Because the negative prediction rule took precedence in this study, those individuals were classified according to that criterion only. Out of the 363 negative prediction rule candidates, 229 (63.09%) were treated off-protocol or given stabilization exercises, and 134 (36.91%) were treated on-protocol or not given stabilization exercises. The remaining 135 prediction rule candidates were treated as follows: 111 (82.22%) were treated on-protocol or given stabilization exercises and 24 (17.78%) were treated off-protocol (no stabilization exercises). Overall the combined sample had 380 (50.67%) on-protocol candidates and 370 (49.33%) off-protocol candidates.

4.3 COMBINED SAMPLE AND SUBGROUP BASELINE CHARACTERISTICS

Characteristics of the 750 subjects in the study are broken down by cohort and are presented in Table 6 below.

Table 6. Initial subject baseline characteristics for the combined sample and broken down by TBC subgroup

	Combined Cohort (N=750)	**Manipulation (N=103)**	**Stabilization (N=498)**	**Specific Exercise (N=149)**	**(Flex / Ext) (N= 74 / 75)**
Age					
Mean	46.1	45.6	45.6	48.0	50.7 / 45.4
SD	11.9	12.1	11.9	11.6	10.4 / 12.2
Range	45.8	45.7	45.8	44.6	39.4 / 44.6
Missing	-	-	-	-	- / -
Gender					
Male	292	40	191	61	24 / 37
Female	456	62	306	88	50 / 38
Missing	(2)	(1)	(1)	-	- / -
CCI Score					

Mean *SD* *Range* *Missing*	0.16 0.44 2.0 (23)	0.15 0.46 2.0 (3)	0.17 0.45 2.0 (15)	0.16 0.40 2.0 (5)	0.22 / 0.10 0.45 / 0.35 2.0 / 2.0 - / (5)
FABQ_WK *Mean* *SD* *Range* *Missing*	 10.8 11.1 42.0 (26)	 9.7 10.2 41.0 (2)	 10.5 11.2 42.0 (22)	 12.6 11.1 42.0 (2)	 13.4 / 11.9 11.8 / 10.5 39.0 / 42.0 (2) / -
FABQ_PA *Mean* *SD* *Range* *Missing*	 13.4 7.0 24.0 (23)	 15.8 6.2 24.0 (2)	 12.5 7.1 24.0 (19)	 14.6 6.3 24.0 (2)	 15.1 / 14.1 6.2 / 6.4 24.0 / 24.0 (2) / -
Oswestry *Mean* *SD* *Range* *Missing*	 38.2 16.2 86.0 (3)	 46.6 11.8 44.0 -	 35.7 16.1 82.0 (3)	 40.6 16.5 86.0 -	 39.7 / 41.5 15.3 / 17.7 64.0 / 86.0
Pain *Mean* *SD* *Range* *Missing*	 7.0 2.4 10.0 (3)	 7.8 2.2 10.0 -	 6.8 2.5 10.0 (3)	 7.3 2.3 10.0 -	 7.3 / 7.2 2.2 / 2.4 9.0 / 10.0 - / -
Active Treatment *Yes* *No* *Missing*	 288 462 -	 34 69 -	 205 293 -	 49 100 -	 31 / 18 43 / 57 - / -
LOB *CM* *MA* *Missing*	 584 144 (22)	 81 19 (3)	 385 99 (14)	 118 26 (5)	 55 / 63 19 / 7 - / (5)

SMC					
Yes	290	34	187	69	41 / 28
No	460	69	311	80	33 / 47
Missing	-	-	-	-	- / -
Study-Months					
Mean	11.2	10.8	11.4	10.9	11.4 / 10.5
SD	4.0	4.3	4.0	3.9	4.0 / 3.8
Range	18.0	18.0	18.0	16.5	15.7 / 15.8
Missing	(22)	(3)	(14)	(5)	- / (5)

Abbreviation: CCI = Charlson Comorbidity Index
LOB = Insurance (Line of Business)
SMC = Sought medical care in past for same condition

4.4 TESTING NORMALITY FOR BASELINE CHARACTERISTIC DISTRIBUTIONS

Shapiro-Wilk tests were used to access the normality of the seven continuous variable baseline characteristic distributions. All continuous baseline variables analyzed in this study had significant p-values < 0.05, therefore the null hypothesis for normal distributions were rejected for each of those characteristics. Therefore non-parametric Wilcoxon Ran-Sum tests were required for any comparisons of those variables. The remaining four categorical variables were compared with Chi-Square ($\aleph^2$) tests. The results of the Shapiro-Wilk tests for normality are summarized in Table 7 below:

Table 7. Normal distribution testing for continuous baseline variables

Baseline Variable	**Shapiro-Wilk Test Statistic**	**p-value**
Age*	0.9604	< 0.0001
FABQ_WK*	0.88963	< 0.0001
FABQ_PA*	0.96405	< 0.0001
Initial Oswestry*	0.98909	< 0.0001
Initial Pain Score*	0.92293	< 0.0001
Charlson Comorbidity Index *	0.408169	< 0.0001
Study-Months*	0.969464	<0.0001
* = *Significant finding (Reject the null of normal distribution)*		

4.5 COMPARISON OF BASELINE CHARACTERISTICS BETWEEN SUBGROUPS WITH COMPLETED OR INCOMPLETE DATA SETS

The original sample (N=942) had 56 (5.94%) subjects with incomplete MDS's. A data set was defined as incomplete if at least one variable listed in Appendix A was missing. Thirty-six (4.8%) subjects had incomplete data sets as opposed to 714 (95.2%) with completed MDS's. Baseline characteristics were compared between subjects with completed versus incomplete data sets using Wilcoxon Rank Sums (continuous variables) and Chi-Square analyses (categorical variables). No statistically significant differences were found between the two groups across those eleven baseline variables. The results of the comparisons between these groups can be viewed in Table 8 below.

Table 8. Comparison of baseline characteristics between subgroups with completed versus incomplete MDS's

Baseline Variable	**Sig.**
Age†	0.2100
FABQ_WK†	0.1684
FABQ_PA†	0.4675
Initial Oswestry†	0.3210
Initial Pain Score†	0.5052
Study-Months†	0.8458
Charlson Comorbidity Index Score†	0.8739
Gender*	0.4721
Active Treatment*	0.4447
SMC*	0.7469
LOB*	0.0820
† = Compared using non-parametric Wilcoxon Rank Sum Test ** = Compared using Chi-Square*	

4.6 COMPARISON OF BASELINE CHARACTERISTICS BY COHORT (ON VERSUS OFF PROTOCOL)

Baseline characteristic comparisons between on versus off protocol for the combined sample and TBC cohorts are presented below in Table 9. There were a total of 9 significant findings out of 66 comparisons (13.6%) for the entire analysis. Statistically significant findings for the overall cohort included the Fear Avoidance Beliefs Questionnaire Physical Activity scale (FABQ_PA) and the Line of Business variable, the later of which separated out commercial from Medical Assistance/Medicaid categories. In addition, sub-group (based on TBC) comparisons resulted in 7 additional variables that were significantly different between the on and off-protocol cohorts. All significant findings are highlighted in bold.

Table 9. Comparison of baseline characteristics by cohort (on versus off protocol)

	Combined Cohort (N=750)	Manipulation (N=103)	Stabilization (N=498)	Specific Exercise (N=149)	(Flex / Ext) (N= 74 / 75)
Age					
Sig.	*0.8204*	*0.2366*	*0.4255*	*0.1271*	***0.0228†*** */ 0.9134*
Gender*					
Sig.	*0.5058*	*0.2474*	*0.7440*	*0.1397*	*0.7318‡ / 0.0849*
CCI Score					
Sig.	*0.7246*	*0.1588*	*0.5321*	*0.5041*	*0.1753 / 0.4139*
FABQ_WK					
Sig.	*0.1938*	*0.4341*	*0.4877*	*0.4745*	*0.1534 / 0.6224*
FABQ_PA					
Sig.	***0.0486†***	*0.7151*	***0.0020†***	*0.7914*	*0.2121 / 0.2875*
Oswestry					
Sig.	*0.6032*	*0.5564*	*0.1143*	*0.0670*	*0.3798 / 0.0864*
Study-Months					
Sig.	*0.7350*	*0.9762*	*0.5891*	***0.0319 †***	*0.0773 / 0.1989*
Pain					
Sig.	*0.8776*	*0.8177*	*0.0739*	*0.0880*	*0.8722 /* ***0.0282†***
Active Treatment*					
Sig.	*0.7731*	*0.5417‡*	*0.9787*	*0.2127*	*0.9352 / 0.0867‡*

LOB* *Sig.*	***0.0373†***	*0.2227‡*	*0.3565*	***0.0068†***	***0.0028†‡*** */ 0.5501‡*
SMC* *Sig.*	*0.1011*	*0.3215‡*	***0.0096†***	*0.6638*	*0.4579 / 0.8895*

Abbreviation: CCI = Charlson Comorbidity Index, LOB = Line of Business (CM=Commercial Insurance Carrier, MA=Medical Assistance), SMC = Sought Medical Care in past for condiditon

** = Variable compared using Chi-Square test unless otherwise indicated*

‡ = Fisher's exact test used for cells with expected counts < 5

† = Significant finding with p-value < 0.05

4.7 SPECIFIC AIM 1 ANALYSIS

Specific Aim 1: *Perform a cost minimization analysis comparing both direct health care and physical therapy costs for subjects classified as on-protocol according to the three major subsets (manipulation, specific exercise, and stabilization exercise) of a TBC approach to those who are considered off-protocol (receiving a usual physical therapy care approach) in the management of patients with LBP in the primary outpatient physical therapy setting. In addition, member total cost burden and physical therapy cost burden will be compared by protocol status to determine if treatment status influences a subject's actual out-of-pocket direct health care expenses.*

4.7.1 Primary outcome descriptives

Overall Costs: In the combined study sample, downstream costs in the on-protocol group were substantively less than the off-protocol group in in all 4 primary outcome measures: total direct net healthcare costs, total member burden, total direct physical therapy costs, & total physical therapy member burden, with total direct healthcare costs for the combined off-protocol group ($941,897.55) 1.4 times higher than the amount spent by the Health Plan for the on-protocol subjects ($658,477.94). This breaks down to $157.82 spent per member month for subjects treated according to the TBC versus $235.69 per member month for those given "usual" physical therapy care. (e.g., care consistent with reimbursement and scope of practice standards but not consistent with best evidence standards).

Physical Therapy Costs: The total amount spent by the Health Plan for physical therapy care was $182,746.85 (27.75% of total healthcare costs) for the combined on-protocol group versus $211,054.57 (22.40% of total healthcare costs) for off-protocol. This equates to $43.80 per member month spent for physical therapy care in the on-protocol group and $52.81 spent per member month for physical therapy care for the off-protocol group. Therefore the payer profited through substantial cost savings in both total dollars spent on LBP care and physical therapy care for subjects treated per TBC protocol definitions defined in this study.

Member Burden: Patient's out of pocket expenses were less if they were treated according to the TBC classification in this study. On-protocol member out of pocket expenses for the combined sample was $90,779.56 over the 18-month study period compared to $118,987.48 off-protocol. This breaks down to a cost savings of just over $8.00 per member month ($21.76 on-protocol vs. $29.77 off-protocol). Additionally, member out of pocket expenses for physical therapy care was cheaper for on-protocol subjects ($43,377.70 or $10.40 per member month) compared to off-protocol subjects ($47,046.95 or $11.77 per member month). Descriptive characteristics of the four primary outcome variables are broken down by TBC cohort in Table 10 below.

Table 10. Primary outcome descriptives by cohort and protocol status

	Total Costs	Member Burden	Total Physical Therapy Costs	Physical Therapy Member Burden
Combined On-Protocol				
Mean	*$1,732.84*	*$238.89*	*$502.05*	*$118.84*
SD	*$3,427.09*	*$325.29*	*$508.07*	*$145.78*
95% CI	*($1,387.16 - $2,078.51)*	*($206.08 - $271.70*	*($449.68 - $554.42)*	*($103.84 - $133.85)*
Combined Off-Protocol				
Mean	*$2,545.67*	*$321.59*	*$601.29*	*$134.04*
SD	*$6,142.66*	*$933.70*	*$786.11*	*$294.28*
95% CI	*($1,917.71 - $3,173.63)*	*($226.14 - $417.04)*	*($518.69 - $683.90)*	*($103.14 - $164.93)*
Manipulation On-Protocol				
Mean	*$1,016.91*	*$242.88*	*$463.96*	*$158.49*
SD	*$1,621.90*	*$228.53*	*$424.62*	*$171.78*
95% CI	*($80.45 - $1,953.37)*	*($110.93 – $374.83)*	*($218.79 - $709.12)*	*($59.30 - $257.67)*
Manipulation Off-Protocol				
Mean	*$1,990.99*	*$329.89*	*$618.93*	*$117.00*
SD	*$3,764.71*	*$1,169.54*	*$826.05*	*$165.83*
95% CI	*($1,197.94 - $2,784.03)*	*($83.52 - $576.26)*	*($436.28 - $801.59)*	*($80.33 - $153.67)*

Stabilization On-Protocol				
Mean	*$1,373.11*	*$220.91*	*$473.92*	*$104.40*
SD	*$2,447.51*	*$340.94*	*$521.51*	*$135.52*
95% CI	*($1,065.11 - $1,681.11)*	*($178.01 - $263.52)*	*($406.75 - $541.09)*	*($86.98 - $121.81)*
Stabilization Off-Protocol				
Mean	*$2,610.12*	*$324.33*	*$578.52*	*$140.94*
SD	*$6,589.89*	*$882.47*	*$759.98*	*$335.40*
95% CI	*($1,794.18 - $3,426.05)*	*($215.06 - $433.59)*	*($482.29 – $674.76)*	*($98.47 - $183.42)*
Specific Exercise On-Protocol				
Mean	*$2,544.04*	*$274.84*	*$563.40*	*$143.32*
SD	*$4,862.28*	*$300.37*	*$487.66*	*$159.18*
95% CI	*($1,668.86 - $3,419.22)*	*($220.77 - $328.90)*	*($473.71 – $653.09)*	*($114.05 - $172.60)*
Specific Exercise Off-Protocol				
Mean	*$3,726.44*	*$270.45*	*$747.10*	*$123.60*
SD	*$7,852.71*	*$430.28*	*$906.79*	*$188.71*
95% CI	*($681.48 - $6,771.40)*	*($103.61 - $437.30)*	*($395.48 - $1,098.72)*	*($50.43 - $196.78)*
Flexion On-Protocol				
Mean	*$2,931.69*	*$295.14*	*$610.25*	*$145.65*
SD	*$5,648.97*	*$316.50*	*$552.37*	*$156.86*
95% CI	*($1,472.41 - $4,390. 97)*	*($213.37 - $376.90)*	*($467.55 - $752.94)*	*($105.13 - $186.17)*

Flexion Off-Protocol				
Mean	*$4,667.67*	*$156.75*	*$648.61*	*$86.60*
SD	*$10,559.77*	*$196.65*	*$471.81*	*$177.83*
95% CI	*($1,429.36 - $10,764.70)*	*($43.21 - $270.29)*	*($376.20 - $921.02)*	*($16.07 - $189.27)*
Extension On-Protocol				
Mean	*$2,162.74*	*$254.87*	*$513.21*	*$140.83*
SD	*$3,950.50*	*$284.82*	*$406.17*	*$163.00*
95% CI	*($1,150.97 - $3,174.51)*	*($181.92 - $327.81)*	*($404.44 - $621.98)*	*($97.18 – 184.49)*
Extension Off-Protocol				
Mean	*$2,785.21*	*$384.16*	*$845.59*	*$160.61*
SD	*$3,828.39*	*$563.92*	*$1,210.09*	*$198.47*
95% CI	*($574.77 - $4.995.66)*	*($58.56 - $709.96)*	*($146.90 - $1,544.27)*	*($46.01 - $275.20)*

4.7.2 Testing normality for primary outcome distributions

Shapiro-Wilk tests were used to access the normality of the four continuous outcome baseline characteristic distributions. All four variables analyzed in this study had significant p-values < 0.05, therefore the null hypothesis for normal distributions were rejected for each of those characteristics. Therefore non-parametric Wilcoxon Ran-Sum tests was used to compare outcome variables statistically between the on and off protocol groups. In addition, since cost data is non-zero and typically skewed, the data was transformed by [log (10)] for comparisons and subsequent regressions. The results of the Shapiro-Wilk tests for normality are summarized in Table 11 below:

Table 11. Normal distribution testing for continuous primary outcome variables

Baseline Variable	**Shapiro-Wilk Test Statistic**	**p-value**
Total Direct Health Care Costs*	0.38917	< 0.0001
Total Member Burden*	0.308569	< 0.0001
Direct Physical Therapy Costs*	0.563277	< 0.0001
Physical Therapy Member Burden*	0.449868	< 0.0001
** = Significant finding (Reject the null of normal distribution)*		

4.7.3 Comparison of primary outcomes by cohort (on versus off protocol)

Probability values (p-values) for the comparisons of each of the four primary outcomes by protocol status can be observed in Table 12 below. The comparisons are provided for both the combined sample and the separated cohorts. The specific exercise cohort is further sub-divided by flexion and extension subgroups. Specifically, the stabilization cohort's total direct health care costs and total physical therapy costs were the only two statistically significant findings among all the comparisons. On-protocol stabilization subjects cost the Health Plan only half as much (50.9%, $336,412.73) as those defined as off-protocol ($660,359.24). In addition, it was significantly cheaper for the Health Plan to provide physical therapy services to stabilization subjects that were on-protocol ($110,896.99) compared to those subjects that were off-protocol ($140,002.31). Those significant findings are highlighted in bold in Table 12 below.

Table 12. Comparison of primary outcome variables by cohort (on versus off protocol)

	Combined Cohort (N=750)	**Manipulation (N=103)**	**Stabilization (N=498)**	**Specific Exercise (N=149)**	**(Flex / Ext) (N= 74 / 75)**
Total Direct Health Care Costs					
Sig.	*0.1210*	*0.1753*	***0.0126†***	*0.3871*	*0.2933 / 0.8833*
Total Member Cost Burden					
Sig.	*0.3711*	*0.3406*	*0.4082*	*0.1238*	*0.0614 / 0.9532*
Direct Physical Therapy Costs					
Sig.	*0.1557*	*0.4403*	***0.0312†***	*0.4987*	*0.7419 / 0.6073*
Physical Therapy Member Cost Burden					
Sig.	*0.5105*	*0.9264*	*0.0946*	*0.9878*	*0.7203 / 0.5325*

† = Significant finding with p-value < 0.05

Note: All comparisons made with cost data transformed by [log (10)]

4.7.4 Hierarchical linear regression analysis to explain variation in primary outcomes

A total of 24 (4 outcome variables * 6 TBC cohorts) hierarchical liner regression models were produced in this analysis and are recorded in Table 13 below. The cohort effect (the status of being on or off protocol) accounted for significant variation after adjustment of all baseline characteristics and interactions in the previous two models whose outcomes were determined to be statistically significant during earlier comparisons. Specifically, the stabilization group's total health care costs and total physical therapy costs were the only two outcomes among all the TBC combined and separated subgroups that were determined statistically significant by earlier rank sum comparisons (See Table 12). These were the probabilities computed for the added cohort effect to those two linear models: Total health care costs (F-value = 10.15, Pr > F = 0.0015), and total physical therapy costs (F-value = 6.50, Pr > F = 0.0111). Both models are highlighted bold in Table 13 below.

Table 13. Final hierarchical linear regression models for primary outcomes by TBC cohort

	Final Model
Combined	
1. Total health care costs	$y = 2.58 - 0.29_{LOB(CM)} + 0.03_{Pain} + 0.0005_{Age} + 0.02_{SM} + 0.006_{FABQ\_WK} - 0.25_{AT(N)} + 0.007_{Age*AT(N)}$
2. Total member burden	$y = 0.66 + 0.99_{LOB(CM)} + 0.005_{Age} + 0.004_{OSW} + 0.01_{SM} + 0.11_{♀}$
3. Total PT costs	$y = 2.75 - 0.01_{OSW} - 0.41_{CCI} - 0.003_{Age} - 0.12_{LOB(CM)} - 0.003_{Age} -$

4. PT member burden	$0.12_{LOB(CM)} + 0.09_{AT(N)} + 0.01_{SM} + 0.02_{OSW*AGE}$ $y = 1.86 – 0.02_{Pain} – 0.02_{SM} + 0.004_{Age} + 0.004_{Pain*SM}$
Manipulation	
1. Total health care costs	$y = 3.17 – 0.35_{LOB(CM)}$
2. Total member burden	$y = 0.83 + 1.11_{LOB(CM)} + 0.03_{SM}$
3. Total PT costs	$y = 2.36 + 0.04_{Pain} – 0.19_{SMC}$
4. PT member burden	$y = 2.39 – 0.08_{Pain} -0.07_{SM} + 0.01_{Pain*SM}$
Stabilization	
1. Total health care costs †	$\boldsymbol{y = 2.36 – 0.30_{LOB(CM)} + 0.04_{Pain} + 0.006_{Age} + 0.01_{SM} + 0.14_{Cohort(Off)}}$
2. Total member burden	$y = 0.71 + 0.94_{LOB(CM)} + 0.005_{Age} + 0.16_{♀} + 0.006_{FABQ\_WK} + 0.02_{SM}$
3. Total PT costs †	$\boldsymbol{y = 2.52 + 0.002_{OSW} + 0.09_{AT(N)} – 0.09_{LOB(CM)} + 0.003_{FABQ\_WK} – 0.08_{Cohort(On)}}$
4. PT member burden	$y = 1.62 + 0.005_{Age} + 0.02_{Pain}$
Specific Exercise	

1. Total health care costs 2. Total member burden 3. Total PT costs 4. PT member burden	$y = 3.73 + 0.01_{FABQ\_WK} - 0.67_{CCI}$ $y = 1.33 + 1.0_{LOB(CM)}$ $y = 1.83 - 0.06_{CCI} + 0.02_{SM} + 0.003_{OSW} + 0.01_{OSW*CCI}$ $y = 1.65 + 0.02_{SM} + 0.005_{OSW}$
Flexion 1. Total health care costs 2. Total member burden 3. Total PT costs 4. PT member burden	 $y = 2.96 + 0.01_{FABQ\_WK}$ $y = 1.31 + 1.07_{LOB(CM)}$ $y = 2.45 + 0.02_{SM}$ $y = 1.77 + 0.009_{OSW}$
Extension 1. Total health care costs 2. Total member burden 3. Total PT costs 4. PT member burden	 $y = 2.85 + 0.01_{FABQ\_WK}$ $y = 1.39 + 0.90_{LOB(CM)}$ $y = 2.42 + 0.005_{OSW}$ $y = 1.81 + 0.02_{SM}$

Linear Model is of general form : $y = \alpha + \beta_{1x1} + \beta_{2x2} + \beta_{3x3} + \ldots \beta_{kxk}$
- Based on cost outcome variables transformed by [log (10)]

† = Cohort (on versus off-protocol) status accounted for a significant amount of variation in final model after all baseline variables and possible interactions were adjusted for.

Subscript Abbreviations: *(1) LOB(CM) = Line of Business (Commercial Insurance), (2) SM = Study-Months, (3) AT(N) = Active Treatment Status (No), (4) OSW = Oswestry, (5) ♀ = Female Gender, (6) CCI = Charlson Cormibidity Index Score, (7) SMC = Sought Medical Care in Past*

4.7.5 Hierarchical logistic regression analysis to predict if protocol status is predictive of unusually high costs and unusually high member burden

Logistic hierarchical regression modeling was performed to assess the ability of protocol status to predict the dichotomous outcomes (Y / N) associated with higher downstream costs defined as being in the top 25% of spending in this sample. To calculate these variables, we examined 4 primary outcomes (top 25% of total direct net healthcare costs, top 25% total physical therapy costs, top 25% total member burden, and top 25% total physical therapy member burden), resulting in 24 separate logistical regression analyses (e.g., a regression analysis for each outcome (four) and for each of the treatment-based classifications (N=6).

The cohort effect (being on or off-protocol) was predictive after adjustment for incurring the Upper 25th percentile of total direct healthcare costs in both the combined sample (Q_3 = \$1,659.40) and in the stabilization cohort (Q_3 = \$1,598.88). After adjustment, off-protocol subjects in this sample were more likely to exceed the top quarter of direct healthcare spending (Pr > Chi-Square = 0.0205, adjusted OR = 1.514, 95%CI (1.066, 2.149) then on-protocol subjects. Similarly, stabilization off-protocol subjects were more likely to exceed the top quarter of direct health care spending (Pr > Chi-Square = 0.0095, adjusted OR = 1.791, 95%CI (1.153, 2.782) after adjustment then those treated as on-protocol. Lastly, protocol status also contributed greatly after adjustment to stabilization subjects incurring member cost burdens in the top 25% (Q3 = $259.81). Again, the off-protocol group were more likely to incur higher costs for this outcome measure (Pr > Chi-Square = 0.0336, adjusted OR = 1.611, 95%CI (1.038, 2.501). All three models are highlighted in bold, and all of the final logistic models for the top spending quartiles can be viewed in Table 14 below.

Table 14. Final hierarchical logistic regression models for top spending quartiles by TBC cohort

	Final Model
Combined	
1. **Top 25% total costs †**	$\boldsymbol{logit_{(p)} = 3.04 - 0.43_{LOB(CM)} + 0.02_{OSW} + 0.06_{SM} + 0.11_{Pain} + 0.21_{Cohort\ (0ff)}}$
2. Top 25% total member burden	$logit_{(p)} = -3.80 + 1.23_{LOB(CM)} + 0.12_{Pain} + 0.12_{Age}$
3. Top 25% PT costs	$logit_{(p)} = -2.56 + 0.09_{Pain} + 0.02_{OSW} - 0.21_{LOB(CM)}$
4. Top 25% PT member burden	$logit_{(p)} = -3.76 + 1.70_{LOB(CM)}\ 0.03_{Age} + 0.20_{AT(N)}$

Manipulation 1. Top 25% total costs	$logit_{(p)} = -2.30 + 0.14_{SM} – 0.57_{LOB(CM)}$
2. Top 25% total member burden	$logit_{(p)} = -2.45 + 0.13_{SM}$
3. Top 25% PT costs	$logit_{(p)} = -0.94 – 0.53_{SMC}$
4.Top 25% PT member burden	$logit_{(p)} = -0.18 – 0.06_{FABQ\_PA}$
Stabilization **1. Top 25% total costs †**	$\boldsymbol{logit_{(p)} = - 3.35 – 0.40_{LOB(CM)} + 0.02_{OSW} +0.07_{SM} + 0.13_{Pain} +0.29_{Cohort(off)}}$
2. Top 25% total member burden †	$\boldsymbol{logit_{(p)} = -3.75 + 1.48_{LOB(CM)} + 0.02_{OSW} + 0.11_{Pain} + 0.24_{Cohort(off)}}$
3. Top 25% PT costs	$logit_{(p)} = -2.64 + 1.95_{LOB(CM)}$
4. Top 25% PT member burden	$logit_{(p)} = -2.47 + 0.02_{OSW} + 0.11_{Pain} – 0.26_{LOB(CM)}$
Specific Exercise 1. Top 25% total costs	$logit_{(p)} = -1.73 + 0.05_{\mathrm{FABQ\_WK}}$
2. Top 25% total member burden	$logit_{(p)} = -2.00 + 1.28_{\mathrm{LOB(CM)}}$
3. Top 25% PT costs	

4. Top 25% PT member burden	$logit_{(p)} = -3.91 + 0.03_{OSW} + 0.12_{SM}$ $logit_{(p)} = -2.04 + 1.18_{LOB(CM)}$
Flexion 1. Top 25% total costs	$logit_{(p)} = -1.53 + 0.87_{♀}$
2. Top 25% total member burden	$logit_{(p)} = -2.45 + 0.12_{SM}$
3. Top 25% PT costs	$logit_{(p)} = -4.81 + 0.17_{SM} + 0.04_{OSW}$
4. Top 25% PT member burden	$logit_{(p)} = -1.80 + 1.08_{LOB(CM)}$
Extension 1. Top 25% total costs	$logit_{(p)} = -1.92 + 0.07_{FABQ\_WK}$
2. Top 25% total member burden	$logit_{(p)} = -1.55 + 0.04_{FABQ\_WK}$
3. Top 25% PT costs	$logit_{(p)} = -0.85 - 0.49_{AT(N)}$
4. Top 25% PT member burden	$logit_{(p)} = 0.38 - 0.20_{Pain}$

Logistic Model is of general form : $Logit_{(p)} = \alpha + \beta_{1x1} + \beta_{2x2} + \beta_{3x3} + \ldots \beta_{kxk}$

† = Cohort (on versus off-protocol) status accounted for significant variation in final model after all baseline variables and possible interactions were adjusted for.

Subscript Abbreviations: *(1) LOB(CM) = Line of Business (Commercial Insurance), (2) SM = Study-Months, (3) AT(N) = Active Treatment Status (No), (4) OSW = Oswestry, (5) ♀ = Female Gender, (6) CCI = Charlson Cormibidity Index Score, (7) SMC = Sought Medical Care in Past*

4.8 SPECIFIC AIM 2 ANALYSIS

Specific Aim 2: *Develop a decision analysis model containing cost data from the retrospective Specific Aim 1 results and outcome / terminal end state data inferred indirectly from the literature in order to obtain the Incremental Cost Effectiveness Ratios (ICER's) for both treatment arms, thus gaining a better assessment of the overall cost-effectiveness of the TBC approach compared to a usual care approach.*

4.8.1 Incremental Cost-Effectiveness Rations (ICER's)

Based on the results of the decision analysis model, the TBC on-protocol strategy essentially dominated[1] the off-protocol strategy. The mean net cost for the on-protocol strategy was $1,732.80 while the net effectiveness was 0.7900 quality adjusted life-years (QALY's). This equates to about $2,193.00 per QALY. Meanwhile, the mean net cost for the off-protocol strategy was $2,545.70 while the net effectiveness for that treatment arm was 0.7560 QALY's. Therefore the cost for this strategy was $3,367.00 per QALY. This is a difference of $1,174.00 per QALY in favor of the on-protocol treatment strategy for this study. Furthermore, it cost the payer $812.00 less per individual while gaining an additional 0.034 QALY for the on-protocol strategy clearly making this the dominant choice in this model. The comparative cost-effectiveness between the two competing strategies can be viewed in Figure 3 below.

[1] The term dominated in cost-effectiveness analysis means a strategy is both cheaper and has more benefit.

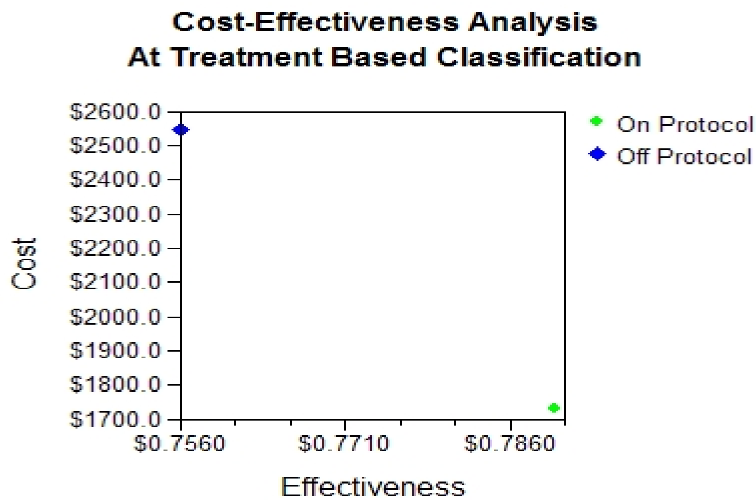

Figure 3. Cost-effectiveness graph of on versus off-protocol strategies

4.8.2 One-Way Sensitivity Analysis

After varying all the potential variables individually across their distributions, only two variables make the ICER positive. In other words, only by changing these two-variables does the off-protocol strategy become more effective. These two variables were 1: The probability that a subject were to get better given that they were on-protocol, and 2: The probability that a subject were to get better given that they were off-protocol. The results of the one-way sensitivity analysis can be observed in the Tornado Diagram in Figure 4 below.

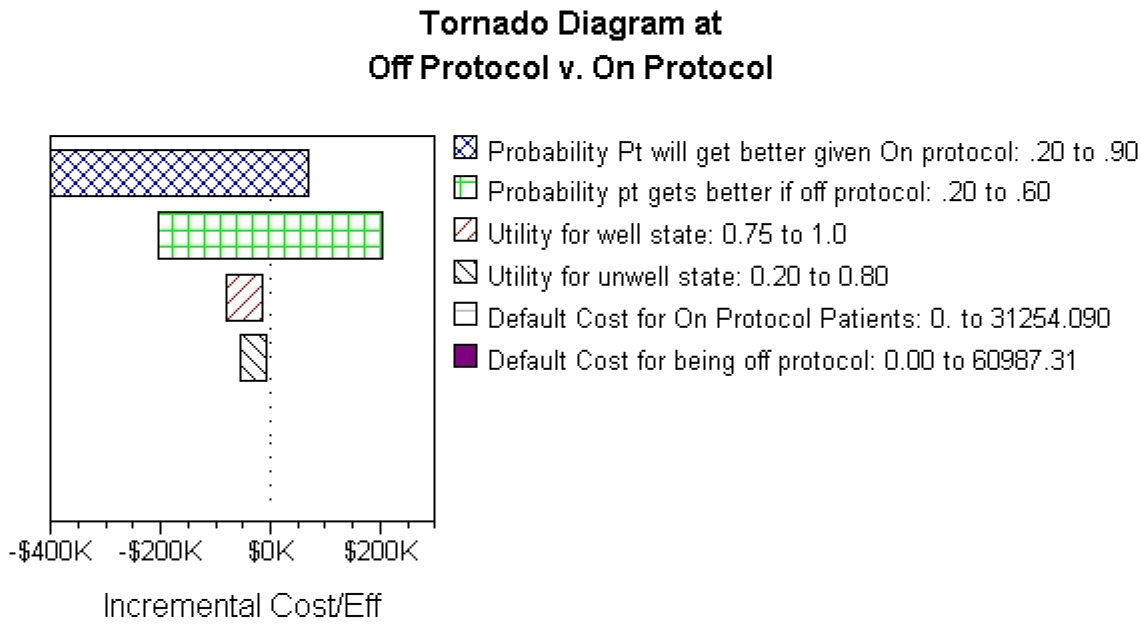

Figure 4. Results of one-way sensitivity analysis

If the probability of getting better given on-protocol status were varied from its' base case of 50% in a negative trend, the on-protocol strategy would continue to dominate until it reached a threshold of 33%. At this level, the off-protocol strategy becomes as effective (0.756 QALY's), but not as cost-effective. To illustrate, at a 33% probability of getting better given an on-protocol status, the off-protocol strategy would cost $3,367.72 per QALY ($2,546 / 0.756 QALY's). Meanwhile, the on-protocol strategy would cost $2,292.33 per QALY ($1,733 / 0.756 QALY's). This is still a $1,075.39 per QALY difference in favor of the on-protocol strategy. In fact, the on-protocol strategy remains more cost-effective even when the probability of getting better is driven all the way to the bottom of its distribution (20%). In this case, the off-protocol strategy is now more effective (0.756 QALY's) compared to the on-protocol strategy (0.744 QALY's), but there is still an associated incremental cost-savings of $813 in favor of being on-protocol. Therefore at 20% probability of getting better given adherence, the on-protocol strategy ends up costing $2,329.33 per QALY ($1,733 / 0.744 QALY's) while the off-protocol strategy costs $3,367.72 per QALY($2,546 / 0.756 QALY's) . Thus, the on-protocol strategy is still cheaper by $1,038.39 per QALY. This means that the probability for getting better given the on-protocol strategy could still be 13% less then the base-case presented for the alternative "usual care"strategy (33%), and the on-protocol strategy would still be considerably more cost-effective.

Similarly, the off-protocol strategy continues to be dominated in this model until the probability of getting better given off-protocol status variable is varied up to 51% (from a baseline 33%). At this point, the off-protocol status is now more effective (0.79 QALY's on versus 0.792 QALY's off), but it is still not more cost-effective then the on-protocol strategy ($2,193 per QALY on versus $3,214 per QALY off). Even when increased to the upper end of its distribution (54%), the influence of this variable still does change the fact that the on-protocol strategy is still more cost-effective. For example if the probability of getting better given an off-protocol classification was increased from it's baseline of 33%to 54%, the cost difference would still be $997 per QALY in favor of being on-protocol.

4.8.3 Two-Way Sensitivity Analysis

By decision modeling convention, the two variables that made the ICER positive (improved outcome effectiveness for off-protocol over on-protocol) during the one-way sensitivity analysis were also introduced into a two-way sensitivity analysis. In this fashion the model performed simulations while varying both distributions at the same time versus one at a time as before. Once again, the following variables included in two-way sensitivity analysis were 1: The probability of getting better given on-protocol status, and 2: The probability of getting better given off-protocol status. The graph in Figure 5 below depicts this analysis with a payer willingness to pay (WTP) for benefit threshold of $50,000. According to the graphic intersections, if the probability of getting better given off-protocol status is < 28% (only 5% below expected baseline), the on-protocol strategy is always favored. Likewise, if the probability of getting better given on-protocol status is > 52% (just 2% above expected baseline), the model still always favors the on-protocol strategy.

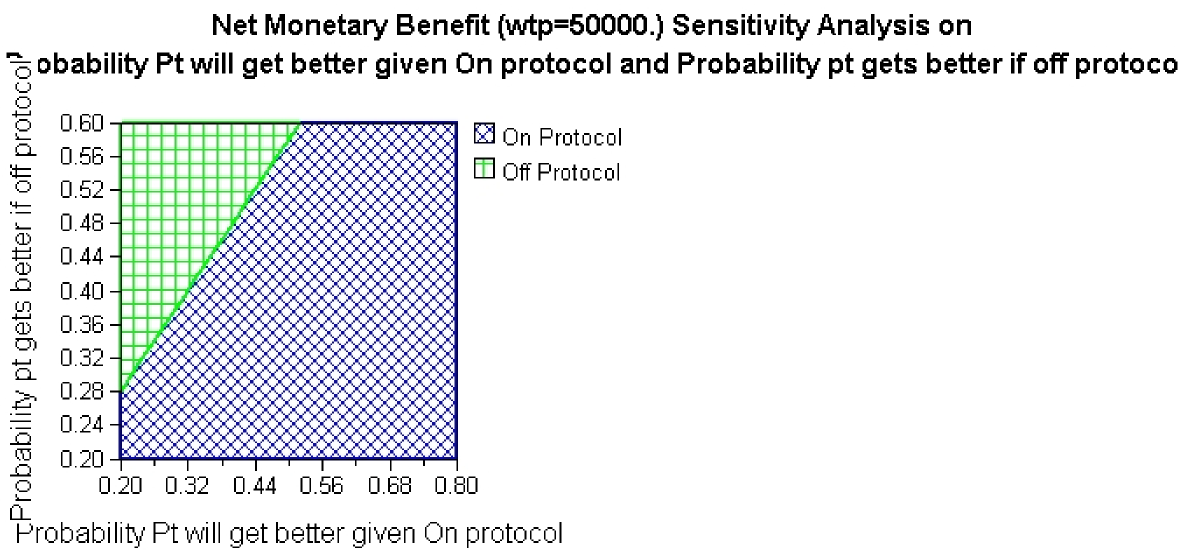

Figure 5. Results of two-way sensitivity analysis

4.8.4 Monte Carlo Sensitivity Analysis

When varying all potential variables simultaneously across each of their distributions within the model, the on-protocol strategy was favored in 100% of the Monte Carlo iterations.

This essentially means it is a superior choice for a payer even when the willingness to pay (WTP) threshold is expanded all the way up to $200,000 per QALY. In other words, when examining across a given dollar range of what a group is willing to spend, the Monte Carlo analysis acceptability curve can present one with the likelihood of a particular treatment strategy being more cost-effective compared to alternative strategies. As indicated, the on-protocol status in this case is likely more cost-effective at all willingness to pay (WTP) thresholds at least up to 200k per QALY. This can be observed in Figure 6 below.

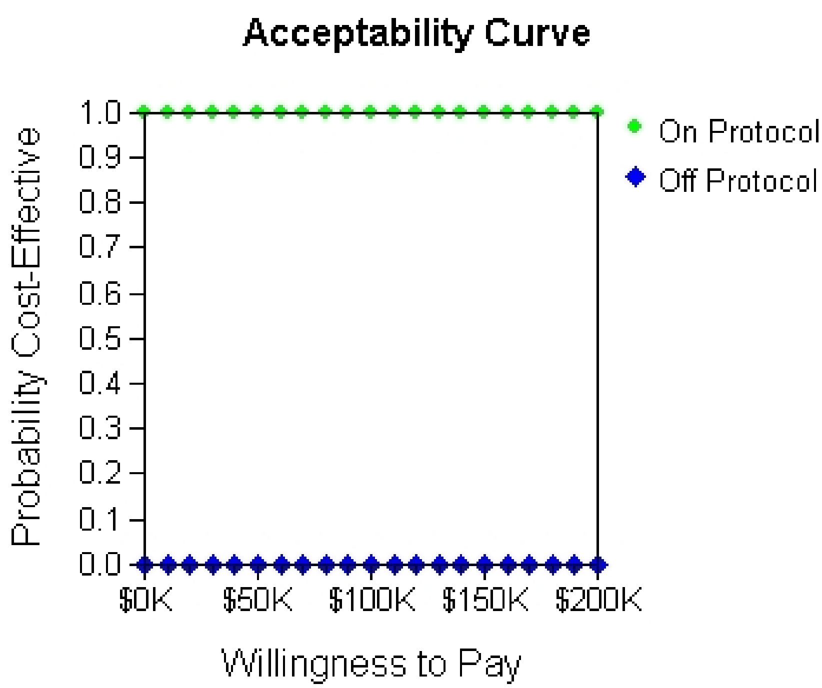

Figure 6. Result of Monte-Carlo sensitivity analysis

5.0 DISCUSSION

5.1 TBC COHORT BREAKDOWN FINDINGS

At least 20% of the eligible sample for this study was not able to be classified and were therefore excluded from the analysis. According to the TBC algorithms, if an individual does not distinctly fit into one of the subgroup classifications, a therapist can examine for other presenting factors that may determine which category "best fits" the patient. (Fritz JM, Cleland JA, & Childs JD, 2007) For example a factor against a manipulation classification beyond the distinctive algorithm recommendations is that a patient has no pain with spring testing, in which case, a patient should not be included in this treatment category. The database and current MDS template used in this study however, was not inclusive enough to capture those additional factors beyond the basic critical variables of the main algorithms, which did not allow decision rules to include additional variables necessary to match all patients. In addition, the TBC approach is still evolving. Current classifications may even need to be expanded or constricted over time and new classifications may arise. Until, the TBC algorithms are optimized, it is not unreasonable to expect that a portion of subjects will be "unclassifiable".

The protocol adherence results are consistent with what is reflected in the literature. Both the manipulation and specific exercise cohorts were unequally distributed and unbalanced in terms of the proportion of subjects classified as on or off-protocol. For example, a very small portion (< 14.0%) of manipulation candidates actually received the recommended treatment. As a general rule of thumb, despite the fact that not only does evidence exist supporting Grade V "thrust" or manipulative procedures, and that providers are supportive of their clinical application, those interventions are consistently underutilized in physical therapy. (Delitto A, & Erhard RE, 2003) On the other hand, recent surveys suggest that therapists most often rely on McKenzie specific exercise type activities, which may lead to broad overutilization of these interventions. (Jette, DU, et al, 2003) Both of these may explain the large proportion of on-protocol subjects in the specific exercise category (>80%) while the opposite finding is true for the manipulation cohort.

Another, reason for discrepancies related to manual therapy and specific exercise may be the relative emphasis placed on training in entry-level professional physical therapy programs. As part of our surveillance procedures in this quality improvement initiative, physical therapists have recently stated "lack of confidence in manual skills" as a reason for non-adherence in the manipulation cohort. Instead of being trained to be more reliant on substitutes for manual therapy such as physical agents and passive treatments, these and previous findings speak to the need to focus training on ensuring that newer graduates are more confident with their manual skills.

Adequate training at the entry-level does not explain fully the non-adherence rates in the manipulation cohort, however. In fact, the American Physical Therapy Association's (APTA) Section of Orthopedics has increased their efforts to incorporate manual therapy at entry level programs across the country since 2001. In addition, the Commission of American Physical Therapy Education (CAPTE) has taken these recommendations into consideration in their accreditation reviews of all entry-level programs. Therefore the evidence has demonstrated a need for better manual therapy strategies, but the overall effect of these efforts has not come to fruition at this time. While these adherence rate findings are low, they are still consistent with the literature which demonstrates an underutilization of manual therapies. In contrast, it may even be plausible that manipulation adherence rates in this study are actually a little higher then the norm given that many of the physical therapists in this sample were graduates the University of Pittsburgh, a program that prides itself on rigorous manual training and comprehensive instruction on the TBC algorithms with all graduates passing competency-based assessments in grade V thrust procedures. In lieu of these finings, this may speak to a further need for implementing QI based adherence initiatives and compliance based programs driven by executive oversight and perhaps reimbursement strategies to ensure that these skills are not lost upon graduation. Preliminary evidence from this study has at least demonstrated the success associated with an adherence-feedback program to improve compliance with obtaining MDS data.

5.2 BASELINE CHARACTERISTIC DIFFERENCES

In this study, only 39% of the sample was males. This may be somewhat surprising because LBP is not a condition that is typically prevalent in one gender over another. For example previous findings have demonstrated fairly even samples sizes of both men and women. (Fritz JM, Cleland JA, & Brennan GP, 2007) The gender variable in this study was acquired from the Health Plan after the 750 subjects had been classified, so the gender utilization rates for LBP before exclusion were not determined for comparison. However, this figure is still relatively comparable to a report published in Spine in 2001 by members of the UPMC Health Plan. (Vogt M, et al., 2005) That study examined the characteristics and usage of 17,148 patients with LBP who made claims in 2001. In that report, 7,477 (43.6%) of subjects were men. Given that information, it is quite possible the unreported incidence for LBP in Southwestern Pennsylvania is relatively equal, but perhaps more women then men actually seek care on a regular basis. A one month point estimate of 24,105 CRS visits in the January of 2008 demonstrated that the gender utilization rate was only 46.1% for men. This figure is lower then Fritz et al. (2007) study, but still higher then these results which cannot completely account for apparent gender bias. A gender bias toward women seeking more treatment for LBP would be consistent with another study that demonstrated surgical treatment rates for LBP in Iowa Blue-Cross Blue Shield subscribers were likely to be increased if the member was female and over 44 years of age. (McGuire, SM, Phillips KT, & Weinstien, 1994) In another recent study, adherence to an active physical therapy treatment strategy was greater with male subjects (71%) versus female subjects (29%). (Fritz JM, Cleland JA, & Brennan GP, 2007) The difference between gender adherence in that study was proposed to be related to higher levels of men with workers compensation claims that emphasized improvement of physical function and return to work. Certainly, all these findings speak to the point that differences in care patterns may exist across genders. For comparisons in the current study however, at least no statistical differences were found examining gender proportions by protocol status for any of the defined cohorts.

According to the CCI scores obtained for this analysis, this sample appears to have been relatively healthy (mean = 0.16, range 0 – 2). This does not appear to be an unusual finding given this group of subjects was all working age patients that were treated in an ambulatory outpatient setting. The CCI is commonly used in studies examining the survival related to the treatment of cancer and other common illnesses. In retrospect, using a different scale, such as the Functional Co-Morbidity Index (FCI), may have been more sensitive in adjusting for baseline levels of co-morbid illness in this study. The FCI has been shown to explain more variance in physical function scores compared to other common indexes such as the Charlson and Kaplan-Feinstein which are designed to predict mortality.(Groll DL, To T, Bombardier C, & Wright JG, 2005) In addition, other studies have demonstrated that the presence of multiple comorbidities as identified by the FCI is associated with lower functional outcome scores and subsequently increased levels of physical disability.(Hart DL, Wang YC, Stratford PW, & Mioduski JE, 2008a, 2008b; Hart DL, Ying-Chih W, Stratford PW, & Mioduski JE, 2008) Unfortunately, it was just not feasible at this time to obtain those scores since the CCI is the common metric that the Health Plan utilizes to access co-morbid illness. Nonetheless, the CCI is still one of the most reliable, valid, and widely used tools to access co-morbid illness.

The proportion of patients treated according to an active treatment strategy (38.4%) as defined in the methodology above was slightly lower but reasonably close compared to earlier published rates of around 41.4%%. (Fritz JM, Cleland JA, & Brennan GP, 2007) It is quite possible that the numbers in this study are more representative of the true population because of a better methodological approach, which allowed the investigator to capture this variable from actual provider treatment behaviour versus CPT or charge based data which is often altered by coders to maximize reimbursement. Nonetheless, by simply knowing that active treatment strategies are correlated with improved outcomes, one can deduce that a reasonable amount of savings might be realized just by slightly improving the low proportion of subjects that are treated adherently to this principle. This provides even further support for QI based adherence initiatives.

Approximately 1/3rd of the study sample had sought medical care (SMC) in the past for LBP. Recidivism reflects chronicity and is highly consistent with existing evidence that demonstrates recurrence rates of 22% - 33% within the first 6 months of LBP onset. (Cassidy JD et al., 2005; Croft PR et al., 1998) If there is any long-term benefit to be gained by the TBC, one should expect at least a small portion of that figure to be minimized which ultimately translates into a significant savings for the payer. This is completely relevant to the UPMC QI initiative which has two main goals: 1. Decrease chronicity transition rates 2. Decrease recurrence rates. This lends even more support for treatment compliance and outcomes monitoring particularly given the findings of this study.

The mean scores for both FABQ subscales in this study were below the threshold cutoffs (PA = 19, WK = 29) for bio-behavioral intervention, which is suggestive of a low incidence of fear avoidance, a known factor that has been associated with a failure to recover. Bi-weekly CRS compliance monitoring has demonstrated that adherence to fear-avoidance recommendations is extremely poor (<10%). Once again, this has substantial relevance to both UPMC QI goals since high levels of fear-avoidance are predictive of long-term disability. So, if these patients can be easily identified, and targeted for intervention to reduce chronicity transition, this could be another area where significant cost savings may be realized if adherence can be improved. Areas to explore include professional development initiatives and other educational endeavors that emphasize a more bio-behavioral approach to patient care when indicated.

The percentage of subjects (19.2 %%) who reported with government sponsored Medical Assistance (MA) in this study was consistent with the overall percentage of the population in Southwestern, PA. For example, the Pennsylvania Department of Public Welfare's Office of Income Maintenance Statistical Reports shows that 15.3% of the population in Allegheny County receives state sponsored Medical Assistance compared to 16.5% statewide. (2009)

Lastly, mean scores for the Oswestry Low Back Pain Disability Questionnaire (OSW) and numeric pain rating (NPR) scores were consistent with other physical therapy LBP related studies. (Fritz JM, Cleland JA, & Brennan GP, 2007) The mean OSW score of 38.2 and the NPR of 7/10 for the combined sample is suggestive of moderate physical disability which is fairly typical of someone seeking physical therapy care for treatment of their LBP.

5.3 DISTRIBUTION NORMALITY TESTING FOR CONTINUOUS BASELINE VARIABLES

All seven baseline characteristic distributions were non-normally distributed. A normal distribution is an important assumption required of independent-*t* testing for statistical comparisons. Though convenient both practically and statistically, the assumption of normality in data such as that used in this study is rarely accurate. (Keselman HJ, Cribbie RA, & Wilcox, R., 2002) All things being equal, statistical significance is more difficult to demonstrate with non-parametric testing when compared to parametric testing. (Greenhalgh T, 1997) The use of non-parametric statistics in this study demonstrates a very conservative approach to the analysis of the data and thus speaks to the robustness of the findings.

5.4 MISSING DATA

A concern with missing data is that this it is not a random phenomenon, and subjects who have incomplete data are more likely to have poorer outcomes, which presents a potential confound. Whenever subjects with missing data are excluded from an analysis, investigators may increase their risk of obtaining biased results which favor the intervention variable. This sample had a missing data rate of 4.8% which is fairly close to a similar study comparing active treatment adherence which demonstrated a missing data rate of 3.5%. (Fritz JM, Cleland JA, & Brennan GP, 2007) Whereas that study excluded individuals with missing data, the choice was made in this case to determine if the subjects with missing data were inherently different at baseline then those with complete data sets. There was no baseline differences between the group with missing data versus the cohort with complete data, which justified assigning a missing code to those incomplete data points for any subsequent analysis throughout this study. As indicated in the methods, all stepwise regressions procedurally eliminated the subjects with any missing variables.

5.5 BASELINE CHARACTERISTIC COMPARISONS

5.5.1 Combined Cohort

For the combined sample, the on-protocol subgroup had significantly higher raw FABQ_PA scores *(mean = 13.97, SD = 6.54, med. = 14.0)* then the off-protocol subgroup *(mean = 12.76, SD = 7.34, med. = 13.0)*. Although this was statistically significant, this does not appear to be a clinically meaningful difference since: (1) the magnitude of the difference is so small (one-point) and (2) both means are well below a score 19.0 which is the cut-off threshold for recommended bio-behavioral intervention. In addition, the combined sample off-protocol group had a slightly higher proportion (22.9%) then the on-protocol group (16.7%) of individuals with Medical Assistance versus commercial insurance. Out of 144 total Medical Assistance Patients, 63 ($43.7%) were classified as off-protocol, and 81 (56.3%) were classified as on-protocol. This proportion of non-adherence for government insured (Medicaid) individuals was higher than that obtained in a recent study examining adherence to active physical therapy treatment, where only 10 (22%) out of 45 subjects were treated according to adherent care. (Fritz JM, Cleland JA, & Brennan GP, 2007) That study did include a much smaller number of subjects with government sponsored insurance (45/1190 or 3.8%), which may explain the overall differences in proportion. Another explanation would be in the geographical variation between Utah (Fritz et al. demographic) versus our sample base of Western Pennsylvania. Regardless of the relatively small differences in proportions between the studies, both findings suggest a troubling fact that individuals with Medical Assistance are receiving disproportionately more non-adherent care on the first visit to physical therapy. In our attempt to explain such a disparity, we believe that t is unlikely that physical therapist choice of management strategies is a result of providers knowing the subject's insurance carrier. Our sources note that in the case of patients who fit the category of Medicaid, there is a high prevalence of documentation errors specific to the requirements needed by government regulations for patients who fit the category of Medical Assistance and the fact that the most common reason for such errors is because the therapist was unaware of the insurance status of the patient. This begs to question of whether there are other associated characteristics such as race, apparent socioeconomic status or geographic location of care that are attributable to this difference. For example, studies have shown that quintiles and costs of services for Medicaid beneficiaries vary markedly among geographic regions with the research demonstrating an association between health care costs, supply of beds, and the number of specialist physicians available. (Bodenheimer T, 2005a, 2005b, 2005c; Fisher ES et al., 2003b) Certainly, this pattern would be worth examining further in future studies focusing on evidence based-guidelines of care.

5.5.2 Manipulation Cohort

No differences were observed between any of the baseline characteristics in the manipulation cohort.

5.5.3 Stabilization Cohort

Similar to the combined sample findings, the on-protocol subgroup of the stabilization cohort had significantly higher raw FABQ_PA scores *(mean = 13.59, SD = 6.73, med. = 14.0)* at baseline compared to those classified as off-protocol *(mean = 11.43, SD = 7.38, med. = 12.0)*. Again, even though this appears to be a statistically significant finding, it is not likely a clinically significant finding as the difference is small in magnitude and both group's scores are well below the threshold cut-off score (19) for bio-behavioral intervention. A significantly higher proportion (43.1%) of stabilization subjects classified as off-protocol sought medical care in the past for their LBP compared to those on-protocol (31.8%). This finding speaks to the point that often physical therapists ask patients about past care and patient responses influence subsequent clinical decisions. Physical therapists are sensitive to individuals that have sought care in the past, and are more likely to exhibit an overreliance on care administered in the past without regard to the patient's present presentation, which may indicated a different strategy. The generic prescription of stabilization exercises can create a further pattern of over-utilization of stabilization exercise in spite of the presence of examination data that would suggest a different approach would be more effective, which was clearly the case in this analysis. Out of the 363 patients who were negative prediction rule candidates (strong evidence that stabilization exercises would not be effective), (Hicks GE et al., 2005), as many as 63.1% (229) received stabilization exercises. Meanwhile, as much as 82% (111) of the 135 positive to the prediction rule candidates received stabilization exercises as well. These findings do not even account for the others in the combined sample who also routinely received stabilization exercises as treatment. So instead of relying on the exam findings and the recommend algorithms that suggest otherwise, clinicians appear to habitually prescribe stabilization exercises which contribute to both over-utilization and improper utilization of services.

5.5.4 Specific Exercise Cohort

The on-protocol group of the combined specific exercise cohort had significantly more study-months *(mean = 11.32, SD = 3.73, med. = 12.30)* of enrollment in the investigation compared to the off-protocol group *(mean = 9.39, SD = 4.38, med. = 8.44)*. With this finding, on-protocol subjects would have been afforded more of an opportunity to accrue charges then those who were off protocol. Based on the results of this study, this group only experienced a slight average increase member burden ($4.39) and physical therapy member burden ($19.72), meanwhile a considerable savings was observed for total costs. Additionally, the number of study-months (SMC) contributed largely to the final linear model explaining the difference in the specific exercise physical therapy member burden costs (F = 6.07, pr > F = 0.0153). Therefore, these findings lend further support to the proper use of the TBC algorithms.

In contrast to the SMC findings, the off-protocol group of the combined specific exercise cohort had a significantly higher proportion of Medical Assistance subjects (35.7%) compared to the on-protocol group (13.79%). These findings are consistent with the combined sample and other evidence suggesting that as a whole, this group of patients is more likely to receive non-adherent care.

5.5.5 Specific Exercise Cohort Subgroups

When the combined specific exercise cohort was broken down into its two main subgroups, the same statistical trend demonstrating a significantly larger portion of Medical Assistance patients was observed. (Flexion: 57.1% off-protocol vs. 18.3% on-protocol; Extension: 14.3% off-protocol vs. 9.8% on-protocol) The difference in the extension subgroup may be smaller because there were only a total of 7 Medical Assistance patients that were included in the analysis, an underrepresentation that questions the stability of this estimate. In a combined cohort where the compliance with the TBC exceeded 80%, only 61.5% of Medical Assistance patients were treated on-protocol. Again, this disparity is consistent with the earlier findings in the overall sample, which suggests underlying reasons exist beyond the scope of this study to explain that this group of patients is more likely to receive non-adherent or poorer care across the health care spectrum. Specific exercise training is the hallmark of physical therapy education and is cited repeatedly as the most often used intervention by physical therapists in treating people with LBP, which illustrates again that other factors are contributing to this disparity.

The off-protocol subgroup of the specific exercise extension cohort had significantly higher initial pain scores *(mean = 8.36, SD = 2.06, med. = 9.0)* then the on-protocol subgroup *(mean = 6.95, SD = 2.43, med. = 8.0)*. Ninety percent, (9 / 10) patients with Medical Assistance who were classified as off-protocol for specific exercise in this study had pain scores > 6. In addition, 40% had FABQ_WK subscale scores > 29, and 60% had FABQ_PA subscale scores > 19. None of these individuals received a bio-behavioral intervention, again adding to the propensity of these subjects to be treated with non-adherent strategies. The higher than normal proportion of bio-behavioral symptoms in patients with Medical Assistance may provide some guidance in explaining the disparate results obtained in this study.

Lastly, the specific exercise flexion cohort was older *(mean = 52.05, SD =9.70, med. = 56.347)* then the off-protocol subgroup *(mean = 44.67, SD = 11.53, med. = 46.62)*. This may be expected because of the increased prevalence of patients with lumbar spinal stenosis, a disease that has increasing prevalence with age and typically leads to interventions that suggest specific exercise with a flexion directional preference.

While some of these findings do demonstrate trends across the cohorts, the likelihood of spurious findings exists because of the multiple comparisons.

5.6 SPECIFIC AIM 1 ANALYS

Specific Aim 1: *Perform a cost minimization analysis comparing both direct health care and physical therapy costs for subjects classified as on-protocol according to the three major subsets (manipulation, specific exercise, and stabilization exercise) of a TBC approach to those who are considered off-protocol (receiving a usual physical therapy care approach) in the management of patients with LBP in the primary outpatient physical therapy setting. In addition, member total cost burden and physical therapy cost burden will be compared by protocol status to determine if treatment status influences a subject's actual out-of-pocket direct health care expenses.*

Potential Cost-Savings: Health Plan

The consistent pattern of cost-savings (Table 10.) demonstrated in this study likely represents an underestimate of the overall potential cost-savings that could be realized by the payer through greater adherence to the quality improvement initiative. In fact, the magnitude of the underestimate is probably great given: (1) the considerable lack of adherence to some of the protocol classification schemas such as the manipulation prediction rule, and the negative stabilization prediction rule guidelines; (2) protocol status was defined only on the initial physical therapy visit and did not even include adherence with subsequent patient visits where active treatment strategies and bio-behavior approaches (when indicated) are critical. We calculated that with a 5.35% increase in physical therapy spending as a proportion of total health care costs in the combined on-protocol group, a total savings of $283,419.61 was observed, indicating that a small percentage increase for physical therapy spending as a proportion of total direct healthcare spending may be a worthwhile trade-off for a payer.

Potential Cost-Savings: Member Burden

According to Table 10 above, the on-protocol groups for each cohort predominantly demonstrated an average cost savings for each of the 4 primary outcomes: total direct net healthcare costs, total member burden, total direct physical therapy costs, & total physical therapy member burden. However, one of the outcome measures that the on-protocol group did not demonstrate a cost savings in was total member burden for the flexion cohort and its parent, the specific exercise cohort. As pointed out earlier, these two groups had a significantly higher portion of Medical Assistance patients who were off-protocol. Generally, one of the benefits of state sponsored programs such as Medical Assistance is that beneficiaries have limited to no co-payments for their health care. A larger percentage of Medical Assistance patients that have little to no co-payments would explain lower out of pocket costs in this cohort.

Lastly, no cost savings was demonstrated by being on-protocol for physical therapy member burden in the combined specific exercise, flexion, and manipulation cohorts. It is unclear though why physical therapy member burden was greater for the manipulation on-protocol group considering there were no significant differences described at baseline. However, similar to the payer results for the combined cohort, it was seemingly a worthwhile trade-off for these subjects to have paid slightly more out of pocket for physical therapy services because they experienced an overall larger cost-savings in total health care member burden for their LBP. In other words, they paid approximately $13.00 more per member for physical therapy care while saving slightly under $109.00 per member in total health care burden, thus netting approximately $96.00 in total savings per member for the treatment of their LBP.

5.6.1 Statistical Comparisons of Primary Outcome Variables

Although a considerable cost savings with the TBC algorithm approach was clearly demonstrated in this study, there were only 2 statistically significant differences among all the cohort comparisons. Both differences were found among the stabilization cohort, which is by far the largest of the 3 TBC subgroups. This cohort experienced a statistically significant cost savings in total direct health care costs *(mean on-protocol = $1,373.11 versus mean off-protocol $2,610.12)* and in total direct physical therapy costs *(mean on-protocol = $473.92 versus mean off-protocol $521.51)* for the on-protocol group.

Others may refute the potential TBC associated cost-savings in this study due to a lack of statistical significance beyond the two aforementioned outcome measures. It is important to note however that a lack of statistical significance should not preclude one from overlooking a potentially more meaningful finding such as a potential cost savings, especially when examining costs from a payer or member perspective. In fact, Karl Claxton, the current co-editor for the Journal of Health Economics argues the following; "*the rules of statistical inference are arbitrary and entirely irrelevant to the decisions which clinical and economic evaluations claim to inform. Decisions should be based only on the mean net benefits per net cost irrespective of whether differences are statistically significant or fall outside a Bayesian range of equivalence. Failure to make decisions in this way by accepting the arbitrary rules of inference will impose costs which can be measured in terms of resources or health benefits forgone.*"(Claxton K., 1999) Therefore, net savings should be discussed only in terms of the net benefits derived (ICER) in order to make a more properly informed policy decision based on value and efficiency versus statistical significance.

5.6.2 Hierarchical Linear Regression Analysis

The results from the linear regression analysis supported all the earlier rank-sum comparisons that demonstrated a statistical difference in outcomes based on protocol status, thus giving further credence to the TBC. Therefore after adjustment a subject's protocol status contributed significantly in explaining the variation between the differences in total direct health care costs and physical therapy costs for the stabilization cohort. Furthermore, it is important to point out that the stabilization cohort was also the largest in combined sample and accounted for the highest percentage of total health care costs (62.3% or $996,771.97) and total member burden (67.8% or $136,178.59). In addition, being off-protocol for stabilization ($82,054.52) was 1.5 times more expensive to members then being on-protocol ($54,124.07) in terms of "out of pocket" costs, even though this was not considered significant statistically.

Next, is also equally important to note that effect of being on or off protocol accounted for none of the variation in the models of the 5 outcomes: total member burden for the flexion group and its parent specific exercise group, and total physical therapy member burden for the same two groups along with the manipulation group. Factors other than protocol status were responsible for explaining the variation in those outcomes. The one characteristic that was consistent in at least 2 of those models was the insurance line of business (LOB) variable. (the combined specific exercise group and its flexion subgroup) In fact, LOB was the only variable that was included in both of those models, meaning that subjects with Medical Assistance accounted for the greatest amount of variation in explaining the cheaper member burden for the off-protocol groups in these two cohorts. Total direct health care costs for those off-protocol groups was higher, which directly translates to more health care provided, and lends support to the rational that member burden was only cheaper for those off-protocol because of the limited co-payment structure for Medical Assistance beneficiaries. To illustrate further, when compared to the extension subgroup, the flexion subgroup had a total member burden that was 1/3 less in proportion to total costs then the extension subgroup. For instance the total costs for the extension subgroup was \$170,920.21 and the total member burden was \$20,925.22 (12.24%). Meanwhile the total costs for the flexion subgroup was \$241,248.67 and the total member burden was \$19,902.67 (8.2%). In addition, the proportion of member burden / total cost was higher for the flexion on-protocol (10.07 %) compared to the flexion off-protocol group (3.0%). Therefore, these findings most plausibly are explained by the greater preponderance of Medical Assistance patients in both the combined specific exercise and flexion off-protocol subgroups. The linear models in Table 13 only lend more credibility to this contention.

Certainly it is becoming clearer through these findings that overall more net health care was diverted to treating the patients with Medical Assistance. One may argue that patients with Medical Assistance may be more chronic, but there was no evidence to suggest through either direct comparisons or through linear regressions that the flexion, the specific exercise, nor the combined off-protocol patients sought medical care more often in the past for their LBP. These were the three cohorts with the largest and statistically significant proportions of Medical Assistance subjects compared to those commercially insured. If this inference is true that more care is given to patients with Medical Assistance, it is not clear from this study that finding is related to increased patient visits or increased services. Typically more visits equate to a higher volume of services, especially in physical therapy. It is quite likely then, this finding is from a combination of both, which may be the focus of a potential follow-on study examining utilization rates for on versus off-protocol status. In addition, this was not one of the aims of this analysis. Given this interpretation and the increased prevalence of LOB variable (45.8%, 11 out of 24) over all other variables in the final linear models, it would further suggest the need for future studies to examine the differences in physical therapy care patterns across insurance types. If perhaps, there are defining characteristics or interaction effects that promote non-adherence to beneficiaries of government subsidized medical insurance, then the next step would be to examine cost-effective methods to improve overall compliance.

5.6.3 Hierarchical Logistic Regression Analysis

Perhaps the most intriguing finding is that off-protocol status was now found to be predictive of being in the top 25% (Q3 $\geq$ 1,659.40) of total direct healthcare spending for the combined sample. Furthermore, after accounting for all other variables, the adjusted OR of 1.514 [95% CI ;(1.066, 2.149)] for protocol status was higher than any other contributors in the final model. Not only was a substantial net cost-savings demonstrated with the TBC given this sample, but this evidence should be even more convincing in regards to the economic merits of the TBC. This makes it all the more important to now target this group with proper treatment because if the top-spending body be contained, the Health Plan stands to realize a steady cost-savings that can now be dedicated to other areas of the QI initiative (e.g., education, professional development, information technology investment, etc.)

Consistent with the previous linear regressions and rank-sum comparisons, the cohort effect (being on or off-protocol) was also predictive of total healthcare spending in the stabilization cohort. In addition, since the stabilization cohort was the biggest of the subgroups, these findings would be expected given the results for the logistic modeling in the combined sample. Thus, not only does a cohort effect account for the difference in total healthcare and physical therapy savings, it has further predictive ability to determine who will be the top spenders or those more likely to exceed the top75% (Q_3 = $1,598.88) of total healthcare expenditures as stabilization candidates with LBP.

Protocol status did not contribute however to the final model as a predictor for the top 25% (Q_3 = $634.00) of total physical therapy costs. This was not consistent given the previous statistical significance associated with the rank-sum comparisons and the linear regression findings. Perhaps the most striking conclusion of this finding is that it does not appear to cost any more to treat a person on-protocol as it does to treat them off-protocol. The relative small proportionate cost of physical therapy compared to overall cost expenditures for back pain must also be taken into consideration. For example, physical therapy spending for the stabilization cohort was only about 25.2% ($250,899.30) of that group's total healthcare expenditure of $996,771.97 thus potentially creating a smaller margin for differences between protocol statuses to exist at the upper quartile of spending for that variable.

Once more, the LOB variable was also a significant predictor in half (12 / 24) of the final regression models. In fact, it was also the sole predictor for being in both the top 25% (Q_3 = $356.73) of total member burden costs [Pr > Chi-Square = 0.0191, adjusted OR = 11.42; 95%CI (1.49, 87.50)] and in the top 25% (Q_3 =$188.10) of physical therapy member burden costs [Pr > Chi-Square = 0.0235, adjusted OR = 10.53; 95%CI (1.37, 80.74)] in the specific exercise cohort. Furthermore it was the sole predictor of being in the top 25% (Q_3 =$193.54) of physical therapy member burden costs [Pr > Chi-Square = 0.0420, adjusted OR = 8.76; 95%CI (1.08, 70.85)] in the flexion subgroup. In both cohorts, it was reported previously that the on-protocol subjects had higher member burdens. However, since the total costs were greater in all cases for the off-protocol subjects, these findings only add to the prevailing evidence from this study that individuals with Medical Assistance are treated with more non-adherent care. Again, an attempt to study both the clinical and social characteristics of Medical Assistance patients may help to explain why, and provide the necessary evidence in order to create meaningful change.

5.7 SPECIFIC AIM 2 ANALYSIS

<u>Specific Aim 2:</u> *Develop a decision analysis model containing cost data from the retrospective Specific Aim 1 results and outcome / terminal end state data inferred indirectly from the literature in order to obtain the Incremental Cost Effectiveness Ratios (ICER's) for both treatment arms, thus gaining a better assessment of the overall cost-effectiveness of the TBC approach compared to a usual care approach.*

5.7.1 Interpretation of Decision Analysis Model

It was clearly evident from this model that a TBC adherent strategy is far superior to a non-adherent strategy. The TBC was demonstrated to be far more cost-effective then a "usual" care approach to treating LBP. Even when the distributions of the two-most ICER sensitive variables were varied individually and simultaneously, creating a net effectiveness (benefit) gain for the off-protocol treatment arm, the on-protocol strategy was still considerably more cost-effective. From a policy or payer perspective, the inference should be fairly obvious that the TBC is the "best" choice. The next step would be then to consider how to improve adherence levels among providers.

5.7.2 Further Support for Implementing TBC Compliance Measures

Investing in measures to improve compliance and adherence to TBC standards can be argued to decrease the cost-effectiveness of this QI initiative. However, when compared to the magnitude of potential gains, any reasonable investment in these areas is than likely to be offset. The preliminary work developing and implementing MDS and surveillance in this study has already demonstrated considerable effectiveness for changing clinical behavior using a combined audit / feedback and incentive based program. Based upon that data, it is logical to believe that one could also significantly improve TBC on-protocol compliance with a similar styled intervention. The question then arises at what threshold does it then become cost-effective for a major payer such as the UPMC Health Plan to drive the implementation of the program. Finally, the question of how much a provider should share in this burden comes to light. For example, how much should a payer versus provider invest in infrastructure, education, surveillance and professional development activities in a QI initiative? To address this problem further, the primary investigator (PI) previously developed a separate decision-analysis model which showed that only an 8% improvement in compliance (baseline = 42%) was required to ensure the cost effectiveness of institutionalizing an adherence program strategy that promoted active therapeutic interventions. (McGee J. Smith K, 2008)

That decision analytic model was used to compare the cost-effectiveness of implementing an educational/audit based guideline adherence program to a usual care approach without an adherence strategy. Probabilities for adherence and transitioning among health states were obtained from the literature and related preliminary data.(Cassidy JD et al., 2005; Croft PR et al., 1998; Jette AM, Smith KS, Haley SM, & Davis KD, 1994; Jette DU & Jette AM, 1996; Martin BI et al., 2008; Smith MT, Carmody TP, & Smith MS, 2000; Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, & Vincent D, 2000) In addition, costs and benefits were acquired from the literature as well as the University of Pittsburgh Medical Center (UPMC) Health Plan.(Fritz JM, Cleland JA, & Brennan GP, 2007; Martin BI et al., 2008; Wasiak R et al., 2006) The analysis was conducted from a payer perspective over a two year time frame. Discounting was not applied because of the limited time horizon. In the base case analysis the adherence program was assumed to cost $8,500 per 1000 patients enrolled. One-way, two-way and probabilistic sensitivity analyses were also performed.

Overall, implementing an adherence program cost $706 less per patient while gaining 0.11 QALY, dominating the alternative strategy of not implementing an adherence program. In one-way sensitivity analyses, results were most sensitive to the probability of guideline adherence with or without the adherence program. The program remained cost saving if total program costs were ≤ $43,000/1000 patients. Finally, probabilistic sensitivity analysis, varying all parameters simultaneously over distributions, found the adherence program cost saving in 91% of the 10,000 model iterations.

Based on this model, an adherence program that promotes active therapeutic interventions is both less costly and more effective than a usual care rehabilitation approach without an adherence program. These results provide further meaningful information to the current research findings in order to help determine the possible parameters of a future TBC adherence QI initiative. The additional cost-savings from improved TBC compliance could provide an even more substantial revenue stream whereby, recouped resources could be put to more efficient future use for the Health Plan and its subscribers.

5.7.3 Limitations of Decision Analysis Model

As with any model, the findings are only as dependable as the assumptions made before the onset. Since this was a simple decision tree, it does not account for repeated health states or outcomes. For example a patient may get better under either arm and then get worse again, only to be treated again under a different strategy. This changing dynamic is normally addressed by adding Markov states and assigning cycle lengths or time horizons so that the model understands certain proportions of movement from state to state is expected over a given time period. Markov states were not utilized because most long term outcome time horizons based on TBC evidence does not exceed 18 months. Some have only been analyzed up to 6 months. Therefore, one would only be speculating on outcome transitions beyond this time period which may threaten the generalizability of the current results. In addition, the aim of this analysis was more focused on accessing overall outcomes based on initial visit classification versus accessing outcomes per cycle. Improved outcome reporting and a larger time horizon from three to five years would help provide a better understanding for this. Additionally, this data would allow future research that could break cost structure down across TBC states instead of having to depend on just the findings of the combined sample. This is very helpful because chronic LBP is often associated with more expensive co-morbidities such as diabetes, obesity and bad habits such as smoking, increased stress, and lack of physical activity. As a result, these suggested modifications should only strengthen the findings more.

An attempt was made to examine the final Oswestry (OSW) scores from the CRS data base in order to see if these findings were consistent with the utilities assigned for the decision analysis model. The on-protocol group had a mean OSW change of 4.91 $\pm$ 11.90, while the off-protocol group had a mean OSW change of 5.49 + 11.2 creating a net change of 0.58 in favor of the non-adherent strategy. The rank-sum comparison for this variable demonstrated no significance with a p-value = 0.5959. It is also important to note that both means are below the minimum clinical important difference (MCID) of 6 for the OSW. Nonetheless, these outcome differences are well below what would be expected for on-protocol given the previous trials that have favored the separate TBC subgroups. However, out of the 750 subjects in the sample, 19 were missing final OSW scores. In addition, 388 (51.7%) individuals only had 1 physical therapy visit accounted for according to the CRS database so the initial OSW scores were carried forward n the analysis thus likely minimizing any potential effect. Out of those, 203 / 380 (53.42%) were on-protocol subjects. Furthermore, it is not reasonable to assume the baseline OSW scores would be accurate for this many subjects at follow-up. It is important to note however, that these figures are not an actual reflection of the true number of physical therapy visits because the CRS data does not match the UPMC charge data beyond the first visit. While it is highly encouraged to capture exam data points for every LBI visit, it is not a practice that is even moderately well adhered to. This evidence provides even further support for a separate outcomes reporting manager or unit for CRS that could capture LBI outcomes at 6 weeks to even 6 months out or more. However this intention is well beyond the aims of this investigation.

Another potential limitation is that this decision analysis model was based solely upon healthcare data from only one geographical region thus possibly limiting external generalizability beyond Southwestern Pennsylvania. Nonetheless, the results from this study may attract other agencies and health plans across the U.S. to collaborate on similar projects in the future.

6.0 CONCLUSION

The evidence from this analysis supports the Aim 1 hypothesis that adherence to the TBC approach on the initial physical therapy visit is less costly or for the payer then an off protocol approach. An incremental cost-savings for all outcome measures was demonstrated predominantly in the combined sample and in the separate TBC cohorts during the initial cost-minimization analysis. The off-protocol group had a cost-savings generally associated with member burden for the specific exercise and flexion cohorts that were primarily explained in this analysis by the type of insurance carrier versus a cohort or protocol status effect. A greater portion of individuals carried Medical Assistance versus commercial insurance in those subgroups which likely skewed the member burden outcomes in favor of the off-protocol cohort due to the lack of co-payments associated with that patient population. Nonetheless, it was obvious from these findings that these subjects received more health care as evidenced by their total direct expenditures, and more often that care was delivered in a non-adherent manner. In general though, a net incremental combined cost-savings was experienced by the on-protocol group for all primary outcome variables. Not all primary outcome variables differed between on and off-protocol cohorts. Inference based on statistical significance should not preclude one from realizing the more important finding here that there was an associated cost-savings with the TBC. The two significant findings however were explained by the cohort effect of being off-protocol after all other variables in this study were adjusted for, thereby adding further credibility to the TBC. Moreover, logistical regression analysis demonstrated that the cohort effect of being off-protocol was predictive of a member incurring charges in the top 25% of direct health care expenditures for thus study. This make targeting this group for improved care all the more essential as it is often a small percentage of the population that generally accounts for the greatest amount of health care expenditures. Lastly, the decision analysis model demonstrated overwhelming support for the TBC as a substantial cost-effective alternative to "usual care" for LBP in the primary outpatient physical therapy outpatient setting. This supported the Aim 2 hypothesis provided at the onset of this study that the compliance with TBC is more cost-effective then "usual" care. Limitations and suggestions for improving the model were also discussed.

Aside from providing support for the TBC, this study has generated further research questions that may help improve processes and reduce further variation which may make treating LBP from the payer perspective even more efficient. Certainly it is clear that other cost-savings may exist simply by improving adherence not only to the TBC, but also to active treatment strategies and to proper fear-avoidance interventions. In addition, findings from this study seem to suggest the importance and feasibility of QI initiatives to improve compliance in all these areas as well as improving outcome reporting processes to further substantiate the claims of this study. At a minimum, the results should be helpful in establishing compliance goals for seemingly underutilized algorithms and guidelines (e.g. manipulation and fear-avoidance behavioral interventions). Next, it would be worthwhile to pursue at least a cross-validation attempt on this same sample with 6-12 additional months of data. This would provide further even further credibility to these findings. The rest will be included in the cross-validation comparison using similar methodology as this study. In addition, a totally separate sample could be collected to validate this analysis but this would likely take a longer time to obtain meaningful results. Moreover, now that the cost-effectiveness of the TBC has been demonstrated, this allows physical therapy to "put its best foot forward" in managing LBP against other comparative strategies such as primary care and chiropractic care from a health economic standpoint. It may be beneficial however to expand the current algorithm TBC definitions to include the Stage II classifications that have been recommended for this approach. It would also be reasonable to examine both the clinical and social characteristics of Medical Assistance patients in this sample. Perhaps then one can determine why this group predominately received more non-adherent care so that corrective measures can be taken as necessary to ensure equality of care and to assist the payer in properly channeling the limited health care resources currently available. This analysis also provides substance for health care economists to consider the bundling of charges in a manner that is rewarding those providers who are exceeding standards of care by compliance with evidence-based guidelines.

APPENDIX A

MINIMUM DATA SET (MDS) VARIABLES AND WEB REPORTING INSTRUCTIONS FOR UPMC LBI

LOW BACK PAIN FORM

DEMOGRAPHICS (Initial Only)

Status: ☐ Licensed PT ☐ Student PT Date (Initial): ____________________

Patient ID: ______________________________ Gender: ☐ Male ☐ Female

---

HISTORY (Initial Only)

Location (check one)
- ☐ LBP
- ☐ LBP and buttock/thigh symptoms (not distal to knee)
- ☐ LBP and leg symptoms distal to knee

Duration
- ☐ ≤ 15 Days
- ☐ > 15 Days

Location of other symptoms (check all that apply)
- ☐ N/A
- ☐ Head/Neck
- ☐ Thoracic Spine
- ☐ Upper Extremity (ies)
- ☐ Hip(s)
- ☐ Knee(s)
- ☐ Foot/Feet

FABQ
PA ________
WK ________

Post Surgical
- ☐ Yes
- ☐ No

Sought medical care for this same episode in the past?
☐ Yes ☐ No

Previous episodes of LBP ☐ 0 ☐ 1-2 ☐ 3-5 ☐ >5

Frequency Increasing ☐ Yes ☐ No

PHYSICAL EXAM: ☐ **Initial** ☐ **Follow-up** **Date:**_______________

Avg SLR
- ☐ ≥ 91
- ☐ < 91

Prone Instability Test
- ☐ Positive
- ☐ Negative

Mobility Testing
- ☐ Hypo
- ☐ Normal
- ☐ Hyper

Directional Preference
- ☐ Extension
- ☐ Flexion
- ☐ No Directional Preference

Aberrant Movements
- ☐ Yes
- ☐ No

Pain (worst): ________ Flexion ROM: ________ Oswestry: ________

TREATMENT CLASSIFICATION (Initial & Weekly)

Stage I (check one)
- ☐ Thrust Manipulation (Grade V)
- ☐ Non Thrust Manipulation (Grade I-IV)
- ☐ Stabilization
- ☐ Flexion Directional Preference
- ☐ Extension Directional Preference
- ☐ Traction

Stage II (check all that apply)
- ☐ Aerobic
- ☐ General Conditioning

FABQW Status (check one)
- ☐ Negative (<29)
- ☐ "At Risk" (29-34)
- ☐ Positive (>34)

FABQPA Status
- ☐ Positive (>14)
- ☐ Negative (≤14)

<u>**NOTE: You must check**</u>
1. One Stage I category ***or*** one or more stage II categories ***and***
2. One FABQ status (initial only; weekly optional)

INTERVENTIONS (Initial & Weekly) (check all that apply)

- ☐ Patient Education/Instruction
- ☐ Flexion Exercises
- ☐ Extension Exercises
- ☐ Flexibility Exercises
- ☐ Stabilization Exercises
- ☐ General Conditioning Exercises
- ☐ Thrust Manipulation (Grade V)
- ☐ Non Thrust Manipulation (Grade I-IV)
- ☐ Aerobic Exercise
- ☐ Functional Training
- ☐ Heat Modalities
- ☐ Cold Modalities
- ☐ Traction – Mechanical
- ☐ Traction – Autotraction
- ☐ De-weighting / Unloading
- ☐ Behavioral Exercise Approach
- ☐ NMES (Strengthening)
- ☐ NMES (Pain Control)
- ☐ Soft Tissue Massage
- ☐ Myofacial Release
- ☐ Craniosacral Therapy
- ☐ Other

Demographics

- Fill out this section at the initial visit only
- **Status**: Licensed PT or Student PT – check only one box based on who is the primary therapist for the patient
- **Date**: Fill out date in the following format: dd / mm / year (for example, 25/03/2008)
- **Age**: Fill in the patient's age in years
- **Gender:** Check either male or female for the patient's sex

---

HISTORY

- **Location:** You must check only 1 of the 3 boxes.
  - **LBP** – This represents symptoms that can extend from T 12 down to the lumbo-sacral junction (L5/S1).
  - **LBP and buttock/thigh symptoms** –Symptoms extend below the lumbo-sacral junction, as far as the popliteal crease of the knee.
  - **LBP and leg symptoms distal to the knee** – Symptoms extend below the popliteal crease of the knee.
- **Duration**
  - You must check only one box: "≤ 15 days" or "> 15 days".
  - This represents the duration of symptoms for this episode of LBP.
- **Location of other symptoms**

- Check boxes representing all areas where the patient is currently experiencing symptoms, even if unrelated to his/her LBP.

– **FABQ**

  - This is where you will report the scores of the Fear Avoidance Beliefs Questionnaire.
  - Insert the score from the Physical Activity subscale after the letters "PA" (the range of scores is 0-24)
  - Insert the score from the Work subscale after the letters "WK" (the range of scores is 0-42)

– **Post-Surgical**

  - Check "Yes" if the patient has had surgery to the lumbar spine. Do not check "yes" if he/she has had surgery to the thoracic or cervical spine regions.

– **Sought medical care for this same episode in the past?**

  - You must check only one box. Check yes only if the medical care was for THIS EPISODE of LBP. If care was sought for a prior episode only, you should check "no".

– **Previous episodes of LBP**

  - The patient is asked about the number of prior episodes of LBP that have caused him/her to miss work or reduce his functional activity level. Check the appropriate box to represent this number.

– **Frequency Increasing**

- You must check only one box. Check “yes” if the frequency of episodes of LBP (that cause the patient to miss work or reduce functional activity levels) are increasing.

PHYSICAL EXAM

- Fill this section out at the initial visit and at follow-up visits. For follow-up visits, enter the date of the visit (in dd / mm / year format), and the visit number.
- **Visit number:**
    - You fill out the actual visit number for the patient. Include the initial visit. For example, if a patient has been to your clinic for an initial visit and 3 follow-up visits, you would enter “4” for the visit number. If there was ever an occasion where the patient came to your clinic and was billed for care, but you did not see the patient, this would still count as a visit.
- **Avg SLR**
    - The patient is supine with the head relaxed. The examiner holds the foot with one hand to maintain the hip in neutral rotation. The inclinometer is positioned on the tibial crest just below the tibial tubercle. The leg is raised passively by the examiner, whose other hand maintains the knee in extension. The leg is raised slowly to the maximum tolerated straight leg raise (not the onset of pain). The maximum straight leg raise is recorded in degrees. The opposite leg is then tested in the same manner. Average straight leg raise is computed by adding the maximum straight leg raise of the left and right legs and dividing by two.
- **Prone Instability Test**

- The patient lies prone with the body on the examining table and legs over the edge and feet resting on the floor. While the patient rests in this position, the examiner applies posterior to anterior pressure (PA) to the lumbar spine. Any provocation of pain is noted. Then the patient lifts the legs off the floor (the patient may hold table to maintain position) and posterior compression is applied again to the lumbar spine.
- Positive Test - If pain is present in the resting position but subsides substantially (either reduces in severity/intensity, or resolves) in the second position, the test is positive. Mild improvement in symptoms does not constitute a positive test.
- Negative Test – If pain is present in the resting position, but does not subside substantially in the second position, the test is negative. Further, if the patient did not have any pain provocation with PAs, then you should mark “negative”.

– **Mobility Testing**
  - Mobility or spring testing is performed by placing the hypothenar eminence (just distal to the pisiform) of the hand over the spinous process of the segment to be tested. With the elbow and wrist extended, the examiner applies a gentle but firm, anteriorly-directed pressure on the spinous process. Interpretation of whether a segment is hypomobile, normal, or hypermobile should be based on the examiner’s anticipation of what normal mobility should feel like at that spinal level and compared to the mobility detected in the spinal segments above and below the segmental level of interest.
  - The following options are available for each level tested:

        - Hypomobility – Passive mobility is judged to be hypomobile at ≥ 1 lumbar spine segmental level
        - Normal - Passive mobility is judged to be normal throughout the lumbar spine (L1-L5)
        - Hypermobility - Passive mobility is judged to be hypermobile at ≥ 1 lumbar spine segmental level
    - Note that you are able to check both “Hypo” and “Hyper” if you find ≥1 lumbar spinal segment that is hypermobile and ≥1 lumbar spinal segment that is hypomobile. However, if you check “normal”, this implies that all segments (from L1-L5) exhibited normal mobility.
- **Directional Preference**
    - This term focuses on selecting a particular direction of exercise that exhibits a centralization of symptoms with lumbar movement testing during the initial examination and can include extension, flexion, or no directional preference. Note that centralization is defined as when a movement or position results in the migration of symptoms from an area more distal or lateral in the buttocks and/or lower extremity to a location more proximal or closer to the midline of the lumbar spine.
    - Extension – Mark this if your patient’s symptoms centralize with repeated extension movements/exercises
    - Flexion – Mark this if your patient’s symptoms centralize with repeated flexion movements/exercises

- No Directional Preference – Mark this if your patient's symptoms do not centralize with either repeated flexion or repeated extension movements/exercises

– **Aberrant Movements**

- Check "yes" if you observe any of the following **aberrant movement** (as defined below) during sagittal plane motion:
    - Instability catch: An instability catch is defined as any trunk movement outside of the plane of specified motion during that particular motion i.e., lateral sidebending during trunk flexion).
    - Painful arc (on descent or return): Symptoms felt during the movement at a particular point in the motion (or through a particular portion of the range) that are not present before or after this point.
    - Thigh climbing: Using the hands on thighs (or some other external support) to push up on when returning from flexion to the upright position.
    - Reversal of lumbopelvic rhythm: The trunk being extended first, followed by extension of the hips and pelvis to bring the body back to upright position.

– **Hip IR ROM**

- The patient lies prone. The examiner places the opposite leg of the leg to be measured in 30° of hip abduction to enable the tested hip to be freely moved into external rotation. The lower extremity of the side to be tested is kept in line with the body (ie neutral abduction/adduction), and the knee on that side is flexed to 90° with the ankle in the neutral position, and the leg in the vertical position. The inclinometer is placed on the distal aspect of the fibula in line with the bone and zeroed. Measurement of hip IR (hip rotated in a lateral direction [leg moved toward the edge of the plinth) is recorded at the point in which the pelvis first begins to move. The measurement should be recorded bilaterally. When measuring hip rotation, be sure that the knee remains in the same place (does not slide inward toward the opposite knee or outward away from the opposite knee).

    - Check whichever box is applicable: "$\geq$ 1 hip IR > 35°" or "No hip IR > 35°**.**

- **Pain (worst)**
    - Record the worst pain the patient has experienced in the past 24 hours (0-10 scale, 0 = no pain, 10 = worst pain imaginable)
- **Flexion ROM**
    - Lumbar range of motion is measured with a fluid-filled inclinometer. The patient stands erect. The inclinometer is held at T_{12}-L_1 and the patient is asked to reach down as far as possible towards the toes while keeping the knees straight. The measurement of total flexion is recorded in degrees.
- **Oswestry**
    - Simply insert the actual percentage score (0-100). <u>Do not enter the raw points obtained out of 50</u>. Use the procedure below to score the Oswestry.

Scoring the Oswestry

a. Assign a score to each section. Each section can be scored from 0-5, based on the selection chosen by the subject. If the subject marks the first response, assign a score of 0, the next response a 1, the next response a 2, and so on, with the final response being assigned a score of 5. Below is an example of the section called “Pain Intensity” with the corresponding score that should be assigned if that response is selected.

Pain Intensity

- I can tolerate the pain I have without having to use pain medication. **(0)**
- The pain is bad but I can manage without having to take pain medication. **(1)**
- Pain medication provides me complete relief from pain. **(2)**
- Pain medication provides me with moderate relief from pain. **(3)**
- Pain medication provides me with little relief from pain. **(4)**
- Pain medication has no affect on my pain. **(5)**

b. Add up the individual scores for each section.

c. Divide this result by 50, and report as a percentage (ex. 30/50 = 60%). In the event a subject does not complete each section adjust the denominator accordingly. For example, if the subject does not answer the question with respect to “Social Life”, divide by 45 instead of 50. Divide by 40 if they leave 2 sections blank, 35 if they leave 3 sections blank, and so on. **(Note: Therapists should always check to ensure all items are completed to minimize having to adjust the score.)**

d. Mark the score on the form and circle it.

TREATMENT CLASSIFICATION

- Fill out at the initial visit and follow-up visits.

- **<u>You must check</u>**:
  - One Stage I category OR one or more stage II categories <u>AND</u>
  - One FABQ status (initial entry for FABQ status is mandatory, weekly entries optional)
- **Stage I – Check only one box (Thrust Manip, Non-Thrust Manip, Stabilization, Flexion Directional Preference, Extension Directional Preference, Traction)**
  - Selection of Stage I is based on the patient meeting the following criteria:
    - Patients with higher levels of disability (Oswestry scores generally greater than 30%) and substantial reported difficulty with basic daily activities such as sitting, standing, and walking.
    - Management goals are to improve the ability to perform basic daily activities, reduce disability, and permit the patient to advance in his or her rehabilitation.
  - Thrust manipulation (Gr V) – Primary initial intervention approach is to improve mobility / decrease pain / decrease disability through the use of thrust manipulation to the lumbo-pelvic region
  - Non thrust Manipulation (Gr I – IV) - Primary initial intervention approach is to improve mobility / decrease pain / decrease disability through the use of non-thrust manipulation to the lumbo-pelvic region
  - Stabilization – Primary initial goal for therapy is to work on lumbo-pelvic stabilization / re-education / "core stability"
  - Flexion Directional Preference – Primary initial focus of intervention is to have the patient perform repeated flexion movements / exercises. For patients in this

classification, symptoms peripheralize with lumbar extension; symptoms centralize with lumbar flexion

  - Extension Directional Preference - – Primary initial focus of intervention is to have the patient perform repeated extension movements / exercises. For patients in this classification, symptoms centralize with lumbar extension; symptoms peripheralize with lumbar flexion
  - Traction - Signs and symptoms of nerve root compression, but no movements centralize symptoms

- **Stage II – Check one or both boxes as applicable: Aerobic or General Conditioning**
  - Selection of Stage II is based on the patient meeting the following criteria:
    - Individuals in the Stage II Classification include those whose symptoms are not acute and who are only having moderate difficulty with ADLs or work activities
    - Management goals are to improve strength, flexibility, and conditioning, or with a work-reconditioning program.
  - Aerobic – Check this box if aerobic conditioning is a primary management goal
  - General Conditioning – Check this box if management goals include working on improving strength, flexibility, or work reconditioning.
- **FABQW Status (check one box only)**
  - Negative – Mark this if the FABQW subscore is < 29 pts
  - At Risk – Mark if FABQW subscore is between 29 and 34 pts
  - Positive – Mark if FABQ subscore is > 34 pts
- **FABQPA Status (check one box only)**

- Positive – Mark if the FABQPA subscore is >14
- Negative – Mark if the FABQPA subscore is $\leq$ 14

INTERVENTIONS

- Check all boxes that apply at the initial visit and weekly
- **Patient Education/Instruction –** Includes verbal or written education/instruction provided to the patient
- **Flexion Exercises** – Includes any exercises designed to repeatedly flex the spine, such as double-knee-to-chest and single-knee-to-chest exercises
- **Extension Exercises–** Includes any exercises designed to repeatedly extend the spine, such as prone press ups or repeated extension exercises in standing
- **Flexibility Exercises –** Includes any exercises designed to improve muscle length or flexibility. Also includes self-mobilization and general mobility exercises (ie, pelvic tilts, hand-heel rocks, etc).
- **Stabilization Exercises** – Includes exercises designed to specifically strengthen the trunk musculature (ie, transversus abdominus, multifidus, lateral abdominal muscles, etc.)
- **General Conditioning Exercises** – Includes general strength and conditioning exercise such as calisthenics, general resistance training (ie, lifting weights), etc.
- **Thrust Manipulation (Gr V)** – Includes only thrust, or grade V, manipulation (also called small amplitude, high velocity manipulation)
    - **Check all regions that apply:** thoracic spine, lumbo-pelvic region, hips
- **Non Thrust Manipulation (Gr I-IV)-** Includes all forms of joint mobilization/manipulation that doesn't include thrust, or high velocity, technique

- **Check all regions that apply:** thoracic spine, lumbo-pelvic region, hips

– **Aerobic Exercise** – Exercise geared to improve the aerobic capacity of the patient (walking, jogging, running, cycling, stairstepper, etc)

– **Functional Training** – Exercises that are designed specifically to improve certain functional or job-related tasks

– **Heat Modalities** – Includes any physical modalities designed to increase the tissue temperature, such as ultrasound (include both pulsed and continuous here), moist heat packs, diathermy, etc

– **Cold Modalities -** Includes any physical modalities designed to decrease tissue temperature

– **Traction – Mechanical** – Includes traction that is performed to the lumbo-pelvic region through the use of mechanical traction device.

– **Traction – Autotraction** – An autotraction device must be used to check this box.

– **De-weighting / Unloading** – Check this box if you use some form of body-weight support / deweighting / unloading device. Typically, these devices support the patient in a harness and "unload" a portion of the patient's body weight. These are typically arranged over a treadmill so that the patient can walk while a portion of his/her body weight is unloaded, or supported, by the unloading device.

– **Behavioral Exercise Approach** – Includes use of principles of cognitive behavioral therapy in a physical therapy setting (graded exercise approach, use of quotas for exercise, focus on function > pain, focus on remaining active during episode of LBP, practicing/confronting fearful activities, etc.

– **NMES (strengthening) –** Any form of electrical stimulation that is selected / designed for strengthening

- **NMES (Pain control) –** Any form of electrical stimulation that is selected / designed to reduce pain
- **Soft Tissue Massage –** Includes soft tissue techniques that are not myofascial release or Craniosacral therapy techniques
- **Myofascial Release** – Includes any techniques used to improve the mobility of the skin and fascia
- **Craniosacral Therapy** – Includes any techniques specifically designed to target the craniosacral system
- **Neural Mobilization** – Includes any techniques specifically designed to "mobilize" or "tension" the neural system (ie slump stretching, "neural flossing", etc).
- **Other** – Check this box if you used a form of intervention that does not fit into any of the above categories

APPENDIX B

[TBC COHORT ON / OFF PROTOCOL DEFINITIONS]

	IF	AND	THEN COHORT: On Protocol vs. Off Protocol
1.	<u>Manipulation</u> Both Checked: a. Symptom duration $\leq$15 days b. LBP or LBP & buttock/thigh pain not distal to the knee OR Any combination of 3 of the following 4 checked : a. Symptom duration $\leq$15 days b. LBP or LBP & buttock/thigh pain not distal to the knee c. "Hypo" checked for mobility testing	Thrust Manipulation (Gr V) is checked as Intervention (Must mark lumbo-pelvic manipulation)	On Protocol

	IF	AND	THEN COHORT: On Protocol vs. Off Protocol
	d. FABQW <u><</u>19 ** In addition, patient must have an initial Oswestry score ≥ 30, and "None" checked under Directional Preference		
2.	<u>Manipulation</u> Both Checked: a. Symptom duration <u><</u>15 days b. LBP or LBP & buttock/thigh pain not distal to the knee OR Any combination of 3 of the following 4 : a. Symptom duration <u><</u>15 days b. LBP or LBP & buttock/thigh pain not distal to the knee c. "Hypo" checked for mobility testing d. FABQW <u><</u>19 ** In addition, patient must have an initial Oswestry score ≥ 30, and "None" checked under Directional Preference	Thrust Manipulation (Gr V) to the lumbo- pelvic region is <u>NOT</u> checked	Off Protocol

	IF	AND	THEN COHORT: On Protocol vs. Off Protocol
3.	<u>Stabilization</u> a. "Positive" checked for prone instability testing OR Any combination of 3 of the following 5 : a. "Positive" checked for prone instability testing b. "Yes" checked for aberrant movements c. Average SLR ≥ 91 d. Age < 40 years e. "Hyper" checked for mobility testing	Stabilization Exercises checked	On Protocol
4.	<u>Stabilization</u> a. "Positive" checked for prone instability testing OR Any combination of 3 of the following 5 : a. "Positive" checked for prone	Stabilization Exercises <u>NOT</u> checked	Off Protocol

	IF	AND	THEN COHORT: On Protocol vs. Off Protocol
	instability testing b. "Yes" checked for aberrant movements c. Average SLR ≥ 91 d. Age < 40 years e. "Hyper" checked for mobility testing		
5.	<u>Stabilization</u> Any combination of 3 of the following 4: a. "Negative" checked for prone instability b. "No" checked for aberrant movements c. FABQ_PA < 9 d. "Hypo" checked for mobility testing	Stabilization Exercises checked	Off Protocol Note : If the requirements for both this rule and the rule # 3 are met, then this rule takes precedence
6.	<u>Specific Exercise (flexion specific)</u> a. Location marked 'LBP and leg symptoms distal to knee' b. "Flexion" checked for Directional	Flexion Exercises Checked	On Protocol

	IF	AND	THEN COHORT: On Protocol vs. Off Protocol
	Preference		
7.	<u>Specific Exercise (flexion specific)</u> a. Location marked 'LBP and leg symptoms distal to knee' b. "Flexion" checked for Directional Preference	Flexion Exercises <u>NOT</u>checked	Off Protocol
8.	<u>Specific Exercise (extension specific)</u> a. Location marked 'LBP and leg symptoms distal to knee' b. "Extension" checked for Directional Preference	Extension Exercises checked	On Protocol
9.	<u>Specific Exercise (extension specific)</u> a. Location marked 'LBP and leg symptoms distal to knee' b. "Extension" checked for Directional Preference	Extension Exercises <u>NOT</u> checked	Off Protocol

B.1 APPENDIX SECTION

Appendix section's first paragraph.

Second paragraph.

B.1.1 Appendix subsection

This is a subsection (level-3 division) of appendix A.

BIBLIOGRAPHY

Abbott JH, McCane B, Herbison P, Moginie G, Chapple C, & Hogarty T. (2005). Lumbar segmental instability: a criterion-related validity study of manual therapy assessment *BMC Musculoskelet Disord, 6*, 56.

Abenhaim L, Rossignol M, Gobeille D, Bonvalot Y, Fines P, & Scott S. (1995). The prognostic consequences in the making of the initial medical diagnosis of work-related back injuries. *Spine, 20*, 791-795.

Adams ME, McCall NT, Gray DT, Orza MJ, & Chalmers TC. (1992). Economic analysis in randomized controlled trials. *Medical Care, 30*(3), 231-243.

Aina A, May S, & Clare H. (2004). The centralization phenomenon of spinal symptoms--a systematic review. *Man Ther, 9*, 134-143.

Al-Obaidi SM, Beattie P, Al-Zoabi B, & Al-Wekeel S. (2005). The relationship of anticipated pain and fear avoidance beliefs to outcome in patients with chronic low back pain who are not receiving workers' compensation. *Spine, 30*, 1051-1057.

American Physical Therapy Association, Guide to Physical Therapy Practice: 2nd edition. (2001). *Phys Ther, 81*, 9 - 746.

Assendelft WJJ, Morton SC, Yu EI, Suttorp MJ, & Shekelle PG. (2003). Spinal manipulative therapy for low back pain: a meta-analysis of effectieness relative to other therapies. *Ann Intern Med, 138*, 871-876.

Aure OF, Nilsen JH, & Vasselien O. (2003). Manual therapy and exercise therapy in patients with chronic low-back pain: a randomized, controlled trial with 1-year follow-up. *Spine, 28*, 525-531.

Axen I, Jones JJ, Rosenbaum A, Lovegren PW, Halasz L, Larsen K, et al. (2005). The Nordic Back Pain Subpopulation Program; validation and improvement of a predictive model for treatment outcomes in patients with low back pain receiveing chiropractic treatment *J Manipulative Physiol Ther, 28*, 381-385.

Balas AE, Kretschmer RAC, Gnann W, West DA, Boren SA, Centor RM, et al. (1998). Interpreting cost analyses of clinical interventions. *JAMA, 279*(1), 54-57.

Beffa R, & Matthews R. (2004). Does the adjustment cavitate the targeted joint? An investigation into the location of cavitation sounds. *J Manipulative Physiol Ther, 27*, e2.

Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, Regtop W, van der Heijden GJ, et al. (1995). Efficacy of traction for non-specific low back pain: a randomised clinical trial. *Lancet, 346*, 1596-1600.

Beurskens AJ, de vet HC, Koke AJ, Lindeman E, Regtop W, van der Heijden GJ, et al. (1997). Efficacy of traction for nonspecific low back pain. 12-week and 6-month results of a randomised clinical trial. *Spine, 22*, 2756-2762.

Binkley J, Finch E, Hall J, Black T, & Gowland C. (1993). Diagnostic classification of patients with low back pain: report on a survey of physical therapy experts. *Phys Ther, 73*, 138-150.

Boden SD, McCowin PR, Davis DO, Dina TS, Mark AS, & Wiesel S. (1990). Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am, 72*, 1178-1184.

Bodenheimer T. (2005a). High and Rising Health Care Costs. Part 1: Seeking and Explanation. *Ann Intern Med, 142*, 847-854.

Bodenheimer T. (2005b). High and Rising Health Care Costs. Part 2: Technologic Innovation. *Ann Intern Med, 142*, 932-937.

Bodenheimer T. (2005c). High and Rising Health Care Costs. Part 3: The Role of Providers. *Ann Intern Med, 142*, 996-1002.

Borkan JM, & Cherkin DC. (1996). An agenda for primary care research on low back pain. *Spine, 21*, 2280 - 2884.

Borkan JM, Koes B, Reis S, & Cherkin DC. (1998). A report from the Second International Forum for Primary Care Research on Low Back Pain. Reexamining priorities. *Spine, 23*, 1992 - 1996.

Bouter LM, van Tulder MW, & Koes BW. (1998). Methodologic issues in low back pain research in primary care. *Spine, 23*, 2014 - 2020.

Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, & Erhard RE. (2006). Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine, 31*, 623-631.

Bridges PH, Bierema LL, & Valentine T. (2007). The propensity to adopt evidence-based practice among physical therapists. *BMC Health Serv Res, 7*(1).

Briggs AH, Goeree R, Blackhouse G, & O'Brien BJ. (2002). Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Medical Decision Making, 22*, 290-308.

Bronfort G, Haas M, Evans RL, & Bouter LM. (2004). Efficacy of spinal manipulation and mobilization for low back and neck pain: a systematic review and best evidence synthesis. *Spine 4*, 335-356.

Browder DA, Childs JD, Cleland JA, & Fritz JM. (2007). Effectiveness of an extension oriented treatment approach in a subgroup of patients with low back pain: a randomized clinical trial. *Phys Ther, 87*(12), 1608-1618.

Cairns MC, Foster NE, & Wright C. (2006). Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain. *Spine, 31*, E670-681.

Campbell EG, Gruen RL, Mountford J, Miller LG, Cleary PD, & Blumenthal D. (2007). A National Survey of Physician-Indstry Relationships. *The New England Journal of Medicine, 356*(17), 1742-1750.

Cassidy JD, Cote P, Carroll LJ, & Kristman V. (2005). Incidence and course of low back pain episodes in the general population. *Spine, 30*(24), 2817 - 2823.

Cherkin DC, Deyo RA, Battie M, Street J, & Barlow WA. (1998). A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. *N Engl J Med, 339*, 1021-1029.

Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. (2004). A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med, 141*, 920-928.

Chiradejnant A, Maher CG, Latimer J, & Stepkovitch N. (2003). Efficacy of "therapist selected" versus "randomily selected" mobilisation techniques for the treatment of low back pain: a randomised controlled trial. *Aust J Physiother, 49*, 233-241.

Chou R, Qaseem A, Snow V, Casey D, Cross JT Jr, Shekelle P, et al. (2007). Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med, 147*(7), 478-491.

Clare HA, Adams R, & Maher CG. (2004). A systematic review of efficacy of Mckenzie therapy for spinal pain. *Aust J Physiother, 50*, 209-216.

Clarke HA, van Tulder M, Bloomberg S, de Vet H, van der Heijden G, & Bronfort G. (2006). Traction for low back pain with our without sciatica: an updated systematic review withing the framework of the Cochrane collaboration. *Spine, 31*, 1591-1599.

Claxton K, (1999).,The irrelevance of inference: a decision-making approach to the stochiastic evaluation of health care technologies, *J Health Econ, 18,* 341-364

Cleland JA, Fritz JM, Childs JD, & Kulig K. (2006). Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: study protocol of a randomized clinical trial. *BMC Musculoskelet Disord, 7*, 11.

Cleland JA, Fritz JM, Whitman JM, Childs JD, & Palmer JA. (2006). The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: a case series. *J Orthop Sports Phys Ther, 36*, 209-214.

Clinical Practice Guideline: Acute Low Back Problems in Adults. (December 1994). Retrieved. from.

Croft PR, Macfarlane GJ, Papageorgiou AC, Thomas E, & Silman AJ. (1998). Outcome of low back pain in general practice: a prospective study. *BMJ, 316*(7141), 1356 - 1359.

Crombez G, Vlaeyen JW, Heuts PH, & Lysens R. (1999). Pain-related fear is more disabling then pain itself: evidence on the role of pain-related fear in chronic back pain disability. *Pain, 80*, 329-339.

Danneels LA, Vanderstraeten GG, Cambier DC, Witvouw EE, Bourgois J, Dankaerts W, et al. (2001). Effects of three different training modalities on the cross sectional area of the lumbar multifidous muscle in patients with chronic low back pain. *Br J Sports Med, 35*, 186-191.

Davis K, Schoen C, Guterman S, Shih T, Schoenbaum SC, & Weinbaum I. (2007). *Slowing the Growth of U.S. Health Care Expenditures: What are the Options?* Washington DC: The Commonwealth Fund Commission on a High Performance Health System.

de Groot V, Beckerman H, Lankhorst GJ, Bouter LM, (2004). How to measure comorbidity: a critical review of available methods, *J. Clin. Epidemiol, 56,* 221-229

Delitto A. (2005). Research in low back pain: time to stop seeking the elusive "magic bullet." *Phys Ther, 85*, 206 - 208.

Delitto A, Cibulka MT, Erhard RE, Bowling RW, & Tenhula JA. (1993). Evidence for Use of an Extension-Mobilization Category in Acute Low Back Pain Syndrome: A Prescriptive Validation Pilot Study. *Physical Therapy, 73*(4), 216-222.

Delitto A, Erhard RE, & Bowling RW. (1995). A treatment-based classification approach to low back pain syndrome: identifying and staging patients for conservative treatment. *Phys Ther, 75*, 470-485.

Detsky AS, & Naglie IG. (1990). A Clinician's Guide to Cost-Effectiveness Analysis. *Ann Intern Med, 113*(2), 147-154.

Dettori JR, Bullock SH, Sutlive TG, Franklin RJ, & Patience T. (1995). The effects of spinal flexion and extension exercises and their associated postures in patients with acute low back pain. *Spine, 20*, 2303-2312.

Deyo RA, Mirza SK, & Martin BI. (2006). Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. *Spine, 31*(23), 2724-2727.

Deyo RA, & Phillips WR. (1996). Low back pain: a primary care challenge. *Spine, 21*, 2826-2832.

Di Fabio RP, & Boissonnault W. (1998). Physical therapy and health-related outcomes for patients with common orthopaedic diagnoses. *J Orthop Sports Phys Ther, 27*, 219-230.

Doubilet P, Weinstein MC, & McNeil BJ. (1986). Use and Misuse of the Term "Cost Effective" in Medicine. *The New England Journal of Medicine, 314*(4), 253-256.

Drummond MF, Brandt A, Luce B, & Rovira J. (1993). Standardising methodologies for economic evaluation in health care. *International Journal of Technology Assessment in Health Care, 9*, 26-36.

Drummond MF, & Jefferson TO. (1996). Guidelines for Authors & Peer Reviewers of Economic Submissions to the BMJ. *BMJ, 313*, 275-283.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, & Stoddart GL. (2005). *Methods for the Economic Evaluation of Health Care Programmes* (3rd ed.). New York, NY: Oxford University Press, Inc.

Dupuis PR, Yong-Hing K, Cassidy JD, & Kirkaldy-Willis WH. (1985). Radiologic diagnosis of degenerative lumbar spinal instability. *Spine, 10*, 262-276.

Eisenberg JM. (1989). A Guide to the Economic Analysis of Clinical Practices. *Clinical Economics, 262*(20), 2879-2886.

Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, & Pinder EL. (2003a). The Impications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care. *Annals of Internal Medicine, 138*(4), 288-296.

Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, & Pinder EL. (2003b). The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care. *Annals of Internal Medicine, 138*(4), 273-287.

Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. (2002). A Clinical Prediction Rule for Classifying Patients with Low Back Pain who Demonstrate Short-Term Improvement with Spinal Manipulation. *Spine, 27*(24), 2835-2843.

Frank JW, Brooker AS, DeMaio SE, & et al. (1996). Disability resulting from occupational low back pain . Part II: What do we know about secondary prevention? A review of the scientific evidence on prevention after disability begins. *Spine, 21*, 2918-2929.

Fritz JM, Childs JD, & Flynn TW. (2005). Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *BMC Fam Pract, 6*, 29.

Fritz JM, Cleland JA, & Brennan GP. (2007). Does adherence to the guideline recommendation for active treatments improve the quality of care for patients with acute low back pain delivered by physical therapists. *Medical Care, 45*(10), 973 - 980.

Fritz JM, Cleland JA, & Childs JD. (2007). Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *J Orthop Sports Phys Ther, 37*(6), 290-302.

Fritz JM, Cleland JA, Speckman M, Brennan GP, & Hunter SJ. (2008). Physical therapy for acute low back pain: Association with subsequent healthcare costs. *Spine, 33*(16), 1800-1805.

Fritz JM, Delitto A, & Erhard RE. (2003). Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine, 28*, 1363-1371.

Fritz JM, Delitto A, Welch WC, & Erhard RE. (1998). Lumbar spinal stenosis: a review of current concepts in evaluation, management, and outcome measurements. *Arch Phys Med Rehabil, 79*, 700-708.

Fritz JM, Erhard RE, & Hagen BF. (1998). Segmental instability of the lumbar spine. *Phys Ther, 78*, 889-896.

Fritz JM, & George S. (2000). The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. *Spine, 25*, 106-114.

Fritz JM, & George SZ. (2002). Identifying psychosocial variables in patients with acute work-related low back pain: the importance of fear-avoidance beliefs. *Phys Ther, 82*, 973-983.

Fritz JM, George SZ, & Delitto A. (2001). The role of fear-avoidance beliefs in acute low back pain: relationships with current and future disability and work status. *Pain, 94*, 7-15.

Fritz JM, & Irrgang JJ. (2001). A comparison of a Modified Oswestry disability questionaire and the Quebec back pain disability scale. *Phys Ther, 81*, 776-788.

Fritz JM, Lindsay W, Matheson JW, Brennan GP, Hunter SJ, Moffit SD, et al. (2007). Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. *Spine, 32*, E793-800.

Fritz JM, Piva SR, & Childs JD. (2005). Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. *Eur Spine J, 14*, 743-750.

Fryback DG, Dasbach EJ, Klein R, Klein BE, Dorn N, Petersen K, et al. (1993). The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. *Medical Decision Making, 13*(2), 89-102.

Frymoyer JW, & Selby DK. (1985). Segmental instability: Rational for treatment *Spine, 10*, 280-286.

Fullen BM, Maher T, Bury G, Tynan A, Daly LE, & Hurley DA. (2007). Adherence of Irish general practitioners to European guidelines for acute low back pain: a prospective pilot study. *Eur J Pain, 11*(6), 614-623.

Gandjour A, Jefferson T, Demicheli V, & Vale L. (2002). Educating authors and reviewers of economic evaluations of health care. *JAMA, 288*, 959.

George SM, Fritz JM, & Childs JD. (2008). Investigation of elevated fear-avoidance beliefs for patients with low back pain: A secondary analysis involving patients enrolled in physical therapy clinical trials. *J Orthop Sports Phys Ther, 38*, 50-57.

George SZ, Bialosky JE, & Fritz JM. (2004). Physical therapist management of a patient with acute low back pain and elevated fear-avoidance beliefs. *Phys Ther, 84*, 538-549.

George SZ, & Delitto A. (2005). Clinical examination variables discriminate among treatment-based classification groups: A study of construct validity in patients with acute low back pain. *Phys Ther, 85*(4), 306-314.

George SZ, Fritz JM, Bialosky JE, & Donald DA. (2003). The effect of a fear-avoidance based physical therapy intervention for patients with acute low back pain: results of a randomized clinical trial. *Spine, 28*, 2551-2560.

George SZ, Fritz JM, & Erhard RE. (2001). A comparison of fear-avoidance beliefs in patients with lumbar spine and cervical spine pain. *Spine, 26*, 2139-2145.

Gillan MG, Ross JC, McLean IP, & Porter RW. (1998). The natural history of trunk list, its associated disability and the influence of McKenzie management. *Eur Spine J, 7*, 480-483.

Godfrey CM, Morgan PP, & Schatzker J. (1984). A randomized trial of manipulation for low-back pain in a medical setting. *Spine, 9*, 301-304.

Gold MR, Franks P, McCoy KI, & Fryback DG. (1998). Toward consistency in cost-utility analyses: using national measures to create condition-specific values. *Medical Care, 36*(6), 778-792.

Gold MR, Siegel JE, Russell LB, & Weinstein MC. (1996). *Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press, Inc.

Goldby LJ, Moore AP, Doust J, & Trew ME. (2006). A randomized controlled trial investigating the effeciency of musculoskeletal physiotherapy on chronic low back disorder. *Spine, 31*, 1083-1093.

Gonzalez UV, Palacio EL, & Lopez MJ. (2003). Routine primary care management of acute low back pain: adherence to clinical guidelines. *Eur Spine J, 12*(6), 589 - 594.

Greenhalgh T, (1997)., How to read a paper, statistics for ht e non-statistician: different types of data need different statistical tests. *BMJ, 315,* 364-366

Greenman PE. (1991). *Osteopathic manipulation of the lumbar spine and pelvis*. Baltimore, MD: Williams & Wilkins.

Groll DL, To T, Bombardier C, & Wright JG. (2005). The development of a comorbidity index with physical function as the outcome. *Journal of Clinical Epidemiology, 58*, 595-602.

Grotle M, Vollestad NK, & Brox JI. (2006). Clinical course and impact of fear-avoidance beliefs in low back pain: prospective cohort study of acute and chronic low back pain:II. *Spine, 31*, 1038-1046.

Haas M, Groupp E, Panzer D, Partna L, Lumsden S, & Aickin M. (2003). Efficacy of cervical endplay assessment as in indicator for spinal manipulation. *Spine, 28*, 1091-1096.

Hart DL, Wang YC, Stratford PW, & Mioduski JE. (2008a). Computerized adaptive test for patients with hip impairments produced valid and sensitive measures of function. *Arch Phys Med Rehabil, 89*, 2129-21339.

Hart DL, Wang YC, Stratford PW, & Mioduski JE. (2008b). Computerized adaptive test for patients with knee impairments produced valid and sensitive measures of function. *Journal of Clinical Epidemiology, 61*(11), 1113-1124.

Hart DL, Ying-Chih W, Stratford PW, & Mioduski JE. (2008). Computerized adaptive test for patients with foot or ankle impairments produced valid and responsive measures of function. *Qual Life Res, 17*, 1081-1091.

Harte AA, Baxter GD, & Gracey JH. (2003). The efficacy of traction for back pain: a systematic review of randomized controlled trials. *Arch Phys Med Rehabil, 84*, 1542-1553.

Hayden JA, van Tulder MW, & Tomlinson G. (2005). Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. *Ann Intern Med, 142*, 776 - 785.

Hayes MA, Howard TC, Gruel CR, & Kopta JA. (1989). Roentgenographic evaluation of the lumbar spine flexion-extension in asymptomatic individuals. *Spine, 14*, 327-331.

Hazard RG, Haugh LD, Reid S, Preble JB, & MacDonald L. (1996). Early prediction of chronic disability after occupational low back pain injury. *Spine, 21*, 945-951.

Heyne P, Boettke P, & Prychitko D. (2003). *The Economic Way of Thinking: 10th Edition* (10th ed.). Upper Saddle River, NJ: Prentice Hall.

Hicks GE, Fritz JM, Delitto A, & McGill SM. (2005). Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil, 86*, 1753-1762.

Hicks GE, Fritz JM, Delitto A, & Mishock J. (2003). Interater reliability of clincial examination measures for identification of lumbar segmental instability *Arch Phys Med Rehabil, 84*, 1858-1864.

Hides JA, Jull GA, & Richardson CA. (2001). Long-term effects of specific stabilizing exercises for first-episode of low back pain. *Spine, 26*, E243-248.

Hodges PW. (2001). Changes in motor planning of feedforward postural responses of the trunk muscles in low back pain. *Exp Brain Res, 141*, 261-266.

Hodges PW, Moseley GL, Gabrielsson A, & Gandevia SC. (2003). Experimental muscle pain changes feedforward postural responses of the trunk muscles in low back pain. *Exp Brain Res, 141*, 261-266.

Hodges PW, & Richardson CA. (1996). Ineffecient muscular stabilization of the lumbar spine associated with low back pain. A motor control evaluation of the transversus abdominis. *Spine, 21*, 2640-2650.

Hodges PW, & Richardson CA. (1998). Delayed postural contraction of transversus abdominus in low back pain associated with movement of the lower limb. *J Spinal Disord, 11*, 46-56.

Homik JE, & Suarez-Almazor M. (2004). An Economic Approach to Health Care. *Best Practice & Research Clinical Rheumatology, 18*(2), 203-218.

Huang GD, Feuerstein M, & Arroyo F. (2001). Back and upper extremity disorders among enlisted U.S. Marines: burden and individual risk factors. *Mil Med, 166*(11), 1007 - 1017.

Jefferson T, & Demicheli V. (2002). Quality of Systematic Reviews of Economic Evaluations in Health Care. *JAMA, 287*(21), 2809-2812.

Jefferson T, Smith R, Lee Y, Drummond M, Pratt M, & Gale R. (1998). Evaluating the BMJ guidelines for economic submissions: Prospective audit of economic submissions to BMJ and The Lancet. *JAMA, 280*, 276-277.

Jensen MP, Turner JA, & Romano JM. (1994). What is the maximum number of levels needed in pain intensity measurement. *Pain, 58*, 387-392.

Jette AM, & Delitto A. (1997). Physical therapy treatment choices for musculoskeletal impairments. *Phys Ther, 77*, 145 - 154.

Jette AM, Smith KS, Haley SM, & Davis KD. (1994). Physical therapy episodes of care for patients with low back pain. *Phys Ther, 74*(2), 101-116.

Jette DU, Bacon K, Batty C, Carlson A, Ferland A, Hemingway RD, Hill JC, Oglivie L, Volk D., (2003)., Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists, *Phys Ther, 83,* 786-805

Jette DU, & Jette AM. (1996). Physical therapy and health outcomes in patients with spinal impairments. *Phys Ther, 76*(9), 930-945.

Kent P, & Keating J. (2004). Do primary-care clinicians think that nonspecific low back pain is one condition? *Spine, 29*, 1022 - 1031.

Kent P, Marks D, Pearson W, & Keating J. (2005). Does clinician treatment choice improve the outcomes of manual therapy for nonspecific low back pain? A metaanalysis. *J Manipulative Physiol Ther, 28*, 312-322.

Keselman HJ, Cribbie RA, Wilcox RR., (2007)., Pairwise multiple comparison tests when data are non-normal, *Educational and Psychological Measurement, 62*, 420-434

Kilpikoski S, Airaksinen O, Kankaanpaa M, Leminen P, Videman T, & Alen M. (2002). Interexaminer reliability of low back pain assessment using the Mckenzie method. *Spine, 27*, E207-E214.

Koes BW, van Tulder MW, Ostelo R, Burton KA, & Waddell G. (2001). Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine, 26*(22), 2504-2514.

Koes BW, van Tulder MW, & Thomas S. (2006). Diagnosis and treatment of low back pain. *BMJ, 332*, 1430 - 1434.

Kominski GF, Heslin KC, Morgenstern H, Hurwitz EL, & Harber PI. (2005). Economic evaluation of four treatments for low-back pain: results from a randomized controlled trial. *Medical Care, 43*(5), 425-427.

Koumantakis GA, Watson PJ, & Oldham JA. (2005). Trunk muscle stabilization training plus general exercise versus general exercise only: randomized controlled trial of patients with recurrent low back pain. *Phys Ther, 85*, 209-225.

Landry MD, & Sibbald WJ. (2002). Changing physician behavior: A review of patient saftey in critical care medicine. *Journal of Critical Care, 17*(2), 138-145.

Leeuw M, Goossens ME, Linton SJ, Crombez G, Boersma K, & Vlaeyen JW. (2007). The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. *J Behav Med, 30*, 77-94.

Lethem J, Slade PD, Troup JD, & Bentley G. (1983). Outline of a Fear-Avoidance Model of exaggerated pain perception. *Behav Res Ther, 21*, 401-408.

Li LC, & Bombardier C. (2001). Physical therapy management of low back pain: an exploratory survey of therapist approaches. *Phys Ther, 81*, 1018 - 1028.

Lincoln AE, Smith GS, Amoroso PJ, & Bell NS. (2002). The natural history and risk factors of musculoskeletal conditions resulting in disability among U.S. Army personnel. *Work, 18*(2), 99 - 113.

Lingard EA, Berven S, Katz JN, & Kinemax Outcomes Group. (2000). Magagement and care of patients undergoing total knee arthroplasty: variations across different health care settings. *Arthritis Care Res, 13*(3), 129 - 136.

Linton SJ. (2000). A review of psychological risk factors in back and neck pain. *Spine, 25*, 1148-1156.

Linton SJ, & Hallden K. (1998). Can we screen for problematic back pain? A screening questionairre for predicting outcome in acute and subacute back pain. *Clin J Pain, 14*, 209-215.

Long A, Donelson R, & Fung T. (2004). Does it matter which exercise? A randomized controlled trial of exercise for low back pain. *Spine, 29*, 2593-2602.

Luijsterburg PA, Lamers LM, Verhagen AP, Ostelo RWJG, van den Hoogen HJMM, Peul WC, et al. (2007). Cost-effectiveness of physical therapy and general practitioner care for sciatica. *Spine, 32*(18), 1942-1948.

Luo X, Pietrobin R, & SX, S. (2004). Estimates and patterns of direct health care expenditures among individuals with back pain in the United States. *Spine, 29*, 79 - 86.

Machado LA, de Souza MS, Ferreira PH, & Ferreira ML. (2006). The McKenzie method for low back pain: a systematic review of the literature with a metaanalysis approach. *Spine, 31*, E254-262.

Maitland G, Hengeveld E, Banks K, & English K. (2000). *Maitland's verterbral manipulation*. Oxford, UK: Butterworth-Heinemann.

Malmivaara A, Hakkinen U, Aro T, Heinrichs ML, Koskenniemi L, Kuosma E, et al. (1995). The treatment of acute low back pain --bed rest, exercises, or ordinary activity? *N Engl J Med, 332*, 351-355.

Malter AD, Larson EB, Urban N, & Deyo RA. (1996). Cost-effectiveness of lumbar discectomy for the treatment of herniated intervertebral disc. *Spine 21*(9), 1048-1054.

Martin BI, Deyo RA, Mirza SK, Turner JA, Comstock MS, Hollingsworth W, et al. (2008). Expenditures and health status among adults with back and neck problems. *JAMA, 299*(6), 656-664.

McCarthy M. (2003). U.S. Health-Care System Faces Cost and Insurance Crisis. *THE LANCET, 362*, 375.

McGill SM. (2001). Low back stability: from formal description to issues for performance and rehabilitation. *Exerc Sport Sci Rev, 29*, 26-31.

McGill SM. (2003). Coordination of muscle activity to assure stability of the lumbar spine. *J Electromyogr Kinesiol, 13*, 353-359.

McGill SM, & Cholewicki J. (2001). Biomechanical basis for stability: an explanation to enhance clinical utility. *J Orthop Sports Phys Ther, 31*, 96-100.

McGuire SM, Phillips KT, Weinstein JN, Factors that effect surgical rates in Iowa, (1994)., *Spine, 19,* 2038 -2040

McKenzie RA. (1989). *The Lumbar Spine: Mechanical Diagnosis and Therapy*. Waikanae, New Zealand: Spinal Publications LTD.

Mikhail C, Korner-Bitensky N, Rossignol M, & Dumas JP. (2005). Physical therapists' use of interventions with high evidence of effectiveness in the management of a hypothetical typical patient with acute low back pain. *Phys Ther, 85*, 1151 - 1167.

Nachemson A. (1985). Lumbar spine instability. A critical update and symposium summary. *Spine, 10*, 290-291.

Neumann PJ, Stone PW, Chapman RH, Sandberg EA, & Bell CM. (2000). The quality of reporting in published cost-utility analyses, 1976-1997. *Ann Intern Med, 132*(12), 964-972.

Niemisto L, Rissanen P, Sarna S, Lahtinen-Suopanki T, Lindgren KA, & Hurri H. (2005). Cost-effectiveness of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain: a prospective randomized trial with 2-year follow-up. *Spine, 30*(10), 1109-1115.

O'Sullivan PB, Phyty GD, Twomey LT, & Allison GT. (1997). Evaluation of specific stabilizing exercise in the treatment of chronic low back pain with the radiologic diagnosis of spondylolysis or spondylolisthesis. *Spine, 27*, 2959-2967.

Patel AT, & Ogle AA. (2000). Diagnosis and management of acute low back pain. *Am Fam Physician, 61*(6), 1779-1790.

Pearlman AS, Ryan T, Picard MH, & Douglas PS. (2007). Evolving Trends in the Use of Echocardiography: A Study of Medicare Beneficiaries. *Journal of the American College of Cardiology, 49*(23), 2283-2291.

Petersen T, Kryger P, Ekdahl C, Olsen S, & Jacobsen S. (2002). The effect of Mckenzie therapy as compared with that of intensive strengthening training for the treatment of patients with sub-acute or chronic low back pain: A randomized controlled trial. *Spine, 81*, 1702-1709.

Philadelphia Panel, Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for low back pain. (2001). *Phys Ther, 81*, 1641 - 1674.

Pincus T, Burton AK, Vogel S, & Field AP. (2002). A systematic review of psychological factors as predictors of chronic/disability in prospective cohorts of low back pain. *Spine, 27*, E109-120.

Pincus T, Vlaeyen JW, Kendall NA, Von Korff MR, Kalauokalani DA, & Reis S. (2002). Cognitive-behavioral therapy and psychosocial factors in low back pain; directions for the future. *Spine, 27*, E133-138.

Richardson CA. (1995). Muscle control-pain control. What exercises would you precribe? *Man Ther, 1*, 2-10.

Riddle DL. (1998). Classification and low back pain: a review of the literature and critical analysis of selected systems. *Phys Ther, 78*, 708-737.

Rivero AO, Gray A, Frost H, Lamb SE, & Stewart SB. (2006). Cost-utility analysis of physiotherapy treatment compared with physiotherapy advice in low back pain. *Spine, 31*(12), 1381-1387.

Rose SJ. (1989). Physical therapy diagnosis: role and function. *Phys Ther, 69*, 535-537.

Rubin DI. (2007). Epidemiology and risk factors for spine pain. *Neurol Clin, 25*(2), 353-371.

Sackett DL, Haynes RB, Guyatt GH, & Tugwell P. (1991). *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston, MA: Little, Brown & Company.

Schers H, Wensing M, Huijsmans Z, van Tulder M, & Grol R. (2001). Implementation barriers for general practice guidelines on low back pain a qualitative study. *Spine, 26*(15), E348 - 353.

Skargren EI, Carlsson PG, & Oberg BE. (1998). One-year follow-up comparison of the cost and effectiveness of chiropractic and physiotherapy as primary management for back pain. Subgroup analysis, recurrence, and additional health care utilization. . *Spine, 23*, 1875-1883.

Smith MT, Carmody TP, & Smith MS. (2000). Quality of well-being scale and chronic low back pain. *Journal of Clinical Psychology in Medical Settings, 7*(3), 175-184.

Soegaard R, Bunger CE, Christiansen T, Hoy K, Eiskaer SP, & Christiansen FB. (2007). Circumferential fusion is dominant over posterolateral fusion in a long-term perspective: cost-utility evaluation of a randomized controlled trial in severe, chronic low back pain. *Spine, 32*(22), 2405-2414.

Staerkle R, Mannion AF, Elfering A, & et al. (2004). Longitudinal validation of the fear-avoidance beliefs questionairre (FABQ) in a Swiss-German sample of low back pain patients. *Eur Spine J, 13*, 332-340.

Strine TW, & Hootman JM. (2007). U.S. National prevalence and correlates of low back and neck pain among adults. *Arthritis Rheum, 57*(4), 656-665.

Stuge B, Bragelien Veierod M, Laerum E, & Vollestad N. The Efficacy of a Treatment Program Focusing on Specific Stabilizing Exercises for Pelvic Girdle Pain After Pregnancy: A Two-Year Follow-up of a Randomized Clinical Trial. *Spine, 29*(10), E197-E203.

Stuge B, Laerum E, Kirkesola G, & Vollestad N. (2004). The efficacy of a treatment program focusing on specific stabilizing exercises for the pelvic girdle pain after pregnancy: a randomized controlled trial. *Spine, 29*, 351-359.

Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, & Vincent D. (2000). Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology, 39*, 783-790.

Taylor RJ, & Taylor RS. (2005). Spinal cord stimulation for failed back surgery syndrome: a decision-analytic model and cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care, 21*(3), 351-358.

The Commonwealth Fund. (2006). Health Care Costs in US Markedly Exceed that of Other Countries, and Increasing More Rapidly: Commonwealth Fund National Scorecard on U.S. Health System Performance.

Torstensen TA, Ljunggren AE, Meen HD, Odland E, Mowinckel P, & Geijerstam S. (1998). Efficiency and costs of medical exercise therapy, conventional physiotherapy, and self-exercise in patients with chronic low back pain. A pragmatic, randomized, single-blinded, controlled trial with 1-year follow-up. *Spine, 23*(23), 2616-2624.

Turner JA, Franklin G, Fulton-Kehoe D, & et al. (2006). Worker recovery expectations and fear-avoidance predict work disability in a population-based workers' compensation back pain sample. *Spine, 31*, 682-689.

Udvarhelyi S, Colditz GA, Rai A, & Epstein AM. (1992). Cost-effectiveness and cost-benefit analyses in the medical literature: Are the methods being used correctly? *Ann Intern Med, 116*, 238-244.

United Kingdom Back Pain Exercise and Manipulation (UK BEAM) Randomised Trial: Cost Effectiveness of Physical Treatments for Back Pain in Primary Care. (2004). *BMJ, doi:10.1136/bmj.38282.607859.AE*.

Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, Flemming DA, McDonnell MK, et al. (1998). Reliability of physical examination items used for classification of patients with low back pain. *Phys Ther, 78*, 979-988.

van Tulder MW, Koes B, & Malmivaara A. (2006). Outcome of non-invasive treatment modalities on back pain: an evidence-based review. *Eur Spine J, 15*, Suppl 1:S64-81.

Vlaeyen JWS, Kole-Snijders AM, Boeren RG, & van Eek H. (1995). Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain 62*, 363-372.

Vlaeyen JWS, Kole-Snijders AMJ, Rotteveel AM, Ruesink R, & Heuts PHTG. (1995). The role of fear of movement/(re)injury in pain disability. *J Occup Rehab, 5*, 235-252.

Vogt MT, Kwoh CK, Cope DK, Osial TA, Culbya M, Starz TW, (2005). Analgesic usage for low back pain: impact on health care costs and service use, *Spine, 30,* 1075-1081

Waddell G, Newton M, Henderson I, Somerville D, & Main CJ. (1993). A Fear-Avoidance Beliefs Questionairre (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain, 52*, 157-168.

Wasiak R, Kim J, & Pransky G. (2006). Work disability and costs caused by recurrence of low back pain: longer and more costly than in first episodes. *Spine, 31*(2), 219 - 225.

Weiner DK, Kim YS, Bonino P, & Wang T. (2006). Low back pain in older adults: are we utilizing healthcare resources wisely? *Pain Medicine, 7*(2), 143 - 150.

Weinstein MC, & Stason WB. (1972). Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. *The New England Journal of Medicine, 296*(13), 716-721.

Werners R, Pynsent PB, & Bulstrode CJ. (1999). Randomized trial comparing interferntial therapy with motorized lumbar traction and massage in the management of low back pain in a primary care setting. *Spine, 24*, 1579-1584.

Whitehurst DG, Lewis M, Yao GL, Bryan S, Raftery JP, Mullis R, et al. (2007). A brief pain management program compared with physical therapy for low back pain: results from an economic analysis alongside a randomized controlled trial. *Arthritis Rheum, 57*(3), 466-473.

Whitman JM, Flynn TW, Childs JD, Wainner RS, Gill HE, Ryder MG, et al. (2006). A comparison between two physical therapy treatment programs for patients with lumbar spinal stenosis: a randomized controlled trial. *Spine, 31*, 2541-2549.

Williams A. (1974). The Cost Benefit Approach. *Br Med Bull, 30*(3), 252-256.