REVISITING THE COST-EFFECTIVENESS OF RECOMBINANT HUMAN BONE MORPHOGENETIC PROTEIN-2 VERSUS ILIAC CREST BONE GRAFT IN LUMBAR SPINAL ARTHRODESIS

A Thesis

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Master's of Science

by

Richard M. Silverman

August 2015

ABSTRACT

Cost-Effectiveness analysis (CEA) studies used to compare the cost-effectiveness of two different surgical approaches, recombinant human bone morphogenetic protein-2 (rhBMP-2) and iliac crest bone graft (ICBG), have typically used up to two years of follow-up data, thus excluding any additional costs and interventions accumulated after that time period. To address this limitation, a future CEA is planned to compare the value of rhBMP-2 versus ICBG in lumbar fusion using five years follow-up from administrative data. This thesis had 2 objectives. The first objective was to perform a literature review that identified studies that provide or supplement the data necessary to conduct the future CEA. The second objective was to use the results from the literature review to perform a preliminary investigation of the planned CEA model. A systematic literature search of published literature through May 2015 was conducted using PubMed/MEDLINE, the Cochrane Collection Library, and the Tuft's CEA Registry. Using a Markov model, a preliminary CEA compared rhBMP-2 to ICBG with regard to incremental cost-effectiveness ratios (ICERs) from both a payer and societal perspective. Transition probabilities derived from second surgery rates were extracted from two studies: one with differing rates between treatment groups, and one with similar rates between groups. Four subanalyses were performed in order to combine each set of second surgery rates with each cost perspective. Two threshold sensitivity analyses were conducted in order to test the effects of second surgery rates and indirect costs on the results of the analysis. Forty publications were identified that could contribute to the future study, and data was extracted from 11 of those 40 articles to perform the preliminary CEA. Knowledge

gaps included lack of long-term follow-up of fusion rates and second surgery rates, and lack of quality reporting of utility and work status. Furthermore, there was disagreement between prospective and retrospective studies in fusion rates and second surgery rates. In the two subanalyses performed from the payer perspective, ICBG was more cost-effective than rhBMP-2, and the ICER values were $251,026/QALY and $849,893/QALY. In the two subanalyses performed from the societal perspective, rhBMP-2 was both less costly and more effective than ICBG, and the ICER was not needed to compare the interventions. From the payer perspective, rhBMP-2 became more cost-effective than ICBG when the difference in second surgery rates was greater than 13% after five years, and from the societal perspective, ICBG became more cost-effective than rhBMP-2 when the difference in indirect costs was less than $4,000 after five years. In conclusion, this thesis identified knowledge gaps in the literature and selected 40 publications that would be useful to a future CEA comparing the cost-effectiveness of rhBMP-2 and ICBG. The preliminary CEA showed that rhBMP-2 was less cost-effective than ICBG from a payer perspective, but more cost-effective from a societal perspective.

BIOGRAPHICAL SKETCH

Richard Martin Silverman was born on March 25, 1991, in St. Louis, Missouri. He graduated Fox Chapel Area High School in Pittsburgh, PA in 2009, and received his Bachelor of Science degree from Cornell University in Ithaca, NY in 2013. While completing his degree at Cornell, Richard majored in biological sciences and minored in applied economics and management. After graduation, he worked as a research technician in the Perlmutter Lab at the Children's Hospital of Pittsburgh. His research resulted in two publications. The first, entitled *The Aggregation-Prone Intracellular Serpin SRP-2 Fails to Transit the ER in Caenorhabditis elegans*, was published in *Genetics* May 2014, while the second article, entitled *Deficient and null variants of SERPINA1 are proteotoxic in a Caenorhabditis elegans model of α1-antitrypsin deficiency*, is currently under revision.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Hassan Ghomrawi for the continuous support of my research and Master's thesis, for his patience, motivation, and valuable knowledge. His guidance helped me in all the aspects of my research and writing of this thesis.

Besides my advisor, I would like to thank Dr. Zachary Grinspan for his insightful comments as part of my thesis committee. I would also like to express my appreciation to Dr. Jessica Ancker and Dr. Stephen Johnson for vastly expanding my knowledge and making this Master's program a success.

Last but not least, my sincere thanks also goes to Dr. Alex Hughes, Dr. Peter Derman, Dr. Jayme Burket, and Dr. Lukas Lampe, who provided an opportunity to join their research team, and increased my understanding of my research topic.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

This thesis conducted a systematic literature review and performed a preliminary cost effectiveness analysis to compare two different surgical approaches, bone morphogenetic protein and iliac crest bone graft, used in lumbar spinal fusion surgery. The review is comprised of four chapters: introduction, methods, results, and discussion & conclusion. This introductory chapter is divided into four sections. Section 1.1 introduces the interventions being compared, and section 1.2 provides an overview of the mechanism and importance of cost-effectiveness analyses. Section 1.3 summarizes the existing cost-effectiveness literature that compares the two treatment options and identify gaps in the literature that provide impetus for this thesis, while section 1.4 explains the study objectives and hypotheses.

1.1 Lumbar Fusion

Chronic lower back is reported in 19% of working US adults, and is known to be associated with abnormal motion at a vertebral segment (1). In cases where rehabilitation with physical therapy has failed or is not possible, surgical intervention is needed. Lumbar spinal fusion is a surgical procedure designed to decrease the motion at the painful vertebral segment thereby decreasing pain generated from the joint (2). Surgeons can perform this operation using an anterior, posterior, lateral, or combined approach. During the operation, the surgeon uses a bone graft that generates a biological response, called osteogenesis, that causes two vertebral elements to fuse (2).

Three types of bone grafts are currently used as a stimulus for osteogenesis: autografts, allografts, or bone graft substitutes. *Autografts* use bone collected from the patient; *allografts* use bone obtained from a cadaver (3); *bone graft substitutes* are products that either assist or replace the need for autograft or allograft in a spinal fusion (3). The gold standard graft option for fusion is the iliac crest bone autograft (ICBG). The graft is harvested from the patient’s pelvis or iliac crest, and requires an additional incision during anterior fusions (4). Despite its clinical success, ICBG has been hindered by graft-site pain and complication rates approaching 39% in some instances (5).

In an effort to decrease invasiveness and eliminate the potential complications of bone harvesting, researchers from Medtronic Inc. designed a bone graft substitute intended to replace ICBG. This product, named InFUSE™ Bone Graft/LT-CAGE™ Lumbar Tapered Fusion Device, was approved by the FDA in 2002 for anterior interbody lumbar fusions (ALIF) (6), that is, a type of fusion that places the graft in the area occupied by the intervertebral disc via anterior abdominal incision (2). The device contains recombinant human bone morphogenetic protein-2 (rhBMP-2), which serves as the graft substitute by stimulating the formation of new bone tissue (osteoinduction) (7).

The FDA approved the use of rhBMP-2 in ALIF based on the results of a randomized clinical trial performed by Burkus et al. between August 1998 and July 1999.(8) The study concluded that rhBMP-2 matched the efficacy of ICBG, and even surpassed the traditional technique in key parameters such as fusion rate (94.5% rhBMP-2 vs. 88.7% ICBG) and secondary surgery rate (7.0% rhBMP-2 vs. 10.3% ICBG) (8). Since its approval, the use of rhBMP-2 in lumbar fusion increased from 1% (1000 of 103,371 procedures) in 2002 to 43% (75,000 of 174,445 procedures) in 2010 on patients with degenerative diagnoses (9). In addition to the initial clinical trial, several other prospective studies have reported higher fusion rates in favor of rhBMP-2 in lumbar fusions (10–20). Furthermore, the majority of these trials also published lower secondary surgery rates for rhBMP-2 patients over 24 months (12,14,16–19). In contrast to this evidence, retrospective studies have shown no difference in fusion rates (21,22) and secondary operation rates (23,24). Moreover, it is known that rhBMP-2 incurs a higher surgical cost of at least $3,000 (25) and may provoke more nonoperative complications than originally reported (26). Since there is uncertainty regarding the use of rhBMP-2 versus ICBG, there is a need for research that quantifies this decision to select one intervention over another.

1.2 Cost-Effectiveness

In 1993, the US Public Health Service (PHS) commissioned the Panel on Cost-Effectiveness in Health and Medicine to address concerns relating to strained budgets in the healthcare sector (27). The panel defined cost-effectiveness analysis (CEA) as a method for evaluating health outcomes and resource costs of health interventions, and maintained that its central function was to show the relative value of alternative interventions for improving health (27). This relative value is represented by the incremental cost-effectiveness ratio (ICER), where $ICER = \frac{C_1 - C_0}{E_1 - E_0}$ (28). C_1 and E_1 are cost and effectiveness for the alternative group, while C_0 and E_0 are cost and effectiveness for the control group. The ultimate significance of CEA is to provide information that is critical to informing decisions about the allocation of health care resources (27).

The cost of care can be determined from a provider, payer, or societal perspective. The societal perspective is the most inclusive, incorporating both the direct costs of care and the indirect costs relating to productivity loss (29). Direct costs are directly attributed to patient care, such as hospital admissions, surgical fees, diagnostic testing, outpatient visits, medications, and physical therapy (29). Indirect costs are not directly related to patient care, and are typically estimated from the patient's loss in workforce productivity via missed time from work (30). The Panel on Cost-Effectiveness in Health and Medicine currently recommends that CEA be performed from the societal perspective (31).

Patient-reported outcome (PRO) measures are commonly used to approximate the effectiveness of health interventions (32). These measures are recorded using questionnaires, and are separated into three general categories: general quality of life, pain scale, and disease-specific outcome measures (32). PROs are then used to estimate health utility, which reflects a patient's preference for a particular health state or health-related outcome in a single index number that ranges from zero (death) to 1 (perfect health) (29). For CEA, effectiveness is measured in quality-adjusted life years (QALY). QALYs are calculated by multiplying the health utility by the number of years spent with that utility. Effectiveness is combined with cost in the ICER to generate an endpoint unit of cost per QALY gained. This value is used as the basis for comparison between interventions when the strategies are non-dominated. A dominant intervention is one that is both clinically superior and cost saving (33), thus an ICER is not needed to compare the two options.

Two types of analysis models are commonly used to calculate the ICER: decision trees and Markov models (34). Decision trees simulate situations in which a sequence of events usually occur once, whereas Markov models are recursive trees that model conditions with repeating events over time (35). An intervention is generally considered cost-effective when the cost per QALY gained is less than $100,000, but the determination of a specific dollar value threshold for appropriate treatment is controversial (36). Clinicians can vary the threshold ICER according to the type of medicine, the type of disease, and the decision-making context (37).

1.3 The Cost-Effectiveness of rhBMP-2 versus ICBG in Lumbar Fusion

Several previous analyses have applied CEAs to evaluate the decision to use rhBMP-2 over ICBG as a graft option during lumbar fusion (25,38–43). These analyses have acknowledged that rhBMP-2 is associated with higher upfront costs, but concurred that those costs are offset due to either increased utility associated with no donor-site complications, decreased long-term healthcare utilization costs, or decreased loss in productivity (25,38–43).

Table 1. Cost-effectiveness studies comparing rhBMP-2 and ICBG

Reference	Perspective, Clinical Data Source	Conclusion	Limitations
Polly et al. (25)	Provider, Burkus et al. (8)	The upfront price of bone morphogenetic protein is likely to be offset to a significant extent by reductions in the use of other medical resources.	Short time horizon, no indirect costs, no discounting, heavily based on expert opinion
Garrison et al. (38)	Societal, Burkus et al. (8)	There was limited evidence showing that BMP is associated with greater improvement in clinical outcomes. The use of BMP for spinal fusion is unlikely to be cost-effective.	Performed for the UK, short time horizon, type of model not described
Alt et al. (39)	Societal, Burkus et al. (44)	Use of rhBMP-2 in anterior lumbar fusion is a net cost-saving treatment from a societal perspective for Germany, France and UK.	Performed for foreign economies, short time horizon, potential conflict of interests
Carreon et al. (40)	Payer, Glassman et al. (18)	There were higher costs and lower improvements in utility seen in patients receiving ICBG compared with rhBMP-2/ACS in this study population.	Short time horizon, no indirect costs, only patients over 60 years old were considered
Ratko et al. (41) (AHRQ)	Payer, Burkus et al. (8)	The evidence for BMP consistently showed similar and possibly increased frequency of fusion and avoidance of bone graft harvest events.	Short time horizon, no indirect costs, unclear method for utility calculation
Virk et al. (43)	Payer, Multiple sources	RhBMP is the most	Focused on only one

	cost-effective graft option for L4–L5 fusion for degenerative spondylolisthesis largely due to the reduced rate of revision spine surgery.	cause of lumbar fusion, did specify on how model parameters were extracted from the literature

Prior CEA studies had limitations that weaken their conclusions concerning cost-effectiveness. These studies are summarized in Table 1. Out of the 6 CEAs comparing rhBMP-2 and ICBG, 4 used the same initial FDA-approved clinical trial data as the basis for their analysis (25,38,39,41). The primary limitation of using that clinical trial was the short follow-up of patients. Since the trial only follows patients for two years, costs or utilities generated after two years would not be included in the analysis. This issue may become significant since revisions of the index procedure do occur after two years (13). Two of these four studies were performed from a foreign economical perspective that may have different costs due to variations in payers and health systems (38,39). One of the four studies was executed by the Agency for Healthcare Research and Quality (AHRQ) in the US, which focused their analysis on the differences in utility between treatment groups as a result of donor site pain (41). The authors do not explain how they estimated the difference in utility values, nor do they acknowledge that their assumptions disagree with studies that report no difference in utility between treatment groups with similar outcomes (8,12,14,16,19,36,45).

The two remaining CEAs that were considered had limitations related to follow-up time, sample demographics, or data analysis (40,43). The first study had the same time horizon as the first four CEAs mentioned, and used a clinical trial that only included patients over sixty years old (40), which produces results that are not representative of the entire population. The second study had an adequate time horizon, and extracted data from several sources instead of a single clinical trial (43). However, those data sources had variable interventions, outcome measures, and costs that were combined in the analysis in a way that was unclear (43). None of the CEAs referenced incorporated indirect costs into their analysis.

1.4 Research Objectives & Hypotheses

The findings from this thesis will contribute to a larger study that will perform a CEA comparing the value of rhBMP-2 versus ICBG in lumbar fusion. The ensuing study will rely on data from both the literature and the New York Statewide Planning and Research Cooperative System (SPARCS). SPARCS is a comprehensive all payer data reporting system that collects patient level detail on patient characteristics, diagnoses and treatments, services, and charges for each hospital inpatient stay and outpatient visits (46). Data from SPARCS should overcome some of the limitations of the available literature.

The objectives for this thesis were to conduct literature review to support the ultimate CEA, and to pilot investigate the CEA model using only data from the literature. In order to support the larger project, the literature review will systematically search for, identify, and evaluate studies that provide or supplement the data necessary to perform the analysis. Since it is known that SPARCS will not provide all of the inputs needed to execute the CEA model, the findings from the literature review will help to supply the values for any missing model parameters, and will serve as a basis of comparison for data generated from SPARCS. The results from the review will also supply the data inputs for the following pilot analysis.

This preliminary base case CEA will test the hypotheses that based on current knowledge, rhBMP-2 is more cost-effective than ICBG. In regard to this premise, published data suggests that clinical outcomes in operations using either graft will have comparable direct costs and utility values (8,12,14,16,19,36,45). As a result, it is anticipated that a significant difference in the ICER would stem from predominantly two variables: rate of secondary intervention or indirect costs estimated through missed time from work. Sensitivity analysis will calculate the value of these variables that are necessary to change the decision if either intervention is found to be more cost-effective. The objectives of this sensitivity analysis are posed as research questions, and are listed with the main hypothesis in Table 2. Results from the CEA and subsequent analysis will be discussed in the context of these concepts.

Table 2. Study Hypothesis and Research Questions

Hypothesis	rhBMP-2 is more cost-effective than ICBG
Question 1	What difference in second surgery rates changes the decision of the CEA?
Question 2	What difference in indirect costs changes the decision of the CEA?

CHAPTER TWO

METHODS

Chapter two outlines the methods used for completing the literature review and preliminary cost-effectiveness analysis. Section 2.1 describes the structure of the literature search that detected studies needed for the larger CEA, while section 2.2 discusses the articles that influenced the design of the preliminary model, and how a simplified CEA version of the model was built for the purposes of this project. Section 2.3 covers the parameters needed to execute the simplified CEA model, how these parameters were identified and extracted from the literature, and how a sensitivity analysis was performed to test the impact of second surgery rates and indirect costs.

2.1 Literature Search

A systematic literature search of all published literature through May 2015 was performed using PubMed/MEDLINE, the Cochrane Collection Library, and the Tuft's CEA Registry. The search strategy was based on characteristics defined by Cooper in his taxonomy of literature reviews (47), and focused on research methods and outcomes that would help design and inform the cost-effectiveness analysis model. Keywords were selected systematically in order to build the model, and are listed in figure 1.

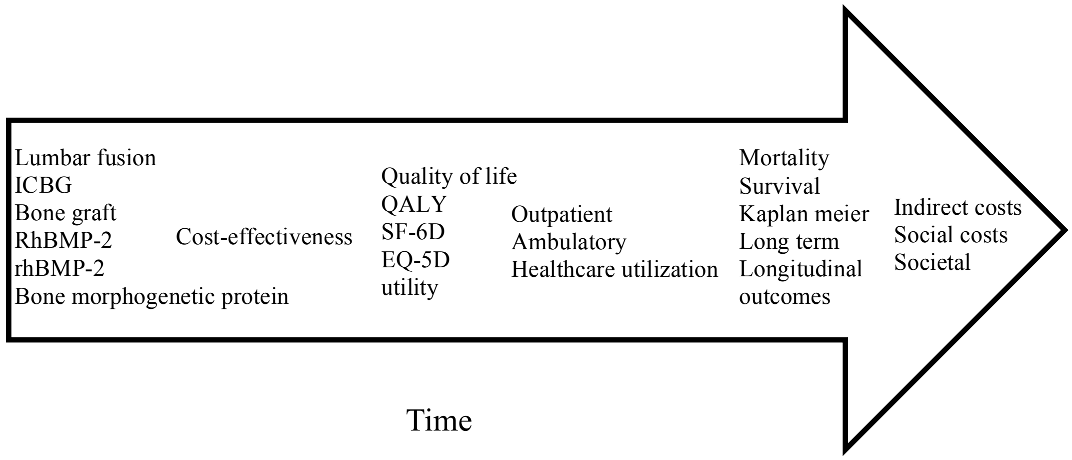

Figure 1. Search Keywords Listed Chronologically

The initial searches concentrated on gaining a better understanding of the general interventions and analyses being performed, and progressed to focusing on the methods of other CEAs for the purposes of model design. The subsequent searches reviewed the literature for clinical outcomes that would ultimately constitute the data set of inputs used in the analysis. Before being included or excluded for data extraction, publications were separated into groups according to the type of study or data they reported. Categories included usage and outcomes, CEA or economic perspectives, clinical trials, registry data, systematic reviews and meta-analyses, and utility or quality of life. Studies were included or excluded according to relevance and quality, and this process will be discussed in the remaining sections of the chapter. In addition to the stated search strategies, backward searching was also used for comprehensiveness. Backwards searching involved reviewing the references of the articles yielded from the keyword search, or reviewing what the authors had published prior to the article (48).

2.2 Model Design

The model for this analysis was designed using elements from three CEA studies. Carreon et al. used a decision tree model to calculate cost-effectiveness, while both Virk et al. and the AHRQ study used a Markov model (40,41,43). The Markov model was more useful in this case since (1) the decision problem involved risk that was continuous over time, (2) the timing of events was important, and (3) important events happened more than once (49). However, the decision tree was still helpful in identifying the series of patient events. Other CEAs did not influence the choice of model since they did not contain clear description of their model design (25,36,38,39).

The model was built using decision-analysis software (TreeAge Pro 2015, R1.0. TreeAge Software, Williamstown, MA). Markov models assume that a patient is always in one of a finite number of discrete health states, called Markov states (49). Patients can shift to a different state according to a transition probability. Each Markov state is associated with specific rewards such as cost and effectiveness, and patients accumulate rewards based on the states they transition to. For this model, the cohort began in the baseline state and underwent lumbar fusion with either rhBMP-2 or ICBG. After the initial decision, patients transitioned to one of three states: clinical success, clinical failure, or death. Patients who survived after the first six months could remain in the same state or flow to a different one relative to transition probabilities. The cohort cycled through the model every six month up to five years. The model is depicted in figure 2.

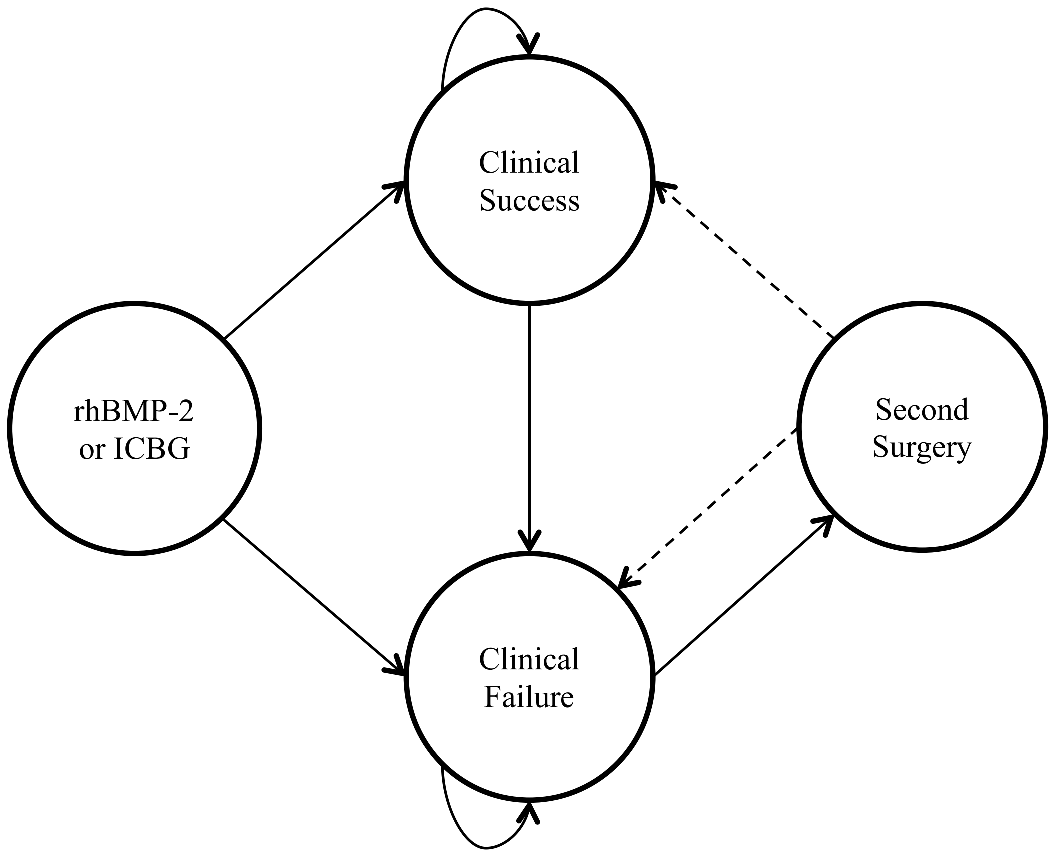

Figure 2. Markov Model showing patient states and transitions.

2.3 Model Parameters & Inputs

This literature-based CEA was estimated from both payer and societal perspectives using data selected from publications. In order to calculate the ICER, the model required total costs, utility, and transition probabilities that correspond to each Markov state. The following subsections will discuss how these model parameters were identified and extracted from published data. Actual input estimates will be described in the results chapter.

2.3.1 Direct Costs

CEAs published by Carreon et al. and Glassman et al. estimated costs [costs for what?] from actual reimbursements used for clinical trials from third party payers (36,40). For both studies, direct costs included the index surgical procedure, treatment of complications, emergency rooms, and outpatient visits. Glassman also published a study specifically recording the perioperative cost of rhBMP-2 versus ICBG (50). This economic analysis included an assessment of total costs, inpatient hospital costs, inpatient physician services, outpatient facility costs, outpatient physician services, and total payer expenditure within three months of the operation (50). In contrast, The AHRQ study used DRG, ICD-9, and CPT codes to identify cost data in the HCUPnet Nationwide Inpatient Sample (NIS) and CMS National Payment Amount-Physician Fee Schedule in order to estimate costs for lumbar fusion (41). DRG codes were used for the index procedure and reoperations, CPT codes for physician fee schedules, and ICD-9 codes other second operations such as removals and supplementary fixations (41). Since the AHRQ determined cost from national data sources, the direct costs for the initial procedure, second operations, and physician fee were used in this analysis. The AHRQ study did not estimate healthcare utilization costs postoperatively, so the remaining expenditures were approximated using data from Glassman et al. (50). These costs were comprised of therapist visits, pharmacy fees, radiology fees, and any outpatient facility or physician services. All extracted dollar values were converted to 2015 dollars.

2.3.2 Indirect Costs

The human capital approach for estimating indirect costs multiplies the time a patient is off work by the sum of the patient's wages and benefits for that period (29). Several clinical trials tracked the length of time needed for patients to return to work after the initial surgery (12,14,16,19,44). However, only two actually recorded the median length of time in days for each treatment group (14,44). Out of those studies, only Burkus et al. had a statistically significant result (44). Therefore, median return to work time was selected from that study. The national average weekly earnings from May 2015 were derived from the US Bureau of Labor Statistics, and then multiplied by return to work time to obtain indirect costs for each treatment group. Because there was no return to work data available for patients who had a second surgery, indirect costs were only estimated for the initial procedure and not for any second operations.

2.3.3 Clinical Outcomes & Utility

Recent clinical trials in lumbar fusion have used health-related quality of life (HRQOL) measures, such as the Short Form Health Survey 36 (SF-36) and the Oswestry Disability Index (ODI), for assessing clinical effectiveness (8,12,15–19,44). In order to incorporate these measures into a CEA, they must be converted into a utility index that is used to calculate QALY. The EuroQol Five Dimension (EQ-5D) questionnaire is the most widely used index for producing utility scores (51). However, none of the clinical trials considered in this analysis contained EQ-5D data. Previous CEAs recognized this issue, and utilized a method developed by Brazier et al. for converting the SF-36 to a 0 to 1 scale called the SF-6D (52). As a result, all utility estimates for this study were derived from the SF-6D. The clinical trial performed by Burkus et al. (44) had the largest sample size for using the types of bone graft (rhBMP-2 and ICBG) and technique (ALIF) most relevant to this analysis. Consequently, utility for each patient state up to two years were extracted from that trial. Utility estimates were still needed for the model past two years in addition to patients who needed and received a second operation. Utility estimates were still needed for patients who needed and received a second operation, and were extracted from articles written by Glassman et al. and Carreon et al. (36,53).

2.3.4 Transition Probabilities

Transition probabilities quantify the chance of a patient shifting from one Markov state to another. The transitions between different states were primarily defined by second surgery rates. Patients who needed a second surgery were classified as clinical failures, while those who did not were considered clinical successes. The patients could regress from a clinical success to a clinical failure, which led to the increase in the second surgery rate over time. Since there was disagreement in the literature regarding second operation rates, the base case CEA was performed using two different studies. The first study was performed by Burkus et al., while the second was a retrospective study that analyzed second operation rates using Medicare claims data (8,23). At each point in the model, there was a probability of death due to procedural or general causes. Perioperative probability was extracted from a study that investigated trends in the NIS, while only one article reported long-term general life expectancy after spinal surgery (54,55).

Assumptions were made for simplicity or to fill in missing data points. Second surgery and mortality rates that were not available at specific time points from the literature were extrapolated using data that was present. In order to extrapolate these rates, time was plotted against available data to generate a trend line. The line was chosen according to highest R^2 value, and the equation for the line was used to calculate the missing data. Due to the lack of follow-up data, patients who underwent a second operation were assumed to achieve successful fusion.

2.4 Sensitivity Analysis

We conducted sensitivity analyses to calculate the threshold values for second surgery rates and indirect costs that were required to change the ICER and thus the decision of the CEA. Since there was not enough data from the literature to generate confidence intervals for either variable, point estimates were of the minimum value needed to change the output of the model were calculated. Sensitivity analysis evaluating second surgery rates was performed since second surgery was as a significant variable since additional operations were the most costly intervention after the initial procedure (36,41,50). These analyses were performed on the payer perspective sub-analyses.

The second sensitivity analysis was performed on indirect costs using the societal perspective subanalyses. Indirect costs were a reflection of return to work status, which two studies have reported to be largely in favor of rhBMP-2 (14,44). As mentioned in chapter one, previous analyses have acknowledged that rhBMP-2 is associated with higher upfront costs, but concurred that those costs are offset due to either increased utility associated with no donor-site complications, decreased long-term healthcare utilization costs, or decreased loss in productivity (25,38–43). The threshold sensitivity analysis helped determine to what extent the addition of indirect costs due to losses in productivity impacted the CEA decision.

CHAPTER THREE

RESULTS

Chapter three contains the results from the literature review and the preliminary CEA. Section 3.1 consists of search strategies and the number of citations they generated, the flow of citations reviewed, and the estimates from data extraction that were used in the analysis. Section 3.2 includes output from the Markov model performed using the decision analysis software. Four different subanalyses were run, two from a payer perspective and two from a societal perspective. Finally, section 3.3 reports the results from the sensitivity analysis that investigates two variables, second surgery rates and indirect costs.

3.1 Literature Search

Only articles that were used in the development of the Markov model and its parameters are included in this chapter. Other references that contributed to general background knowledge were not considered part of the review.

3.1.1 Search Strategy

The literature search was executed using 26 keywords across three different databases. Thirteen combinations of keywords generated 2,484 citations. The most inclusive search produced 492 citations, while the most exclusive search yielded 16. There were parallel results between searches and databases, hence not all citations were unique. The search strategies and the number of results for each search are listed in Table 3.

Table 3. Search strategy and number of results

Search Strategies	Number of Results
(rhBMP-2 OR BMP OR bone morphogenetic protein) AND lumbar fusion	492
(rhBMP-2 OR BMP OR bone morphogenetic protein) AND cost-effectiveness	59
(rhBMP-2 OR BMP OR bone morphogenetic protein) AND lumbar fusion AND cost-effectiveness	16
lumbar fusion AND cost-effectiveness	116
(quality of life OR QALY OR SF-6D OR EQ-5D OR utility) AND lumbar fusion	415
(quality of life OR QALY OR SF-6D OR EQ-5D OR utility) AND lumbar fusion AND (rhBMP-2 OR BMP OR bone morphogenetic protein)	22
(quality of life OR QALY OR SF-6D OR EQ-5D OR utility) AND lumbar fusion AND (ICBG or "bone graft")	26
(outpatient OR ambulatory OR healthcare utilization) AND spinal fusion	178
mortality AND lumbar fusion	222
mortality AND lumbar fusion NOT perioperative	183
(mortality OR survival OR kaplan meier) AND lumbar fusion NOT perioperative	288
lumbar fusion AND (long term OR longitudinal OR outcomes)	427
(indirect costs OR social costs OR societal) AND lumbar fusion	40

Total	2,484

3.1.2 Literature Review

Studies were initially screened for duplications, which eliminated 1,502 references. The remaining 982 articles were screened for title or abstract, and 889 had keywords or results that did not match inclusion or exclusion criteria, which left 93 publications left for full-text evaluation. These studies were categorized according to results or study design, which included usage and outcomes, CEA or economic perspectives, clinical trials, registry data, systematic reviews and meta-analyses, and utility or quality of life. The full-text screen excluded 53 references, resulting in 40 publications left for data extraction. As described in previous chapter, point estimates were carefully selected from the literature for use in the Markov model. The flow of citations is illustrated in Figure 3.

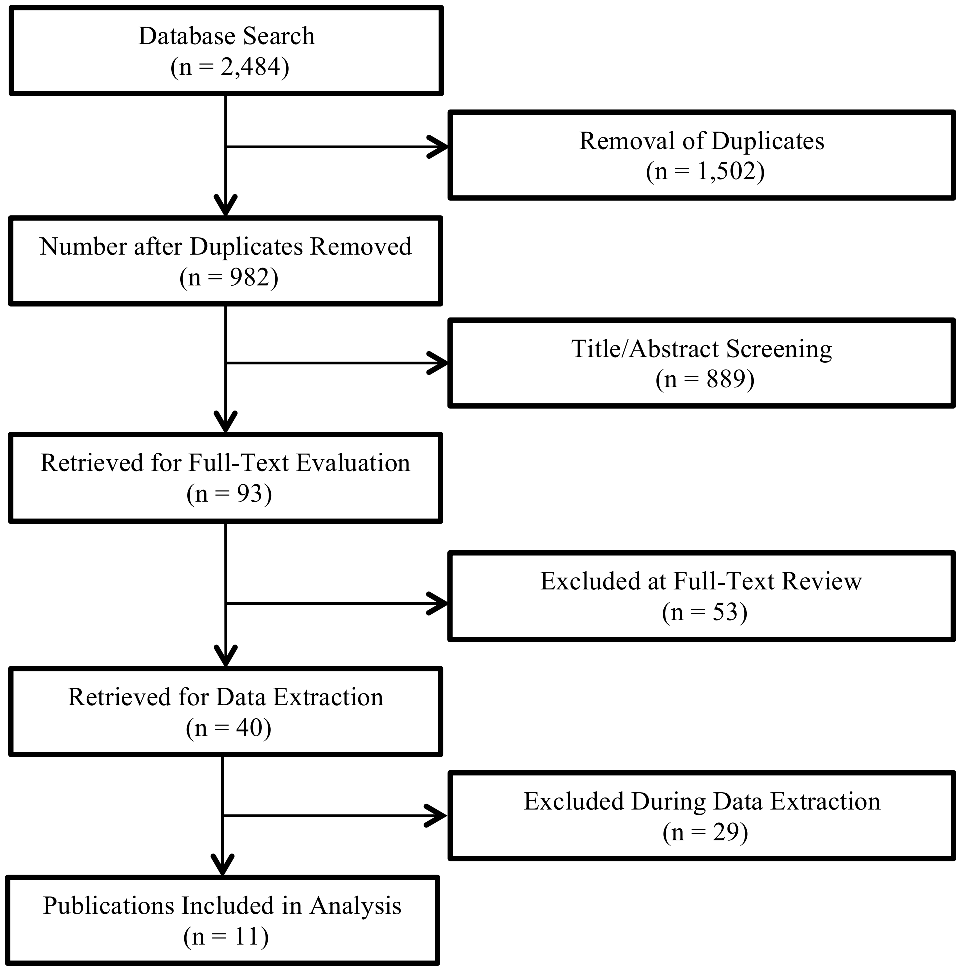

Figure 3. Flow diagram presenting results of the literature review

3.1.3 Data Extraction

Point estimates for each model parameter were extracted from 11 studies. These publications and the associated inputs are summarized in Table 4.

Table 4. Studies used for determining input point estimates

Reference	Model Input	Data Source(s)
Ratko et al. (41) (AHRQ)	Direct Costs	HCUPnet NIS, MPFS
Glassman et al. (36,50)	Direct Costs	Clinical trial reimbursements

Burkus et al. (44)	Indirect Costs	Clinical trial return to work
US Bureau of Labor Statistics (56)	Indirect Costs	CES survey
Burkus et al. (44)	Utility	Clinical trial PRO
Glassman et al. (36)	Utility	Clinical trial PRO
Carreon et al. (53)	Utility	Clinical trial PRO
Burkus et al. (8,13)	Second operation rates	Clinical trial
Deyo et al. (23)	Second operation rates	Medicare claims data
Pumberger et al. (54)	Mortality	HCUPnet NIS
Kim et al. (55)	Mortality	Clinical trials

Direct costs were categorized by accumulation in the inpatient or outpatient setting. Inpatient costs included the index lumbar fusion, surgeon's fee, rhBMP-2 kit, rehabilitation, or any second surgery. The initial inpatient costs totaled $35,442 for the rhBMP-2 group and $30,904 for the ICBG group. The difference was due to the cost of the bone graft substitute. There were four different types of secondary operations, with costs ranging from $13,822 to $30,904. Since the proportion of each surgery was not known, the costs were averaged to generate a mean second surgery cost equal to $21,427. This total was combined with the cost of inpatient rehabilitation to obtain the value of receiving additional interventions. Outpatient costs that included the typical physician visits, imaging, consults, lab fees, medications, and physical therapy corresponding to lumbar fusion amounted to $8,263. Patients who were clinical failures and had yet to receive additional intervention were expected to incur more of these outpatient healthcare utilization costs.

Indirect costs were calculated by multiplying the national average weekly earnings for May 2015 by the median return to work time for each treatment group. The mean indirect cost was $14,270 for rhBMP-2 patients, and $21,098 for ICBG patients. All cost estimates are listed in Table 5.

Table 5. Cost parameter estimates

Direct Costs			
	Inpatient		
		Lumbar fusion surgery (initial)	$29,374
		Surgeon fee	$1,530
		rhBMP-2 kit	$4,518
		Reoperation (secondary)	$30,904
		Device removal (secondary)	$13,980
		Complex supplemental fixation	$27,000
		Simple supplemental fixation	$13,822
	Outpatient		
		Mean non-physician reimbursement (facility fees, labs, imaging, medications, physical therapy)	$7,914
		Mean physician reimbursement (specialist consults, visits)	$349
Indirect Costs			
	Average weekly earnings (May 2015)	$861.12	
	Median Return to Work		
		rhBMP-2: 116 days (16.6 weeks)	$14,270
		ICBG: 171.5 days (24.5 weeks)	$21,098

Markov state utilities for each state in this analysis were kept constant. All clinical failures were assigned a utility value of 0.49, while clinical success after the first surgery equaled 0.64. Patients who attained clinical success after a second operation had a lower value of 0.54. Because the model cycled every 6 months, utility was multiplied by 0.5 to obtain QALYs. The values are recorded in Table 6.

Table 6. Utility estimates and QALYs

Markov State	SF-6D value	QALY (for 6 Months)
Baseline	0.54	0.27
Clinical Success	0.64	0.32
Clinical Failure	0.49	0.245
Clinical success, post second surgery	0.54	0.27

Transition probabilities were derived from second surgery and mortality rates. The second surgery rates for rhBMP-2 patients extracted from Burkus et al. included values ranging from 0.025 to 0.14 for each time point during the five-year analysis (8,13). The same study only provided a two-year rate of 0.103 for ICBG patients, thus the other values were extrapolated using the trend line created by the rhBMP-2 data. Since Deyo et al. supplied rates up to four years for both treatment groups, only rates for the fifth year were extrapolated (23). Second surgery rates for rhBMP-2 (0.108) and ICBG (0.113) after five years in the retrospective study by Deyo et al. were lower than the rates for rhBMP-2 (0.14) and ICBG (0.173) in the clinical trial performed by Burkus et al. (8,13,23). Perioperative mortality was equal to 0.002, while non-perioperative mortality was extracted from a survival curve published in a study researching life expectancy after spinal surgery (55). Table 7 contains the transition probability estimates.

Table 7. Model transition probability point estimates (†extrapolated value)

Time (months)	Second Surgery Rate (Burkus)		Second Surgery Rate (Deyo)		Mortality	
	rhBMP-2	ICBG	rhBMP-2	ICBG	rhBMP-2	ICBG
6	0.025	0.058	0.012	0.012	0.01	0.01
12	0.043	0.076	0.027	0.029	0.01	0.01
18	0.060	0.093†	0.045†	0.044†	0.015	0.015
24	0.070	0.103	0.063	0.06	0.018	0.018
30	0.085	0.118†	0.072†	0.071†	0.02	0.02
36	0.089	0.122†	0.092	0.085	0.02	0.02
42	0.097	0.130†	0.092†	0.092†	0.023	0.023
48	0.100	0.133†	0.108	0.105	0.023	0.023
54	0.130	0.163†	0.108†	0.108†	0.025	0.025
60	0.140	0.173†	0.108†	0.113†	0.025	0.025

3.2 Preliminary Cost-effectiveness Analysis

The results of the base case CEA are listed in Table 8. All cost and effectiveness estimates were discounted at an annual rate of 3%. Two subanalyses employed the second surgery rates derived from Burkus et al. (13,44). The first model was performed from a payer perspective. Cumulative cost and effectiveness over five years for rhBMP-2 were $50,968 and 2.85 QALY, while the cumulative cost and effectiveness for ICBG were $47,258 and 2.84 QALY. The ICER was $251,026 per QALY between the treatment groups, and ICBG was more the cost-effective option. The second subanalysis was performed from a societal perspective by incorporating indirect costs. For this scenario, rhBMP-2 was dominant (See Chapter one for definition of dominance). Cumulative costs and effectiveness were $62,573 and 2.85 QALY for rhBMP-2, and $65,405 and 2.84 QALY for ICBG.

The other subanalyses used second surgery rates observed by Deyo et al. (23). From the payer perspective, ICBG was more cost-effective. Cumulative costs were $49,465 for rhBMP-2 and $45,482 for ICBG. Cumulative effectiveness was 2.85 QALY for both groups, and the ICER was $849,893 per QALY. From the societal perspective, rhBMP-2 was dominant. Cumulative costs and effectiveness totaled $61,567 and 2.87 QALY for rhBMP-2, and $64,162 and 2.86 QALY for ICBG.

Table 8. Results from the Base Case CEA (*in favor of rhBMP-2)

Payer Perspective	Cumulative Cost ($)		Cumulative QALY		$/QALY		ICER
	rhBMP-2	ICBG	rhBMP-2	ICBG	rhBMP-2	ICBG	($/QALY)
Burkus et al.	50,968	47,258	2.85	2.84	17,881	16,668	251,026
Deyo et al.	49,465	45,482	2.86	2.86	17,277	15,912	849,893
Societal Perspective	**Cumulative Cost ($)**		**Cumulative QALY**		**$/QALY**		**ICER**
	rhBMP-2	ICBG	rhBMP-2	ICBG	rhBMP-2	ICBG	($/QALY)
Burkus et al.	62,572	65,4045	2.85	2.84	21,917	23,042	dominant*
Deyo et al.	61,567	64,162	2.87	2.86	21,466	22,405	dominant*

3.3 Sensitivity Analysis

The base case results found that ICBG was more cost-effective from a payer perspective, while rhBMP-2 was more cost-effective from a societal perspective. The first threshold sensitivity analysis was performed on the payer perspective analyses using second surgery rates to change the optimal strategy to rhBMP-2. The minimum difference in second surgery rate after five years required to achieve this outcome was approximately 0.13. In other words, the second surgery rate for rhBMP-2 had to be 13% less than ICBG for the treatment to become cost-effective. The second threshold sensitivity analysis was performed on the societal perspective analyses using indirect costs to change the optimal strategy to ICBG. The minimum difference in indirect costs required to achieve this outcome was approximately $4,000. Therefore, if the variation in median return to work between treatment groups was under five weeks, rhBMP-2 was no longer cost-effective from a societal perspective. The results from the sensitivity analysis are located in Table 9.

Table 9. Sensitivity Analysis Results

Payer Perspective	Variable	Value Required to Change Decision from ICBG to rhBMP-2
Burkus et al.	Second Surgery Rates	0.130 at five years
Deyo et al.	Second Surgery Rates	0.135 at five years
Societal Perspective	Variable	Value Required to Change Decision **from** rhBMP-2 to ICBG
Burkus et al.	Indirect Costs	$3,990.19 (4.6 weeks)
Deyo et al.	Indirect Costs	$4,227.64 (4.9 weeks)

CHAPTER FOUR

DISCUSSION & CONCLUSION

Chapter four discusses the findings from the literature review and the preliminary cost-effectiveness analysis. Sections 4.1-4.2 summarize the results of the CEA and sensitivity analysis in the context of the hypothesis and research questions listed in chapter one. Section 4.3 re-examines limitations of the existing literature, and how those elements affected inputting published data into the initial model. Section 4.4 describes limitations & strengths of this CEA, while section 4.5 explains how the results of the analysis and literature review relate to the larger project and future research. Lastly, section 4.6 explores the research implications, and section 4.6 presents the final conclusions.

4.1 Preliminary Cost-effectiveness Analysis

Hypothesis: rhBMP-2 is more cost-effective than ICBG

Although ICBG was slightly more cost-effective, neither strategy was dominant from both payer perspective subanalyses. The ICER values for those analyses were $251,025.98 per QALY and $849,893.21 per QALY, which were far above the $50,000 to $100,000 threshold mentioned in chapter one. The reason for this large value was small denominator in the ICER, which corresponded to the small difference in cumulative QALY gained. In this case, the ICER only served as comparison between rhBMP-2 and ICBG, which made it seem that neither intervention was cost-effective. However, the decision to perform lumbar fusion was already proven to be cost-effective versus nonoperative treatment with $ per QALY values less than 50,000 over five years(36,43,57). This result supports the idea that clinicians can vary the threshold ICER according to the decision-making context (37). The undominated result favoring ICBG disagreed with the three other CEAs comparing rhBMP-2 and ICBG performed from a payer perspective. Carreon et al. revealed higher costs and lower improvements in utility seen in patients over 60 years of age receiving ICBG (40). In this work, we used outcomes from patients whose mean age was closer to 40 years old, which could explain the difference in costs and utility. It is possible that older patients do not respond as well to the increased invasiveness of ICBG, and have increased healthcare utilization as a result. Virk et al. favored rhBMP-2 because of decreased costs associated with reduced revision rates (40,43). These rates were dissimilar enough to make up for the difference in initial costs, while the second surgery rates used for this study were not. Ratko et al. preferred rhBMP-2 due to assumed decreases in utility resulting from ICBG donor site pain (41), but this analysis did not make those assumptions since those values were not observed in the literature (8,12,14,16,19,36,45).

From both societal perspectives, rhBMP-2 was more cost-effective. Even though the decrease in lost productivity for the rhBMP-2 group was enough to displace the higher initial cost, the differences in cumulative cost (≈ $3,000) and cumulative QALYs gained (0.01) were modest. This result was difficult to compare to two earlier CEAs who used a societal perspective since both used foreign cost estimates (38,39). However, both studies used the same return to work data from Burkus et al. (44), which they interpreted differently. The interpretations of this return to work data will be discussed further in the limitations section.

4.2 Sensitivity Analysis

Research Question 1: What difference in second surgery rates changes the decision of the CEA?

The sensitivity analysis determined that from a payer perspective, rhBMP-2 needed to decrease second surgery rates by approximately 13% after five years to become more cost-effective than ICBG. However, this difference in rates is unlikely to occur based on the literature. The studies performed by Burkus et al. and Deyo et al. that were used to estimate second surgery rates over a five-year period had final differences of 3.3% and 0.52% in favor of rhBMP-2 respectively (13,23,44). No other study had rates that could be extrapolated past 24 months, and only one that compared rhBMP-2 and ICBG patients had a rate difference greater than 10% during that time period (18). Furthermore, this data may have been affected by the clinical trial only including patients older than 60 years of age, whose risk for complications from graft harvest and prolonged surgery are elevated (18).

Research Question 2: What difference in indirect costs changes the decision of the CEA?

The base case analyses performed from a societal perspective had a difference in indirect costs of $6,828 in favor of rhBMP-2, which converts to a difference of 8 weeks in median return to work time between surgical approaches. The threshold sensitivity analysis determined that rhBMP-2 would no longer be cost-effective compared to ICBG if the variation in indirect costs were reduced to approximately $4,000, which translates to 4.6 weeks in median return to work. Since only indirect costs from the initial procedure were included, cost estimates were probably understated. There was only one article that calculated indirect costs for reoperations (58), but none that estimated costs for the other secondary surgeries. As a result, indirect costs associated with secondary procedures were excluded from the analysis.

4.3 Limitations of the Existing Literature

This analysis had inherent limitations due to its reliance on the secondary use of published data. One important limitation was the lack of long-term follow up of second surgery rates, which was also noticed by Virk et al. (43). Only one article reported prospective second surgery rates for rhBMP-2 up to five years (13), and zero studies reported ICBG second surgery rates past two years. Consequently, the prospective ICBG rates were extrapolated based on trends observed in the rhBMP-2 data, and the actual trend may diverge from those predictions. The paper written by Deyo et al. was the only study to report reoperations past two years for both rhBMP-2 and ICBG (23), but those rates were specific to Medicare beneficiaries who underwent fusion for lumbar stenosis. Furthermore, the study was performed retrospectively and reported no difference in rates, which disagreed with clinical trials that discovered a variation in second surgery rates favoring rhBMP-2. The retrospective design may have suffered from selection bias that influenced its results, but there were no other retrospective studies available for comparison that could confirm this notion.

Another literature related limitation was the lack of detailed reporting on return to work status after surgery. Several studies either did not report return to work or reported the percentage of patients who returned to work at a single time point such as two years, but did not mention the actual time missed by people who received initial or secondary interventions (10–12,16,17,19,20,59). The return to work data used for this analysis, published by Burkus et al., was previously used in two other CEAs. Garrison et al. found that this data was sometimes difficult to interpret because of unclear or inappropriate methods used for data analysis and results presentation (38), while Alt et al. also noted incomplete presentation of results (39). As a result, both studies requested access to the raw data, and recalculated the return to work time by incorporating preoperative work status. This recalculation significantly reduced the difference in return to work between the treatment groups, with Alt et al. finding rhBMP-2 patients returning to work 43 days earlier opposed to 55 days (39), and Garrison et al. calculating no difference between rhBMP-2 and ICBG. Haid et al. was the only other clinical trial to report median return to work time, noting a statistically insignificant 94-day difference in favor of rhBMP-2 (14). Ultimately, there were deficiencies in the return to work data used for indirect costs, which could explain why the majority of CEAs comparing rhBMP-2 and ICBG excluded those parameters from their analyses.

The last significant limitation pertaining to the literature involved utility values. The pilot clinical trial performed by Burkus et al. used the SF-36 as a HRQOL primary measure to assess clinical effectiveness (8), which set the standard for subsequent trials. In order to convert clinical effectiveness to QALYs, Brazier et al. developed a method for indexing SF-36 scores to a 0 to 1 scale called the SF-6D (52). Five out of six of the published CEAs summarized in chapter one used the SF-6D utility index for generating QALYs in their analyses. In another CEA evaluating the economic value of spinal stenosis surgery, Tosteson et al. recognized that QALY gains were somewhat lower when estimated with the SF-6D instead of the EQ-5D, which was found to be less responsive to low back treatment (60). This discrepancy in utility estimates was noticed when comparing the results clinical trials using difference effectiveness measures. Glassman et al. estimated a cumulative 2-year gain in utility of 0.13 using the SF-6D (36), while Adogwa et al. estimated a cumulative 2-year gain in utility of 0.35 using the EQ-5D (61). Moreover, another study performed by Brazier et al. discovered that the EQ-5D and SF-6D are not interchangeable and have shown significantly different results when applied to patients with similar disorders (62). Perhaps there is a larger disparity in QALYs gained between rhBMP-2 and ICBG that the SF-6D is not detecting, which the EQ-5D would be able to recognize. This idea is also relevant to donor site pain in ICBG patients, which is a well known side effect of surgery that is not reflected in SF-6D scores, hence Ratko et al. could be justified in their utility assumptions (41). However, this result could also be due to the reporting of the measure, when surveys are not collected until 6 or 12 months after surgery when the donor site pain has become manageable.

4.4 CEA Strengths & Limitations

Even with the exploratory approach to the analysis, this CEA had some discernable strengths. One advantage was using a Markov model to analyze the data. This method helped to better simulate a situation where events or patients states can recur over time, in comparison to a decision tree model that only considers one-time events. Another strength was performing multiple subanalyses that incorporated different perspectives and transition probabilities. Although the return to work data was not ideal, it was still important to understand how performing the analysis from a societal perspective might influence the results. Additionally, inputting different second surgery rates acknowledged the disagreement between prospective and retrospective data concerning rhBMP-2 and ICBG, which resulted in a more comprehensive analysis.

A further potential advantage may have been generating transition probabilities from second surgery rates, thus establishing the incidence of additional surgical intervention as the predictor of clinical success or failure. At their most basic meanings, lumbar fusion success is the radiographic union of vertebrae, while lumbar fusion failure is the radiographic non-union of vertebrae. However, radiographic outcomes are not completely predictive of clinical outcomes as indicated in Burkus et al., who reported that 90% of patients receiving one type of second surgery previously had radiographic evidence of fusion (8). It is reasonable to expect a patient who has a successful operation from a clinical perspective would consume the minimum amount of medical resources associated with that intervention. For this model, that success resulted not having a second surgery. The opposite would be true for a clinical failure, where the first procedure does not alleviate the symptoms, leading to the utilization of more medical resources that culminates in a second surgery. From a cost-effectiveness standpoint, the most significant outcomes were those that accurately predicted increases or decreases in healthcare utilization.

There were limitations in this CEA that were specific to the model and not a product of the literature. These limitations were largely an effect of assumptions related to the preliminary nature of the analysis, and the larger ensuing CEA will not necessarily make the same assumptions. One limitation involved excluding costs associated with non-operative complications, which are common in surgical procedures. It was decided that complications requiring addition surgical intervention were more indicative of overall costs, thus non-operative complications were not incorporated. In order to simplify utility corresponding a particular Markov state, it was assumed that one state was limited to one utility value. In actuality, the incremental utility of a state may better or worsen over time. Another limitation was the combination of costs and rates of second surgeries. The second surgery rate included four different types of operations that all had different costs. Since it was difficult to decipher the probabilities for each surgery by treatment group from the literature, the rates were combined by treatment group into a single second surgery rate, and costs into a single mean second surgery cost. This process could have overestimated or underestimated costs based on the distribution of second surgery type. The final limitation was the assumption that each patient was a clinical success after a second surgery. This outcome is probably not the case in reality, and incorporating those costs and probabilities into the model might have influenced the results.

4.5 Future Research

Further research is needed to address the limitations mentioned in section 4.3. As the data quality in spine registries and inpatient databases such as SPARCS improves, additional retrospective studies should be performed on the long-term follow up of second surgery rates. These studies would further elucidate the discrepancies between prospective and retrospective data, and hopefully provide more consensus regarding the differences in rates between rhBMP-2 and ICBG. Return to work data would be difficult to collect retrospectively, hence additional prospective studies would be required to measure work status for the calculation of indirect costs. It was clear that indirect costs can have a significant effect on the CEA, and the use of higher quality return to work data would more accurately demonstrate the actual importance of that effect. Similarly, utility values would be hard to estimate from retrospective data, and additional prospective research would be necessary to obtain new HQROL data. It is recommended that those studies use the EQ-5D as the primary measure of utility rather than the SF-36 or SF-6D to ensure a more complete recording of changes in quality of life.

One example of future research that addresses some of these limitations is the larger CEA that will rely on data from SPARCS. Since SPARCS will mainly track inpatient costs and outcomes, any outpatient model inputs will be estimated from the literature. As a result, the primary purpose of the literature review aspect of this thesis was to identify studies that can be used to estimate model parameters that cannot be extracted from SPARCS. Forty studies were ultimately retrieved for data extraction, which will help estimate direct costs attributed to outpatient healthcare utilization, indirect costs, and utility, while SPARCS will be used to acquire direct costs associated with inpatient care and transition probabilities that are derived from fusion rates or second surgery rates. The pilot investigation of the CEA model established which inputs may have a significant impact on the model, while indicating how the results might emerge depending on the estimation of transition probabilities and indirect costs.

4.6 Research Implications

The disagreement between the societal and payer perspectives demonstrated in the preliminary CEA has significant implications. The Panel on Cost-Effectiveness in Health and Medicine currently recommends that CEA be performed from the societal perspective (31), since it serves the goal of facilitating comparisons across interventions and patient groups (27). However, CEAs performed from the societal perspective may not provide the most appropriate information to specific decision-makers. Third-party payers are concerned with the direct costs of care, but not with how soon a patient returns to work after surgery, when deciding to cover a certain intervention. In this case, payers would be more likely to choose ICBG over rhBMP-2 based on direct costs. On the other hand, if they recognize that their choice is affecting more downstream costs, payers may reconsider their decision. Even though the societal perspective does not show all parties exactly what they want to know to make choices best suited to their interests, it asks they be aware of and consider the interests of others (27,63). By incorporating all costs and health effects, a CEA performed from the societal perspective can provide the basis for decisions that are fair to all parties, an agenda for negotiating such decisions, and information essential for designing compensation and incentives to support them (63).

4.7 Conclusion

The objectives for this thesis were to conduct a literature review to inform a larger CEA comparing rhBMP-2 and ICBG, and to pilot investigate the CEA model using only data from the literature. Studies were systematically searched, reviewed, and identified for the primary goal of contributing to future research, and the secondary goal of completing the preliminary analysis. The exclusively literature-based CEA discovered that rhBMP-2 was less cost-effective than ICBG from a payer perspective, but more cost-effective from a societal perspective. Furthermore, sensitivity analysis revealed that the payer perspective decision changed when the difference in second surgery rates approached 13% over five years, and that the societal perspective decision changed when the difference in indirect costs was $3,000 or less.

REFERENCES

1. Phillips FM, Slosar PJ, Youssef JA, Andersson G, Papatheofanis F. Lumbar spine fusion for chronic low back pain due to degenerative disc disease: a systematic review. Spine. 2013 Apr 1;38(7):E409–22.

2. Peter F. Ullrich, Jr. Lumbar Spinal Fusion Surgery [Internet]. Spine-Health. [cited 2015 May 18]. Available from: http://www.spine-health.com/treatment/spinal-fusion/lumbar-spinal-fusion-surgery

3. Ali Araghi, Peter F. Ullrich, Jr. Bone Graft Options for Spine Fusion [Internet]. Spine-Health. [cited 2015 May 19]. Available from: http://www.spine-health.com/treatment/spinal-fusion/bone-graft-options-spine-fusion

4. Peter F. Ullrich, Jr. Autograft: The Patient's Own Bone [Internet]. [cited 2015 May 20]. Available from: http://www.spine-health.com/treatment/spinal-fusion/autograft-patients-own-bone

5. Sasso RC, LeHuec JC, Shaffrey C, Group SIR, others. Iliac crest bone graft donor site pain after anterior lumbar interbody fusion: a prospective patient satisfaction outcome assessment. J Spinal Disord Tech. 2005;18:S77–81.

6. U.S. Food and Drug Administration. Recently-Approved Devices > InFUSE™ Bone Graft/LT-CAGE™ Lumbar Tapered Fusion Device - P000058 [Internet]. FDA. [cited 2015 May 20]. Available from: http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm083423.htm

7. Urist MR. Bone: formation by autoinduction. 1965. Clin Orthop. 2002 Feb;(395):4–10.

8. Burkus JK, Gornet MF, Dickman CA, Zdeblick TA. Anterior lumbar interbody fusion using rhBMP-2 with tapered interbody cages. J Spinal Disord Tech. 2002;15(5):337–49.

9. Martin BI, Lurie JD, Tosteson ANA, Deyo RA, Farrokhi FR, Mirza SK. Use of bone morphogenetic protein among patients undergoing fusion for degenerative diagnoses in the United States, 2002 to 2012. Spine J. 2015 Apr;15(4):692–9.

10. Boden SD, Zdeblick TA, Sandhu HS, Heim SE. The use of rhBMP-2 in interbody fusion cages: definitive evidence of osteoinduction in humans: a preliminary report. Spine. 2000;25(3):376–81.

11. Boden SD, Kang J, Sandhu H, Heller JG. Use of recombinant human bone morphogenetic protein-2 to achieve posterolateral lumbar spine fusion in humans: a prospective, Randomized clinical pilot trial 2002 Volvo Award in clinical studies. Spine. 2002;27(23):2662–73.

12. Burkus JK, Sandhu HS, Gornet MF, Longley MC. Use of rhBMP-2 in combination with structural cortical allografts: clinical and radiographic outcomes in anterior lumbar spinal surgery. J Bone Joint Surg Am. 2005 Jun;87(6):1205–12.

13. Burkus JK. Six-Year Outcomes of Anterior Lumbar Interbody Arthrodesis with Use of Interbody Fusion Cages and Recombinant Human Bone Morphogenetic Protein-2. J Bone Jt Surg Am. 2009 May 1;91(5):1181.

14. Haid RW, Branch CL, Alexander JT, Burkus JK. Posterior lumbar interbody fusion using recombinant human bone morphogenetic protein type 2 with cylindrical interbody cages. Spine J. 2004 Sep;4(5):527–38.

15. Dimar JR, Glassman SD, Burkus KJ, Carreon LY. Clinical outcomes and fusion success at 2 years of single-level instrumented posterolateral fusions with recombinant human bone morphogenetic protein-2/compression resistant matrix versus iliac crest bone graft. Spine. 2006;31(22):2534–9.

16. DimarII JR. Clinical and Radiographic Analysis of an Optimized rhBMP-2 Formulation as an Autograft Replacement in Posterolateral Lumbar Spine Arthrodesis. J Bone Jt Surg Am. 2009 Jun 1;91(6):1377.

17. Slosar PJ, Josey R, Reynolds J. Accelerating lumbar fusions by combining rhBMP-2 with allograft bone: a prospective analysis of interbody fusion rates and clinical outcomes. Spine J. 2007 May;7(3):301–7.

18. Glassman SD, Carreon LY, Djurasovic M, Campbell MJ, Puno RM, Johnson JR, et al. RhBMP-2 versus iliac crest bone graft for lumbar spine fusion: a randomized, controlled trial in patients over sixty years of age. Spine. 2008;33(26):2843–9.

19. Dawson E. Recombinant Human Bone Morphogenetic Protein-2 on an Absorbable Collagen Sponge with an Osteoconductive Bulking Agent in Posterolateral Arthrodesis with Instrumentation<sbt aid=“1392156”>A Prospective Randomized Trial</sbt> J Bone Jt Surg Am. 2009 Jul 1;91(7):1604.

20. Hurlbert RJ, Alexander D, Bailey S, Mahood J, Abraham E, McBroom R, et al. rhBMP-2 for Posterolateral Instrumented Lumbar Fusion: A Multicenter Prospective Randomized Controlled Trial. Spine. 2013 Dec;38(25):2139–48.

21. Guppy KH, Paxton EW, Harris J, Alvarez J, Bernbeck J. Does Bone Morphogenetic Protein Change the Operative Nonunion Rates in Spine Fusions?: Spine. 2014 Oct;39(22):1831–9.

22. Rihn JA, Patel R, Makda J, Hong J, Anderson DG, Vaccaro AR, et al. Complications associated with single-level transforaminal lumbar interbody fusion. Spine J. 2009 Aug;9(8):623–9.

23. Deyo RA, Ching A, Matsen L, Martin BI, Kreuter W, Jarvik JG, et al. Use of Bone Morphogenetic Proteins in Spinal Fusion Surgery for Older Adults With Lumbar Stenosis: Trends, Complications, Repeat Surgery, and Charges. Spine. 2012 Feb;37(3):222–30.

24. Adams CL, Ogden K, Robertson IK, Broadhurst S, Edis D. Effectiveness and Safety of Recombinant Human Bone Morphogenetic Protein-2 Versus Local Bone Graft in Primary Lumbar Interbody Fusions: Spine. 2014 Jan;39(2):164–71.

25. Polly DW, Ackerman SJ, Shaffrey CI, Ogilvie JW, Wang JC, Stralka SW, et al. A cost analysis of bone morphogenetic protein versus autogenous iliac crest bone graft in single-level anterior lumbar fusion. Orthopedics. 2003 Oct;26(10):1027–37.

26. Carragee EJ, Hurwitz EL, Weiner BK. A critical review of recombinant human bone morphogenetic protein-2 trials in spinal surgery: emerging safety concerns and lessons learned. Spine J. 2011 Jun;11(6):471–91.

27. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Jama. 1996;276(14):1172–7.

28. Tosteson ANA, Skinner JS, Tosteson TD, Lurie JD, Andersson GB, Berven S, et al. The cost effectiveness of surgical versus nonoperative treatment for lumbar disc herniation over two years: evidence from the Spine Patient Outcomes Research Trial (SPORT). Spine. 2008 Sep 1;33(19):2108–15.

29. Rihn JA, Currier BL, Phillips FM, Glassman SD, Albert TJ. Defining the value of spine care. J Am Acad Orthop Surg. 2013;21(7):419–26.

30. Fayssoux R, Goldfarb NI, Vaccaro AR, Harrop J. Indirect costs associated with surgery for low back pain—A secondary analysis of clinical trial data. Popul Health Manag. 2010;13(1):9–13.

31. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996 Oct 23;276(16):1339–41.

32. McCormick JD, Werner BC, Shimer AL. Patient-reported outcome measures in spine surgery. J Am Acad Orthop Surg. 2013;21(2):99–107.

33. Cohen DJ, Reynolds MR. Interpreting the results of cost-effectiveness studies. J Am Coll Cardiol. 2008 Dec 16;52(25):2119–26.

34. Benjamin P Geisler. An Overview of Cost Modeling and Cost-Effectiveness Analysis. 2010 Aug 12; MIT.

35. TreeAge Pro 2015 User's Manual. 2015 TreeAge Software, Inc.;

36. Glassman SD, Polly DW, Dimar JR, Carreon LY. The Cost Effectiveness of Single-Level Instrumented Posterolateral Lumbar Fusion at 5 Years After Surgery: Spine. 2012 Apr;37(9):769–74.

37. Simoens S. How to assess the value of medicines? Front Pharmacol [Internet]. 2010 [cited 2015 Jul 13];1. Available from: http://journal.frontiersin.org/article/10.3389/fphar.2010.00115/abstract

38. Garrison KR, Donell S, Ryder J, Shemilt I, Mugford M, Harvey I, et al. Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review. Health Technol Assess Winch Engl. 2007 Aug;11(30):1–150, iii – iv.

39. Alt V, Chhabra A, Franke J, Cuche M, Schnettler R, Le Huec J-C. An economic analysis of using rhBMP-2 for lumbar fusion in Germany, France and UK from a societal perspective. Eur Spine J. 2009 Jun;18(6):800–6.

40. Carreon LY, Glassman SD, Djurasovic M, Campbell MJ, Puno RM, Johnson JR, et al. RhBMP-2 versus iliac crest bone graft for lumbar spine fusion in patients over 60 years of age: a cost-utility study. Spine. 2009;34(3):238–43.

41. Ratko TA, Belinson SE, Samson DJ, Bonnell C, Ziegler KM, Aronson N. Bone Morphogenetic Protein: The State of the Evidence of On-Label and Off-Label Use [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010 [cited 2015 May 13]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK284865/

42. Hsu WK, Hashimoto RE, Berven SH, Nassr A. Biological Substitutes/Extenders for Spinal Arthrodesis: Which Agents Are Cost-effective? Spine. 2014 Oct;39:S86–98.

43. Virk S, Sandhu HS, Khan SN. Cost effectiveness analysis of graft options in spinal fusion surgery using a Markov model. J Spinal Disord Tech. 2012;25(7):E204–10.

44. Burkus JK, Heim SE, Gornet MF, Zdeblick TA. Is INFUSE bone graft superior to autograft bone? An integrated analysis of clinical trials using the LT-CAGE lumbar tapered fusion device. J Spinal Disord Tech. 2003;16(2):113–22.

45. Lubelski D, Alvin M, Torre-Healy A, Abdullah K, Nowacki A, Whitmore R, et al. Quality-of-Life Outcomes following Thoracolumbar and Lumbar Fusion with and without the Use of Recombinant Human Bone Morphogenetic Protein-2:

Does Recombinant Human Bone Morphogenetic Protein-2 Make a Difference? Glob Spine J. 2014 Oct 10;04(04):245–54.

46. Statewide Planning and Research Cooperative System [Internet]. [cited 2015 Jun 8]. Available from: https://www.health.ny.gov/statistics/sparcs/

47. Cooper HM. Organizing knowledge syntheses: A taxonomy of literature reviews. Knowl Soc. 1988 Mar;1(1):104–26.

48. Levy Y, Ellis TJ. A systems approach to conduct an effective literature review in support of information systems research. Informing Sci Int J Emerg Transdiscipl. 2006;9(1):181–212.

49. Sonnenberg FA, Beck JR. Markov Models in Medical Decision Making: A Practical Guide. Med Decis Making. 1993 Dec 1;13(4):322–38.

50. Glassman SD, Carreon LY, Campbell MJ, Johnson JR, Puno RM, Djurasovic M, et al. The perioperative cost of Infuse bone graft in posterolateral lumbar spine fusion. Spine J. 2008 May;8(3):443–8.

51. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health Qual Life Outcomes. 2009;7(1):27.

52. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002 Mar;21(2):271–92.

53. Carreon LY, Djurasovic M, Canan CE, Burke LO, Glassman SD. SF-6D Values Stratified by Specific Diagnostic Indication: Spine. 2012 Jun;37(13):E804–8.

54. Pumberger M, Chiu YL, Ma Y, Girardi FP, Vougioukas V, Memtsoudis SG. Perioperative mortality after lumbar spinal fusion surgery: an analysis of epidemiology and risk factors. Eur Spine J. 2012 Aug;21(8):1633–9.

55. Kim H-J, Lee H-M, Kim H-S, Moon E-S, Park J-O, Lee K-J, et al. Life expectancy after lumbar spine surgery: one-to eleven-year follow-up of 1015 patients. Spine. 2008;33(19):2116–21.

56. Bureau of Labor Statistics. Table B-3. Average hourly and weekly earnings of all employees on private nonfarm payrolls by industry sector, seasonally adjusted [Internet]. [cited 2015 Jun 17]. Available from: http://www.bls.gov/news.release/empsit.t19.htm

57. Cahill KS, Wang MY. Cost effectiveness of lumbar fusion improves with time. Neurosurgery. 2012;70(2):N21.

58. Adogwa O, Parker SL, Shau D, Mendelhall SK, Aaronson O, Cheng J, et al. Cost Per Quality-adjusted Life Year Gained of Revision Fusion for Lumbar

Pseudoarthrosis: Defining the Value of Surgery. J Spinal Disord Tech. 2015 Apr;28(3):101–5.

59. Vaccaro AR, Anderson DG, Patel T, Fischgrund J, Truumees E, Herkowitz HN, et al. Comparison of OP-1 Putty (rhBMP-7) to iliac crest autograft for posterolateral lumbar arthrodesis: a minimum 2-year follow-up pilot study. Spine. 2005;30(24):2709–16.

60. Tosteson ANA, Lurie JD, Tosteson TD, Skinner JS, Herkowitz H, Albert T, et al. Surgical treatment of spinal stenosis with and without degenerative spondylolisthesis: cost-effectiveness after 2 years. Ann Intern Med. 2008 Dec 16;149(12):845–53.

61. Adogwa O, Carr RK, Kudyba K, Karikari I, Bagley CA, Gokaslan ZL, et al. Revision lumbar surgery in elderly patients with symptomatic pseudarthrosis, adjacent-segment disease, or same-level recurrent stenosis. Part 1. Two-year outcomes and clinical efficacy: clinical article. J Neurosurg Spine. 2013;18(2):139–46.

62. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004 Sep;13(9):873–84.

63. Russell LB, Fryback DG, Sonnenberg FA. Is the societal perspective in cost-effectiveness analysis useful for decision makers? Jt Comm J Qual Improv. 1999 Sep;25(9):447–54.