Title
Investigating Cost Variation Within Spinal Fusion Payment Groups

Author
Wright, David

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Investigating Cost Variation Within Spinal Fusion Payment Groups

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Biomedical and Translational Science

by

David Wright

Thesis Committee:
Assistant Clinical Professor Dr. Samuel Bederman, Chair
Professor Dana B. Mukamel
Professor Sheldon Greenfield

2015

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AHRQ = Agency for Healthcare and Research Quality

CC = Complication or Comorbidity

CCR = Cost-to-Charge Ratio

CDM = Charge Description Master

CMS = Centers for Medicare and Medicaid Services

CV = Coefficient of Variation

DRG = Diagnosis Related Group

GDP = Gross Domestic Product

HCUP = Healthcare Cost and Utilization Project

ICD-9 = The International Classification of Diseases, Ninth Revision

IPPS = Inpatient Prospective Payment System

IQR = Interquartile Range

MCC = Major Complication or Comorbidity

MS-DRG = Medical Severity Diagnosis Related Group

NIS = Nationwide Inpatient Sample

PPACA = Patient Protection and Affordable Care Act

THA = Total Hip Arthroplasty

TJA = Total Joint Arthroplasty

TLIF = Transforaminal Lumbar Interbody Fusion

LIST OF FIGURES

LIST OF TABLES

Page

ABSTRACT OF THE THESIS

Investigating Cost Variation Within Spinal Fusion Payment Groups

By

David Wright

Master of Science in Biomedical and Translational Science

University of California, Irvine, 2015

Assistant Clinical Professor Samuel Bederman, Chair

Background: Medicare reimbursement to hospitals for spinal fusion surgery is provided as a fixed payment for each admission based on Diagnosis Related Group (DRG). This system is predicated on the assumption that patients can be grouped into relatively homogenous units of resource use such that a single payment will adequately cover the costs of inpatient hospitalization for most patients within a given DRG. However, a previous study in total joint arthroplasty (TJA) showed that variation within DRGs can lead to differences between hospital costs and Medicare reimbursement, resulting in predictable financial losses to hospitals and hindering access to care for some patients. No study to our knowledge has investigated cost variation within current spinal fusion DRGs.

Aims: The aims of this thesis are to investigate cost variation within spinal fusion DRGs to determine whether variation within these groups meets an acceptable standard and to propose alternative spinal fusion DRGs, defined according to surgical invasiveness, that attempt to decrease the cost variation within current spinal fusion DRGs.

Methods: Direct hospital costs were obtained from the 2011 Nationwide Inpatient Sample (NIS) using cost-to-charge ratios. Our primary outcome was the coefficient of variation (CV), defined as the ratio of the standard deviation (SD) to the mean (CV=SD/mean x 100), for all hospital costs within a given DRG. CVs for spinal fusion DRGs (453-460) were compared to an established benchmark of TJA "DRG 209" (aggregate of DRGs [466-470]) to determine if cost variation within current spinal fusion DRGs is, in fact, "acceptable." We then modified a previously validated measure of surgical invasiveness and used it to re-categorize patients into new spinal fusion DRGs. Finally, we calculated the CV for these new DRGs to determine if this re-categorization resulted in a decrease in cost variation relative to existing spinal fusion DRGs.

Results: CV for costs within spinal fusion DRGs ranged from 44.16 for DRG 460 (spinal fusion except cervical w/o CC/MCC) to 52.6 for DRG 456 (spinal fusion except cervical w spinal curvature/malignancy/infection or 9+ fusion levels with MCC). The benchmark group, DRG 209, was found to have a CV of 38.2. The CVs for costs within spinal fusion DRGs were all significantly higher than the benchmark group ($p < 0.0001$). Re-categorizing patients into new DRGs according to surgical invasiveness resulted in a decrease in CV for costs within most groups.

Discussion: Our findings demonstrate that cost variation within current spinal fusion DRGs is unacceptably high. As in TJA, this variation may be leading to differences between costs and reimbursement that places undue financial burden on some hospitals and potentially compromises access to care for some patients. Re-categorizing patients into new DRGs according to surgical invasiveness results in decreased cost variation and may offer an alternative strategy for defining spinal fusion payment groups that ensures more equitable hospital reimbursement and improved patient access to care.

INTRODUCTION

Spinal fusion is one of the most common procedures performed by spine surgeons in the Untied States and is an effective treatment option for a wide range of spinal conditions (1-8). Since the 1990's, rates of spinal fusion procedures have increased substantially for a number of reasons including improved biomechanical understanding of the human spine, improved diagnostic imaging techniques, advances in spinal fixation device technology, and the overall increase in the life expectancy of the population (9-14). Over a similar period, average hospital charges associated with spinal fusions have more than tripled, resulting in a significant impact on total U.S. healthcare costs (9).

Medicare reimbursement for spinal fusion surgery is provided as a fixed payment for each admission by assigning patients to a Diagnosis Related Group (DRG) based on the principal procedure performed, as well as patient comorbidities or complications encountered during the admission (15-16). This prospective payment system is predicated on the assumption that patients and procedures can be grouped into relatively homogenous units of resource use such that a single payment will adequately cover the costs of inpatient hospitalization for most patients within a given DRG (16). However, several factors, including procedural complexity and unique patient characteristics, may contribute to variation within DRGs that can lead to differences between hospital costs and Medicare reimbursements. Predictable financial losses to hospitals may result in disincentives for the provision of care, potentially leading to disparities in access for some patients.

A recent study of total hip arthroplasty (THA) demonstrated that procedural differences between primary and revision surgeries are one potential source of variation within payment groups that may not be adequately accounted for in current DRGs (17). Results from this study showed that mean total hospital cost, operative time, estimated blood loss, and length of stay were all significantly higher for revision THA than for primary THA, even though both procedures were reimbursed equally under DRG 209 (Major Joint/Limb Replacement or Reattachment Procedure of Lower Extremity). The cost variation within DRG 209, explained by procedural differences between primary and revision THA, raised concern for patient access to care as hospitals were deterred from performing revision procedures in an attempt to limit ongoing financial losses (18). Following this study, the Centers for Medicare and Medicaid Services (CMS) "split" DRG 209 into DRGs 544 (Major Joint Replacement or Reattachment of Lower Extremity) and 545 (Revision of Hip or Knee Replacement) in an attempt to create more homogenous payment groups and allow for more equitable reimbursement, thus establishing an effective benchmark for "unacceptable" cost variation (19-20). In 2007, these payment groups were further refined into the DRGs 466 through 470 to accommodate for patient-specific major complication/comorbidity (MCC) and simple complication/comorbidity (CC) designations (see Table 1 for details) (21).

A recent simulation model of bundled payments in spine surgery suggested that there might also be a high degree of cost variation within current spinal fusion DRGs (22). No study to our knowledge has examined current spinal fusion DRGs to determine if there may, in fact, be an unacceptable degree of cost variation within each group. As in total joint arthroplasty (TJA), a high degree of cost variation within spinal fusion DRGs, if present, may lead to discrepancies between hospital costs and Medicare reimbursement that places undue financial burden on some hospitals and potentially compromises access to care for some patients.

The purpose of this research is to investigate cost variation within spinal fusion payment groups to ensure that variation within payment groups meets an acceptable standard. The first aim of this thesis is to establish whether the cost variation within current spinal fusion DRGs is acceptable, using TJA as a benchmark for "unacceptable" cost variation. We expect that cost variation within current spinal fusion DRGs is, in fact, unacceptable when compared to the established TJA benchmark. The second aim of this thesis is to explore alternative spinal fusion payment groups in an effort to decrease the cost variation observed within current spinal fusion DRGs. Defining payment groups according to procedural complexity may decrease cost variation and create more homogenous payment groups that could serve as an alternative to existing spinal fusion DRGs.

This thesis is organized into 4 chapters. Chapter 1 includes as a general overview of hospital and physician reimbursement in the United States, a conceptual framework for healthcare utilization, and a review of relevant literature. Chapter 2 addresses the first aim of this thesis, in which we seek to evaluate cost variation within current spinal fusion payment groups. Chapter 3 proposes new spinal fusion payment groups that attempt to improve cost variation relative to current DRGs. Chapter 4 includes a discussion of our findings, strengths and limitations of this thesis, and areas of future research.

Table 1. Chronological evolution of total joint arthroplasty DRGs.

DRG	Title	Years in use
209	Maj Joint/Limb Reattach Procs of Low Extremity	Prior to 2004
544	Major Joint Replacement or reattachment of lower extremity	2005 thru 2007
545	Revision of Hip Or Knee Replacement	2005 thru 2007
466	Revision of hip or knee replacement with MCC	2008 thru present
467	Revision of hip or knee replacement with CC	2008 thru present
468	Revision of hip or knee replacement without MCC or CC	2008 thru present
469	Major joint replacement or reattachment of lower extremity with MCC	2008 thru present
470	Major joint replacement or reattachment of lower extremity without MCC	2008 thru present

Prior to 2005, all total joint arthroplasty procedures were previously categorized and reimbursed an equal amount under DRG 209, regardless of procedural complexity (primary vs. revision) or patient comorbid factors. Between 2005 and 2007, the CMS modified the total joint arthroplasty coding to account for procedural factors, distinguishing between revision and primary procedures. In 2008, they also adjusted for patient comorbid conditions and post-operative complications by dividing DRGs 544 and 545 into the current DRGs, 466-470. These DRGs divide patients into payment groups based on whether the procedure was a primary or revision surgery and whether there were major complications or comorbidities (MCC), complications or comorbidities (CC), or no complications or comorbidities (without MCC or CC).

CHAPTER 1

BACKGROUND

Rates and Costs of Spinal Fusion Surgery

Spinal fusion is one of the most common procedures performed by spine surgeons and is used in the management of a variety of spinal disorders including scoliosis, lumbar stenosis, spinal tumors, vertebral fractures, and degenerative disk disease (1-8). In recent years, the rate of spinal fusion procedures in the United States has increased as much as 15-fold and was found to be 40% higher than in other developed countries (9-14). A variety of factors have been suggested that may have contributed to this increase including improved biomechanical understanding of the human spine, improved diagnostic imaging techniques, increased availability of spinal fixation devices, and the overall increase in life expectancy in the population (9). A recent study of the Nationwide Inpatient Sample demonstrated that rates of lumbar spinal fusion have seen the greatest increase in recent years, with a 2.7 fold increase from 1998 to 2008, followed by cervical fusions and thoracic fusions, respectively (9). During the same period, average hospital charges associated with all types of spinal fusion have more than tripled, from $24,676 in 1998 to $81,960 in 2008 (9). Combined, these factors have resulted in more than a 690% increase in the national aggregate hospital charges for spinal fusion surgery in the United States, from $4.3 billion in 1998 to more than $33.9 billion in 2008 (9). As far as hospital reimbursements that were actually paid out, Medicare alone spent over $2.5 billion on reimbursements for spinal fusion surgeries in 2011(23).

The rising costs of care are not unique to spinal fusion surgery. According to data from The World Bank, total healthcare expenditures in the United States rose from 13.1% of the gross domestic product (GDP) in 2000 to 17.1% of the GDP in 2013 (24). In contrast, countries such as France, Switzerland, and the United Kingdom spent approximately 11.7%, 11.5%, and 9.1% of their GDP on healthcare in 2013, respectively (24). In fact, the U.S. spends more on healthcare as a percentage of GDP than any other developed country in the world (24). In 2013, U.S. healthcare spending per capita averaged $9,146 compared to $4,864 and $3,598 per capita in France and the U.K., respectively (25). To give a different perspective, the U.S. spends approximately 3.8% of its GDP on military expenditures and 5.2% of its GDP on education (26-27).

Given the rising costs of healthcare, it is not surprising that there have been concerns about the affordability and access to healthcare for many people in the United States. While the Patient Protection and Affordable Care Act (PPACA) attempts to address these concerns by ensuring that all Americans are covered by health insurance that meets minimum essential health benefit requirements, the PPACA does relatively little to control the rising costs of care (28). According to a recent report by the CMS, by 2022, the PPACA is projected to reduce the number of uninsured by 30 million, but increase cumulative health spending by roughly $621 billion to an estimated 19.9% of the U.S. GDP (29). In addition, Medicare is projected to spend approximately $456 billion on inpatient hospital care alone in 2022, a 1.8 fold increase over 2012 spending (29).

Models for Healthcare Reimbursement

In an effort to control rising costs of healthcare, several payment schemes have been introduced in recent decades that either modify or provide an alternative to the traditional fee-for-service system that has dominated hospital and physician reimbursement for decades. These include reimbursement methods such as capitation, bundled payments, shared savings, and Medicare's Inpatient Prospective Payment System. These systems place an increased level of financial risk on hospitals and providers; the thought being that with more "skin in the game" providers will be incentivized to make more cost-effective and efficient use of healthcare resources. While an exhaustive review of hospital and physician reimbursement in the United States is beyond the scope of this thesis, it is important to have a general understanding of how some of these systems work, as well as the advantages and disadvantages of each system.

Fee-For-Service

Traditionally, healthcare services in the United States have been reimbursed using a fee-for-service model. Under this system, hospitals and providers are reimbursed separately for individual services provided. For example, hospital inpatient services such as lab tests, imaging, intravenous medication administration, intravenous fluid administration, and hospital supplies would appear as separate line-items on a hospital bill that is submitted to the patient or to the patient's health insurance provider. Physicians are then reimbursed separately for providing services such as routine office visits or procedures, such as posterior lumbar spinal fusions.

For example, consider a patient insured by Anthem Blue Cross who is scheduled for a lumbar fusion procedure. Anthem Blue Cross will usually provide one reimbursement to the physician for performing the procedure, and one reimbursement to the hospital determined by the individual services the patient received such as lab testing, imaging, intravenous fluids, and facility fees for use of the hospital bed and operating room (30). Most nursing staff and other allied health personnel such as physical and occupational therapists are employed by the hospital, so hospital reimbursement often covers the cost of paying these employees as well. Both hospitals and physicians negotiate annually with insurance providers to set the reimbursement rates for individual services and supplies.

Typically, hospitals set the “charge” for each service in a document called the charge description master (CDM), or chargemaster, which is a list of the hospital prices for each service performed in the hospital and for every supply item used during those procedures (31). Interestingly, except in California, hospitals are not required to post their chargemasters for public view (31). The chargemaster serves as the starting point for negotiations between hospitals and individual patients or insurance companies. An individual hospital might be paid by many distinct third-party payers (i.e. Anthem Blue Cross, Health Net, Aetna, etc.) each with its own distinct set of rules and levels of payment, which are negotiated separately with each private insurer once a year (31). It is important to realize that “charges” listed in the chargemaster are usually much higher than the reimbursement that hospitals actually receive. For example, in 2004, U.S. hospitals were actually paid only about 38 percent of their billed “charges” by patients and/or insurance (32). This highlights an important distinction between how much hospitals pay to provide supplies and services, how much hospitals charge for those supplies and services, and how much hospitals are actually reimbursed for those supplies and services. These concepts will be addressed in further detail in our definition of healthcare “costs” later in this thesis. For now, it is important to recognize that the term “cost” can have different meanings depending on the perspective of the parties involved.

There are several advantages and disadvantages to consider in the traditional fee-for-service model. From a patient and physician perspective, it imposes very few restrictions on the doctor-patient relationship and allows physicians to manage patient care without much concern for whether or not services will be reimbursed. However, this also highlights that the fee-for-service model poses very little financial risk to providers and offers few incentives to consider the financial impact of care decisions. In fact, it is often argued that, instead, the fee-for-service model offers incentives to provide the maximum number of services in order to receive the maximum reimbursement for patient care. Furthermore, the medical-legal consequences of missed diagnoses and poor patient outcomes might offer additional incentives to order more tests and services in an attempt to practice "defensive medicine." However, extra testing can also lead to an increased rate of false positive test results that themselves require further exploration, often in the form of more invasive tests and procedures that may be associated with significant morbidity.

By the late 1970s, the fee-for-service model had become the dominant means for financing both public and private medical care (33). As Bradford Gray, a research associate at the Urban Institute points out, "Third-party payers (both private and public) played their financing role passively, reluctant to interfere with medical decision-making and the doctor-patient relationship. They paid for medical care by reimbursing for costs incurred or charges billed by healthcare providers and did little to control which services were provided or how much they cost" (34). Echoing this, Dr. Rick Mayes, Ph.D. in the Department of Political Science at the University of Richmond in Virginia notes that "rampant medical inflation in the 1970's forced policy makers to search for ways to control Medicare's rapidly escalating costs" (33). However, it was not until the early 1980's that rising costs of healthcare and unchecked hospital and physician charges led to the first major changes in hospital and physician reimbursement.

Prospective Reimbursement

The first state to provide a promising alternative to the traditional fee-for-service model was New Jersey. In 1980, New Jersey introduced a new prospective payment model that assigned hospitalized patients to Diagnosis Related Groups (DRGs) according to the patient's primary diagnosis or procedure. Hospital reimbursement was then provided as a fixed, predetermined amount for each DRG (35). In one of the first articles outlining the conceptual framework for this new prospective payment system, William Dowling noted that "Prospective reimbursement…is a method of paying hospitals in which 1) amounts or rates of payment are established in advance for the coming year and 2) hospitals are paid these amounts or rates regardless of the costs they actually incur…[This] differs from retrospective cost reimbursement in that payment rates are specified in advance rather than determined after the fact and are not based on costs actually incurred…Providers face firm fixed prices for their services. If they are able to keep their costs below these prices, they will make a surplus; if not they will suffer a loss. Thus providers are definitely at risk" (36).

In 1984, only three years after New Jersey instituted this system, Medicare adopted a similar DRG-based reimbursement scheme known as the Inpatient Prospective Payment System (IPPS). As outlined by the CMS, "section 1886(d) of the Social Security Act sets forth a system of payment for the operating costs of acute care hospital inpatient stays under Medicare Part A (Hospital Insurance) based on prospectively set rates. Under the IPPS, each case is categorized into a DRG. Each DRG has a payment weight assigned to it, based on the average resources used to treat Medicare patients in that DRG" (37). DRG payments are calculated by multiplying the DRG weight by a "standardized amount," representing the average price per case for all

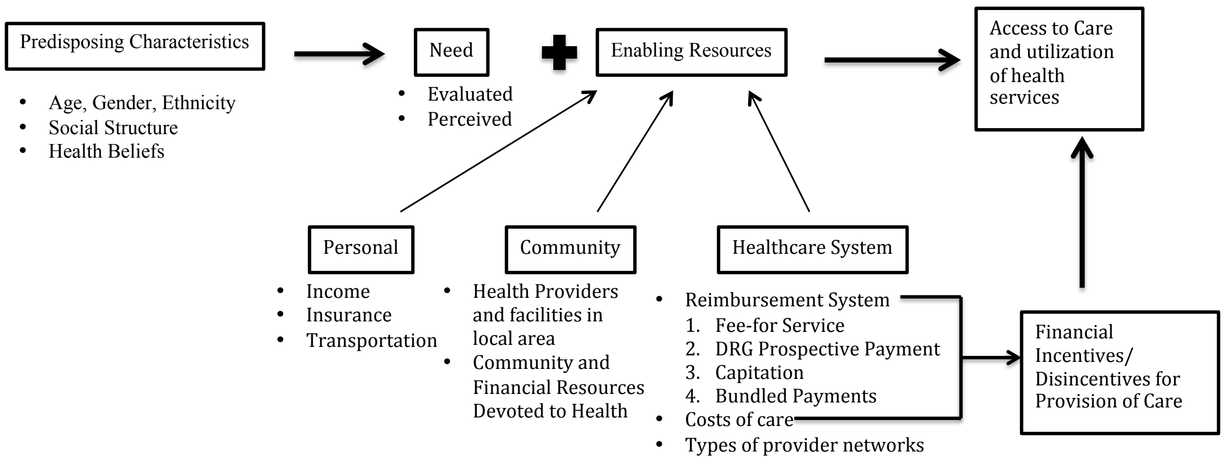
Medicare cases during the year. To calculate DRG weight, the average standardized charge for each DRG is calculated by removing the effects of regional area wage differences, indirect medical education costs, and additional payments to hospitals that treat a large percentage of low-income patients (referred to as "disproportionate share payments"). The average charge for each DRG is divided by the national average standardized charge-per-case to determine the weighting factor (38). Medicare adopted this system in 1984 despite a paucity of research examining how it had affected hospital and healthcare practices in New Jersey.

In 1986, a study of the New Jersey DRG-based payment system conducted by Hsiao et al. demonstrated that while hospital cost per admission was reduced (mostly due to decreased length of stay), there were no annual cost savings, largely due to an increase in the rate of hospital admissions (35). In addition, preliminary studies to determine the appropriate reimbursement rates showed wide variation across hospitals in reported cost for each DRG. However, it was difficult to determine how much of this variation was due to differences in cost by hospital geographic location (which accounts for differences in things such as cost of living, wage index, costs of supplies, and costs of transportation) and how much was due to patient heterogeneity within DRG payment groups (35). As William Hsiao, Professor of Economics at Harvard School of Public Health states, "DRGs were designed to be homogeneous units of hospital activity to which binding prices could be attached. The expectation of the DRG project managers in New Jersey was that there would be one rate established per DRG for all hospitals in the state. Studies conducted to determine the appropriate rates, however, showed wide variation in costs across hospitals for each DRG. It was impossible to determine how much of the variation reflected real cost differences and how much was due to DRGs classifying unlike patients together" (35). This last point introduces a critical concept. Medicare's IPPS is predicated on the assumption that patients and procedures can be grouped into relatively homogenous units of resource use such that a single payment will adequately cover the costs of inpatient hospitalization for most patients within a given DRG. However, as noted in the study by Hsiao et al., several factors contribute to cost variation within DRGs that can lead to important discrepancies between direct hospital costs and Medicare reimbursement. Hospitals make financial gains by treating patients for whom hospital costs are lower the fixed DRG reimbursement rate. Conversely, hospitals take financial losses when treating patients whose costs exceed the fixed DRG reimbursement rate. Predictable financial losses to hospitals may result in disincentives for the provision of care, potentially leading to disparities in access for some patients.

Patient Access to Care

The matter of patient access to care is complex, and differences between direct hospital costs and Medicare reimbursement is just one of many factors that can have an impact on a patient's ability to access care. Below is a conceptual framework that I have modified from an original framework proposed by sociologist Dr. Ronald Andersen (Figure 1). His original "Behavioral Model of Health Services Use," outlines a framework for understanding the complex issue of healthcare utilization and access to care (39). According to his model, access to and utilization of healthcare services is a function of a person's predisposition to use these services, a person's need for care, and factors that enable or impede their use (39).

Figure 1. Conceptual framework of factors affecting patient access to care.



This figure is a conceptual model that has been modified from Andersen's "Behavioral Model of Health Services Use" and outlines the factors that influence patient access to care and utilization of healthcare services. This thesis focuses on the way in which "hospital reimbursement" and "costs of care" act as enabling resources within the healthcare system. If costs of care exceed hospital reimbursement, there are financial disincentives for provision of care that can ultimately have a negative impact on patient's ability to access care.

Predisposing characteristics include unique patient factors, such as age and genetic characteristics, which predispose a person to being more or less likely to have a need for health services. Predisposing characteristics also include elements of social structure which determine a person's status in the community, their ability to cope with health problems, their ability to command resources to solve those problems, the environment in which they live, as well as their education, occupation, and ethnicity. Finally, personal and cultural health beliefs, such as whether or not someone trusts in western medical practices, can also affect patient's predisposition for utilizing health services (39).

Although these predisposing characteristics determine a patient's likelihood for need, patients may not seek care until a "need" actually arises. Here, we make a distinction between “perceived need” and “evaluated need.” “Perceived need” refers to a patient’s perception and understanding of their own health and needs for care. In contrast, “evaluated need” refers to physician or provider perception a patient’s health and their recommendations to the patient for healthcare services.

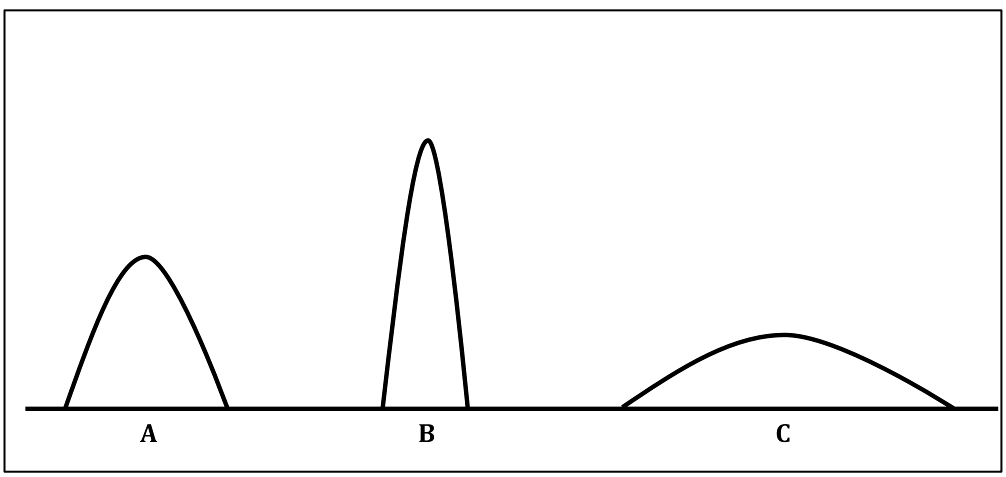
“Enabling resources” include factors that enable or impede use of health services. Personal enabling resources include things such as income, health insurance status, and means of transportation. Community enabling resources refer to the number of healthcare providers in the local community as well as whether or not there are any services or programs such as health education classes dedicated to health in the community. Finally, the organization and flow of resources within the healthcare system can also enable or impede the use of healthcare services. For example, even though a patient may have insurance, if providers in their community do not accept this insurance, then the patient is unlikely to utilize healthcare services in their community, or may have to travel greater distances to seek services where their insurance is accepted. Under the umbrella of enabling resources within the healthcare system, this thesis focuses on the way in which cost variation within current DRGs for spinal fusion surgery may lead to discrepancies between hospital costs and Medicare reimbursement that result in financial disincentives for provision of care to some patients, thereby limiting patient’s ability to access those resources. Andersen argues that “inequitable access occurs when social structure (e.g., ethnicity), health beliefs, and enabling resources determine who gets medical care” as opposed to provision of care purely based on “need” (39).

Cost Variation Within DRGs

To illustrate the way in which cost variation can impede access to care, consider the three curves presented below (Figure 2) representing differing degrees of cost variation for a single DRG "X," where direct hospital costs are on the x-axis and frequency of hospital admissions is on the y-axis. Assuming that all three distributions have the same mean cost, group B has the tightest distribution of costs, group C has the widest distribution of costs, and group A falls somewhere between groups B and C. Also recall that Medicare sets its prospective reimbursement rates for each DRG according to the average cost for patients in that group. For those patients whose costs fall above the DRG mean, the hospital takes a loss. Conversely, for patients whose costs fall below the DRG mean, the hospital will make a profit. Given these observations, group B represents the ideal cost distribution for DRG "X" because Medicare's fixed reimbursement will essentially cover the costs for the vast majority of patients admitted with that diagnosis. Even for the most costly patients in distribution B, the hospital experiences only minimal loss.

Now consider group C, which has a wide distribution of costs around the mean. In this case, the most costly patients within DRG "X" will result in significant financial losses for hospitals while the least costly patients will result in significant financial gains. In theory, if all hospitals treat equal numbers of high- and low-cost patients, these gains and losses should average out. However, the reality is that certain hospitals, such as academic centers, may have highly specialized facilities and providers that take referrals for more complex patients from surrounding hospitals. From a patient-care perspective, these referrals are entirely appropriate. However, the higher degree of patient and procedural complexity at these institutions may result in a higher proportion of high-cost patients within DRGs relative to low-cost patients. This imbalance may lead to significant financial losses for these tertiary care centers. Conversely, community hospitals that refer their most complex patients to these tertiary centers may realize significant financial gains from treating a greater proportion of low-cost patients within a DRG. From a payer perspective, while Medicare experiences some cost savings by underpaying for more costly patients, they experience a loss by over-paying hospitals for the least costly patients. Given these observations, it is easy to see how hospitals may be incentivized to seek out or "cherry-pick" those patients for whom costs are expected to fall below the DRG mean. Conversely, there is an incentive to try and identify those patients whose costs will likely fall above the DRG mean and refer them elsewhere or transfer them to another hospital for a "higher level of care." The wider the distribution of costs within a DRG, the greater these incentives become.

Figure 2. Potential cost distributions within a DRG.



These three curves represent potential cost distributions within a given DRG. Cost is represented along the x-axis and frequency of admissions on the y-axis. Assume that all three curves have the same mean cost. Group B has the lowest cost variation, group C has the greatest, and group A is somewhere in between. Medicare DRG reimbursements are based on the mean cost for a given DRG. This works well for group B, which has a very narrow distribution. However, if Medicare reimburses the mean for group C, a large portion of patients will have costs that fall well above the reimbursement rate while others will have costs that fall well below. This effect may incentivize hospitals to seek out or "cherry-pick" low-cost patients within the DRG while referring high-cost patients elsewhere.

As discussed above, the cost of an individual patient admission is influenced by a number of factors. Therefore, CMS has included some of these factors when calculating hospital DRG reimbursement rates. These factors include adjustments in wage index, indirect medical education costs, cost outliers, and disproportionate share payments. Also, in an attempt to better control for some of the patient specific factors that influence cost variation, CMS developed the Medical Severity-DRG (MS-DRG) classification system in 2007 (40). This system further divides most DRGs into 3 sub-groups based on whether the patient carried additional diagnoses that qualified as either major complications or comorbidities (MCC), simple complications or comorbidities (CC) or was without MCC or CC. For example, the DRG for “major chest procedures” was divided into DRG’s 163 (major chest procedures with MCC), 164 (major chest procedures with CC), and 165 (major chest procedures without MCC or CC) (40).

A recent study of total hip arthroplasty (THA) demonstrated that procedural differences are another potential source of cost variation within payment groups that may not be adequately accounted for in current DRGs. Results from this study showed that mean total hospital cost, operative time, estimated blood loss, and length of stay were all significantly higher for revision THA than for primary THA, even though both procedures were reimbursed equally under DRG 209 (Major Joint/Limb Reattachment Procedure of Lower Extremity) (17). This variation within DRG 209, explained by procedural differences between primary and revision procedures, raised concern for patient access to care as hospitals were deterred from performing revision procedures in an attempt to limit ongoing financial losses (18). Following this study, CMS performed a detailed review of claims data for DRG 209 and found that hospital charges for revision procedures were 20-30% higher than those for primary total joint replacement (20). The CMS review also found that although 70 percent of hospitals that performed primary TJA procedures also performed at least one revision TJA per year, only 5 percent of TJA hospitals performed at least 30 percent revisions (18). These findings supported previous concerns that certain hospitals (i.e. those performing more than 30% revision procedures) were incurring an undue financial burden due to appropriately high referral rates from other hospitals that did not specialize in revision procedures. Following these findings, CMS effectively 'split' DRG 209 into DRGs 544 (Major Joint Replacement or Reattachment of Lower Extremity) and 545 (Revision of Hip or Knee Replacement) in an attempt to create more homogenous units of resource use and allow for more equitable reimbursement (19-20). In the Federal Register published on August 12th 2005, CMS wrote, "To address the higher resource costs associated with hip and knee revisions relative to the initial joint replacement procedure, we proposed to delete DRG 209, create a proposed new DRG 544 (Major Joint Replacement or Reattachment of Lower Extremity), and create a proposed new DRG 545 (Revision of Hip or Knee Replacement)…We believe that the creation of the new DRGs for revisions of hip and knee replacements should resolve payment issues for hospitals that perform the more difficult revisions of joint replacements" (20). In 2007, CMS further refined DRGs 544 and 545 into DRGs 466 through 470 to accommodate for patient-specific major complication/comorbidity (MCC) and simple complication/comorbidity (CC) designations (see Table 1 for details).

Currently, thoracolumbar spinal fusion DRGs are already separated into three procedural categories that include "combined anterior/posterior spinal fusion," "spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels," and "spinal fusion except cervical" (41). However, a recent simulation model of bundled payments in spine surgery suggested that there might still be a high degree of cost variation within current spinal fusion DRGs, implying that these current procedural categories are inadequate (22). To date, no study has examined current spinal fusion DRGs to determine if there is, in fact, an acceptable degree of cost variation within each group. As in total joint arthroplasty (TJA), this variation, if present, may lead to discrepancies between hospital costs and Medicare reimbursement that places undue financial burden on some hospitals and potentially compromises patient access to care.

Aims of this Thesis

The first aim of this thesis is to establish whether the observed cost variation within current spinal fusion DRGs is acceptable. This will be accomplished by comparing the coefficient of variation (CV) for costs within each spinal fusion DRG to the CV for costs within a TJA benchmark DRG. The second aim of this thesis is to propose alternative spinal fusion payment groups, defined according to procedural complexity, that attempt to decrease the cost variation within current spinal fusion DRGs. The following chapters will address each of these aims in greater detail.

CHAPTER 2

Investigating Cost Variation within Current Spinal Fusion DRGs

Medicare reimbursement to hospitals is provided as a fixed payment for each admission based on Diagnosis Related Group (DRG). Although a recent simulation model of bundled payments in spine surgery found a high degree of cost variation within current spinal fusion DRGs, no study to our knowledge has determined whether this cost variation is acceptable (22). A study of TJA DRG 209 in 2005 demonstrated that procedural factors are one source of cost variation within DRGs that can cause differences between hospital costs and Medicare reimbursement (17). Differences in hospital costs between primary and revision TJA, representing two distinct procedural groups within DRG 209, raised concern for patient access to care as hospitals were deterred from performing revision procedures in an attempt to limit ongoing financial losses (18). Following this study, Medicare separated DRG 209 (which previously included all TJA procedures) into primary and revision TJA DRGs to reduce cost variation, establishing an effective "benchmark" for excessive variation. This chapter investigates cost variation within spinal fusion DRGs. We expect that cost variation within current spinal fusion DRGs is, in fact, unacceptable when compared to the established TJA benchmark.

METHODS

Data Source

This chapter describes a retrospective analysis of data from the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS) for 2011 (42). These data are publicly available through the Agency for Healthcare and Research Quality (AHRQ) and includes a 20% stratified sample of de-identified data from U.S. community hospitals. In addition to conventional community hospitals, the NIS also includes specialty hospitals such as obstetrics-gynecology, ear-nose-throat, orthopedic, and pediatric institutions, public hospitals and academic medical centers. Starting in 2005, the NIS also began to include long term acute care facilities. Excluded from the NIS are short-term rehabilitation hospitals, long-term non-acute care hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities.

Sampling is performed at the hospital level such that the data provided comprises 100% of patient discharges from each sampled hospital. Each year of data is sampled independently. The NIS provides information on approximately 8 million inpatient stays from about 1,000 hospitals each year (43). The NIS is commonly used in cost and utilization studies of spinal fusion surgery and is widely accepted as being a representative sample of U.S. hospitals (10,13, 44). The database includes demographic information, MS-DRG codes, hospital charges, lengths of stay, International Classification of Diseases Ninth Revision Clinical Modification (ICD-9-CM) diagnosis and procedure codes, as well as a system of unique hospital identifier codes (43). The decision was made to use national data rather than state, local, or institutional hospital data because previous studies in spinal fusion have demonstrated significant variation in cost by geographic region (11,12). Using national data allowed for better control of cost variation due to these geographic effects.

Nationwide Inpatient Sample Design

To better understand the statistical analysis that follows, as well as its strengths and limitations, it is important to understand the sampling design and weighting scheme of the NIS. The NIS is designed to reflect a hospital universe defined by AHRQ as "all hospitals open during any part of the calendar year and designated as community hospitals in the American Hospital

Association (AHA) Annual Survey of Hospitals" (51). To ensure that the 20% sample is representative of this universe, the NIS stratifies hospitals by census region (Northeast, North Central, South, and West) location/teaching status (urban or rural; Council of Teaching Hospital [COTH] membership or American Medical Association-approved residency program), bed size category (small, medium, large), and ownership (government nonfederal, private non-profit, private investor-owned). Within each stratum, a systematic random sample of hospitals is drawn that is equal to 20% of all hospitals in the universe for that stratum. For example, if there are 1000 hospitals in the universe for stratum "1," then a systematic random sample of 200 hospitals will be drawn from that stratum. Systematic random sampling involves ordering the sample prior to selection, randomly selecting a starting point for the sample, and sampling at regular intervals from that point. The "systematic" portion of the NIS design refers to the fact that hospitals within each stratum are sorted by zip code prior to sampling. This ensures a more even sampling of communities by location within each stratum. The NIS includes 100% of the discharges from these sampled hospitals (45).

As a result of this sampling design, discharge weights must be applied to the NIS in order to generate estimates that accurately reflect the hospital universe. The discharge weights are calculated according to the inverse sampling probability for discharges in each stratum (Discharge weight = 1/[sampled discharges in the stratum/total number of discharges in the stratum universe]). This results in most discharge weights being close to a value of 5, because the hospital sampling rate in each stratum is 20%, and all discharges from hospitals in each stratum are included, whenever possible. Therefore, each observation sampled in the NIS represents approximately 5 discharges in the universe for that stratum (45).

In 1988, the inaugural year of the NIS, only 8 states participated, making only 1,200 hospitals available for sampling. The 2011 NIS used for this study includes data from 46 states, making up a sampling frame of 5,132 hospitals from which 1,049 were actually sampled. The only states missing from the frame were Idaho, Alabama, Delaware, and New Hampshire (45). The result is a sample that, when weighted, provides estimates that reflect community hospital trends across the entire nation.

Patient Selection

Patients in this study were grouped according to MS-DRG, and included all patients assigned DRGs 453 to 460 for spinal fusion surgery and DRGs 466 to 470 for total TJA (see Table 2 for details). An aggregate of DRGs 466 through 470 were used to re-create "DRG 209" from 2005, which served as a historical benchmark for cost variation in this study.

As in previous studies of cost variation, we have excluded patients with outlier costs for index hospitalization (22). For all analyses, patients with costs below the 1st percentile and above the 99th percentile for each payment group were excluded. Given the de-identified nature of the NIS, limited patient demographic characteristics are available. However, age, sex, location of residence, and expected payer are available.

Table 2. List of DRGs included in this study.

DRG	Title
209†	Maj Joint/Limb Reattach Procs of Low Extremity
453	Combined anterior/posterior spinal fusion w MCC*
454	Combined anterior/posterior spinal fusion w CC**
455	Combined anterior/posterior spinal fusion w/o CC/MCC***
456	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w MCC
457	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w CC
458	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w/o CC/MCC
459	Spinal fusion except cervical w MCC
460	Spinal fusion except cervical w/o MCC
466	Revision of Hip or Knee Replacement w MCC
467	Revision of Hip or Knee Replacement w CC
468	Revision of Hip or Knee Replacement w/o CC/MCC
469	Major Joint Replacement or Reattachment of Lower Extremity w MCC
470	Major Joint Replacement or Reattachment of Lower Extremity w/o MCC

† 209 only prior to 2004. This DRG was reconstructed by combining all current total joint arthroplasty procedure DRGs back into one DRG.
*MCC = Major Complication or Comorbidity
** CC = Complication or Comorbidity
*** w/o CC/MCC = without complication or comorbidity/major complication or comorbidity

Defining "Cost"

Previous studies in spinal fusion surgery have demonstrated a wide range of definitions for the term "cost," therefore a brief discussion of the term "cost" is relevant, as it is defined in this study (46). The first consideration when calculating costs is whether costs are being determined from the perspective of the hospital, the payer, or the patient. This study defines costs from the hospital perspective, which includes direct costs, costs of staff, supplies, utilities, and rent, and excludes costs of the surgeon or anesthesiologist (46-47). "Direct" costs are defined as a costs related to the provision of inpatient or outpatient medical care (47). These are in contrast to the "indirect" costs to the patient related to missed time from work or decreased productivity due to the intervention (47).

A recent systematic review of cost-effectiveness research in spine surgery also showed wide variation in cost calculation methodology between studies (47). Among the 37 studies reviewed, 13 (35%) used Medicare reimbursements, 12 (32%) used a case-costing database, 3 (8%) used cost-to-charge ratios (CCRs), 2 (5%) used a combination of Medicare reimbursements and CCRs, 3 (8%) used the United Kingdom National Health Service reimbursement system, 2 (5%) used a Dutch reimbursement system, 1 (3%) used the United Kingdom Department of Health data, and 1 (3%) used the Tricare Military Reimbursement system (47).

Raw cost data in the NIS is provided as "total charges" for each admission. This represents the amount that hospitals billed for services, but does not reflect direct costs of hospital services or the specific amounts that hospitals received in payment. However, each year of NIS data contains hospital-specific CCRs that are based on all-payer inpatient costs for nearly every hospital in the NIS database. The AHRQ obtains cost information from the hospital accounting reports collected by the CMS, which are then used to calculate CCRs (48). This study uses CCRs provided by NIS to calculate the direct hospital costs associated with inpatient hospitalization for each patient in the study.

Analysis

Quantifying Cost Variation within DRGs

The primary outcome in this study was the coefficient of variation (CV), defined as the ratio of the standard deviation (SD) to the mean (CV=SD/mean x 100), for all direct hospital costs within each DRG. Other measures of spread including the standard deviation and interquartile range were also considered as possible primary outcomes. However, these measures are inherently tied to the absolute value of the mean or median cost. Those DRGs with higher absolute mean costs are more likely to have higher absolute values of standard deviation or interquartile range. Therefore, the CV was chosen as the primary outcome because it is a standardized measure of variation that can be compared between groups with very different absolute mean cost values.

CVs were compared to the established benchmark of TJA "DRG 209" (an aggregate of primary and revision DRGs 466-470) to determine if cost variation within current spinal fusion DRGs is, in fact, "acceptable." Weighted estimates of the population mean and standard deviation of cost for each DRG were calculated using the SAS version 9.3 SURVEYMEANS procedure to account the sample design described above. Although weighted estimate of the population mean is included in the SURVEYMEANS procedure, the population estimate of the standard deviation is not. Therefore, the population estimate of the standard deviation was calculated using methods outlined on the SAS website and described briefly here (49). In summary, the variable of interest is the standard deviation of a finite population. However, the NIS is a sample from a total population. The estimate of the standard deviation of the total population can be described by the expression

$$\hat{S} = \left(\frac{1}{\hat{N} - 1} \sum_{k=1}^{n} \frac{(y_k - \hat{\bar{y}})^2}{\pi_k} \right)^{\frac{1}{2}}$$

where $\hat{N} = \sum_{k=1}^{n} \frac{1}{\pi_k}$ is an estimator of the population total N, $\hat{\bar{y}} = \frac{1}{\hat{N}} \sum_{k=1}^{n} \frac{y_k}{\pi_k}$ is an estimator of the population mean, n is the number of elements in the sample, and π_k is the probability that element k is observed in the sample. In the case of the NIS, π_k is accounted for by the discharge weights provided in the data, which are equivalent to the inverse sampling probability for discharges in each stratum. To estimate $\hat{S}$, we first estimate both $\hat{N}$ and $\hat{\bar{y}}$ with PROC SURVEYMEANS. Next, we generate a variable "*z*" such that each observation z_k is equal to

$$z_k = \frac{1}{\hat{N}-1}(y_k - \hat{\bar{y}})^2 \qquad k = 1, \ldots, n$$

We then use the SURVEYMEANS procedure to estimate the total of "*z*" which allows us to account for the sampling design and discharge weights. This sum of z across all observations is equal to the finite population variance. It follows that the square root of the weighted population variance is equal to $\hat{S}$, the weighted estimate of the total population standard deviation.

In addition, because the sampling design includes 100% of discharges from all sampled hospitals, each patient does not represent a truly "independent" case, but is rather part of a cluster of patients discharged from that hospital. Without accounting for this clustering or "hospital fixed effect," there would be and artificial decrease in the variance associated with estimation of the population parameters, resulting in overly precise estimates (i.e. tighter 95% confidence intervals than would be seen if each patient were truly an "independent" observation). Therefore, the precision of the reported population estimates in this study accounts for the clustering in accordance with the HCUP special report "Calculating Nationwide Inpatient Sample Variances"(50). This is achieved by including the hospital ID under the PROC SURVEYMEANS "CLUSTER" procedure as well as a "RATE=0.20" option in the PROC SURVEYMEANS command line to account for a sampling rate of 20% in each hospital stratum.

After estimating the mean and standard deviation for each spinal fusion DRG, the CV was calculated and compared to the CV for TJA DRG 209. The rationale for the use of TJA DRG 209 as a benchmark for "unacceptable" variation stems from the fact that the cost differences observed within this group prior to 2005 were large enough that Centers for Medicare and Medicaid Services decided to change the DRG coding in order to avoid disparities in access for TJA patients requiring revision procedures. However, rather than quantifying the standard deviation of DRG 209 in 2005 and using this as a benchmark, we will instead “reconstruct” DRG 209 (which can be represented by an aggregate of all current TJA DRGs) from current data. This reconstructed DRG will serve as a more accurate control than the historical control from 2005 data since many trends in TJA are likely to have changed in the past 9 years. It is important to realize that the goal of this study is not to comment on the absolute value of the cost variation within spinal fusion DRGs, but rather establish the acceptability of this variation when compared to a reference group. Therefore, we will not control for regional cost variation that is caused by differences in costs of living, transportation, and wage index because we assume that both TJA and spinal fusion DRGs are subject to these factors to the same degree across the United States.

Bootstrapping was then applied to obtain a sampling distribution of estimates of the CV for each DRG. This allowed for the mean CV for each spinal fusion DRG to be compared to the mean CV for DRG 209 using a one-way ANOVA with a pre-planned Dunnett’s T3 correction for multiple comparisons. The Dunnett’s T3 is a multiple comparison procedure designed specifically to correct for comparison of several treatment groups to a single control group. In this study, DRG 209 will serve as the “control” group, with spinal fusion DRGs 453-460 serving as the “experimental” groups. Bootstrapping was performed to draw 100 simple random samples with replacement at a sampling rate of 50% from each DRG.

RESULTS

Patient characteristics are presented in Table 3. Of the spinal fusion DRGs in 2011, DRG 460 (Spinal fusion except cervical w/o MCC) had the greatest number of patient discharges with 194,477, followed by DRG 455 (Combined anterior/posterior spinal fusion w/o CC/MCC) with 15,005. There were 973,366 discharges in the benchmark group, DRG 209 (aggregate of primary and revision total joint arthroplasty DRGs [466-470]). Approximately 50% of patients in the "combined anterior/posterior spinal fusion" procedural category were between 45 and 64 years of age. Of those patients in the "spinal fusion except cervical" procedural category, 65-84 year-olds made up the greatest proportion of patients in DRG 459 (51%) while 45-64 year-olds made up the greatest proportion in DRG 460 (45%). Patients in the "spinal fusion except cervical w spinal curvature/malignancy/infection or 9+ fusion levels" category were younger, with up to 46% of patients in DRG 458 between 1-17 years of age. This is consistent with higher rates of scoliosis in this age group.

On average, there were a greater proportion of females (~58%) in all categories, and whites made up the large majority (~81%) of the patients across all DRGs. Patients in in central metropolitan and surrounding fringe counties combined for approximately 53% of discharges across all groups. The proportion of patients receiving spinal fusion surgery who had Medicare as their primary expected payer ranged from 22.81% in DRG 458 (spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w/o CC/MCC) to 58.41% in DRG 459 (spinal fusion except cervical with MCC). The other leading payer in spinal fusion DRGs was private insurance, including HMOs, which ranged from 28.06% in DRG 459 to 54.43% in DRG 458.

Table 3. Patient Characteristics

DRG Category	"Combined Anterior/Posterior Spinal Fusion"			"Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus"			"Spinal fusion except cervical"		"Aggregate Benchmark"
DRG (Severity Designation)	453 (MCC)	454 (CC)	455(No MCC/CC)	456 (MCC)	457 (CC)	458 (No MCC/CC)	459 (MCC)	460 (No MCC)	209 (466-470 combined)
Estimated Total Discharges	3,614	13,084	15,005	3,490	9,501	7,445	8,683	194,577	973,366
Age in yrs (mean, SD)	60 (15)	58 (14)	54 (14)	50 (25)	53 (23)	37 (26)	64 (14)	58 (14)	67 (12)
Age Category (%)									
(1) 1-17	2.55	0.91	0.83	25.03	18.69	46.45	0.63	0.49	0.05
(2) 18-44	9.45	14.11	22.15	8.54	9.24	10.91	8.74	18.43	2.73
(3) 45-64	44.97	50.20	52.39	26.93	30.86	19.89	35.83	45.92	40.09
(4) 65-84	41.94	34.20	24.03	37.84	39.29	21.66	51.46	34.08	50.97
(5)85 or more	1.09	0.54	0.6	1.54	1.87	0.78	3.35	1.07	6.11
Gender (F, %)	51.64	57.95	52.15	53.14	64.08	68.9	50.34	55.64	61.06
Race (%)*									
(1) White	79.30	83.24	83.05	76.94	81.91	76.08	80.45	82.27	83.8
(2) Black	9.90	6.30	5.25	7.9	7.38	10.06	9.27	7.33	7.78
(3) Hispanic	6.12	6.37	8.11	7.51	5.56	8.19	6.32	7	5.3
(4) Asian/PI	1.82	1.66	1.31	2.72	2.05	1.81	0.95	0.93	0.96
(5) Native American	0.65	0.26	0.35	0.65	0.36	0.54	0.37	0.26	0.27
(6) Other	2.21	2.17	1.92	4.27	2.73	3.31	2.63	2.2	1.9
Location of Patients Residence (%)**									
(1) Central Metro >= 1million population	37.15	33.82	30.42	31.56	27.87	25.37	30.89	25.68	24.58
(2) Fringe of Metro >= 1million Population	23.74	21.90	23.79	26.12	27.18	24.63	24.24	24.94	25.47
(3) Counties in Metro areas of 250K-999,999 pop	15.25	14.04	14.4	16.08	17.3	22.13	14.61	18.23	19.44
(4) Counties in Metro areas of 50,000-249,999 pop	8.73	10.35	11.52	10.28	10.88	9.18	9.68	10.82	10.52
(5) Micropolitan	8.86	10.49	10.53	8.46	9.54	11.62	11.14	12.3	11.7
(6) No metro or micro counties	6.27	9.40	9.33	7.5	7.24	7.06	9.43	8.03	8.3
Expected Payer (%)***									
(1) Medicare	49.88	40.07	29.17	45.9	43.84	22.81	58.41	39.2	56
(2) Medicaid	6.81	4.75	5	15.1	8.83	14.26	5.54	5.38	3.16
(3) Private including HMO	34.67	43.23	49.94	32.1	42.52	54.43	28.06	44.06	37.18
(4) Self-pay	2.07	1.15	0.84	2.26	0.85	1.05	1.7	0.93	0.52
(5) no charge	0.12	0.14	0.06	0.83	0.17	0.47	0.15	0.14	0.15
(6) other	6.45	10.66	14.99	3.8	3.8	6.97	6.13	10.28	2.99

DRGs presented by procedural category "combined anterior/posterior spinal fusion," "spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels," "spinal fusion except cervical," and "aggregate benchmark" which is equivalent to "major joint/limb reattachment procedure of lower extremity" DRG 209 from 2005.

Coefficients of variation for cost within spinal fusion DRGs ranged from 44.16 (95% CI, 44.00-44.31) for DRG 460 (spinal fusion except cervical w/o CC/MCC) to 52.6 (95% CI, 51.01-54.05) for DRG 456 (spinal fus exc cerv w spinal curv/malig/infec or 9+ fus with MCC). The benchmark group, DRG 209, was found to have a CV of 38.2 (95% CI, 37.66-38.74). Bootstrap analysis demonstrated that when compared to this benchmark, all spinal fusion DRGs had significantly higher CVs (all p-values < 0.0001) using one-way ANOVA with pre-planned Dunnett's T3 correction for multiple comparisons. Population estimates of mean cost for spinal fusion DRGs ranged from $27,153 (95% CI, $26,230–$28,076) for DRG 460 to $77,965 (95% CI, $71,976–$83,954) for DRG 456. The estimated mean cost for DRG 209 benchmark was $15,903 (95% CI, $15,495–$16,311). Population estimates of standard deviation ranged from $11,992 (95%CI, $11,542–$12,441) for DRG 460 to $41,044 (95% CI, $36,713–$45,375). Results are presented in Table 4.

Table 4. Cost variation within current spinal fusion DRGs.

DRG	Description	N*	Mean Cost†		SD Cost†¥		CV†† (95% CI)		
			Mean ($)	95% CI	SD ($)	(95% CI)	CV	(95% CI)	P-value**
209	Maj Joint/Limb Reattach Procs of Low Extremity	973,366	$15,903.00	($15,495–$16,311)	$6,077	($5,835–$6,318)	38.2	(37.66-38.74)	ref.
453	Combined anterior/posterior spinal fusion w MCC	3,614	$75,435.00	($70,622–$80,248)	$37,426	($33,086–$41,765)	49.6	(46.85-52.05)	<0.0001
454	Combined anterior/posterior spinal fusion w CC	13,084	$52,745.00	($49,754–$55,736)	$25,233	($23,081–$27,385)	47.8	(46.39-49.13)	<0.0001
455	Combined anterior/posterior spinal fusion w/o CC/MCC	15,005	$39,271.00	($36,880–$41,662)	$17,660	($16,060–$19,260)	45.0	(43.55-46.23)	<0.0001
456	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w MCC	3,490	$77,965.00	($71,976–$83,954)	$41,044	($36,713–$45,375)	52.6	(51.01-54.05)	<0.0001
457	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w CC	9,501	$52,297.00	($48,581–$56,013)	$25,806	($23,089–$28,523)	49.3	(47.53-50.92)	<0.0001
458	Spinal fus exc cerv w spinal curv/malig/infec or 9+ fus w/o CC/MCC	7,445	$43,324.00	($39,543–$47,105)	$21,834	($19,028–$24,639)	50.4	(48.12-52.31)	<0.0001
459	Spinal fusion except cervical w MCC	8,683	$44,635.00	($42,544–$46,726)	$23,279	($22,005–$24,552)	52.2	(51.72-52.55)	<0.0001
460	Spinal fusion except cervical w/o MCC	194,577	$27,153.00	($26,230–$28,076)	$11,992	($11,542–$12,441)	44.2	(44.00-44.31)	<0.0001

*Estimated total discharges
**P-values calculated using one-way ANOVA with Dunnett's T3 correction for multiple comparisons
†Weighted national estimate of mean cost
†¥ Weighted national estimate of standard deviation.
††CV = coefficient of variation = (standard deviation/mean *100)

A box and whisker plot demonstrating the interquartile range (IQR) of costs for each DRG is presented in Figure 3. When compared to the DRG 209 benchmark, all spinal fusion DRGs demonstrate a higher IQR. DRG 456 demonstrates the largest IQR ($49,976) while DRG

460 demonstrates the smallest ($14,898). Meanwhile, the IQR for DRG 209 is $6,680. In addition, 1st and 99th percentiles for costs varied widely among spinal fusion DRGs. The 1st percentile of costs for DRG 456 was $20,566 while the 99th percentile was just over $249,723 (difference of $229,157). Again, amongst spinal fusion DRGs, DRG 460 had the smallest difference between 1st and 99th percentiles with a 1st percentile of $6,909 and a 99th percentile of $76,087 (difference of $69,177). In contrast, the 1st percentile of costs for DRG 209 was $6,745 while the 99th percentile was $46,606 (difference of $39,860). It is important to note that these figures are included for descriptive purposes only. Given that the mean costs of each DRG are different across DRGs, it is not entirely accurate to make comparisons of cost variations in terms of absolute values of dollar amounts of IQR. Again, it was for this reason that the CV was chosen as the primary outcome of this study because it represents a standardized, unit-less metric of variation that allows for accurate comparison of variation between groups with different means.

Figure 3. Interquartile range of hospital costs within DRGs.

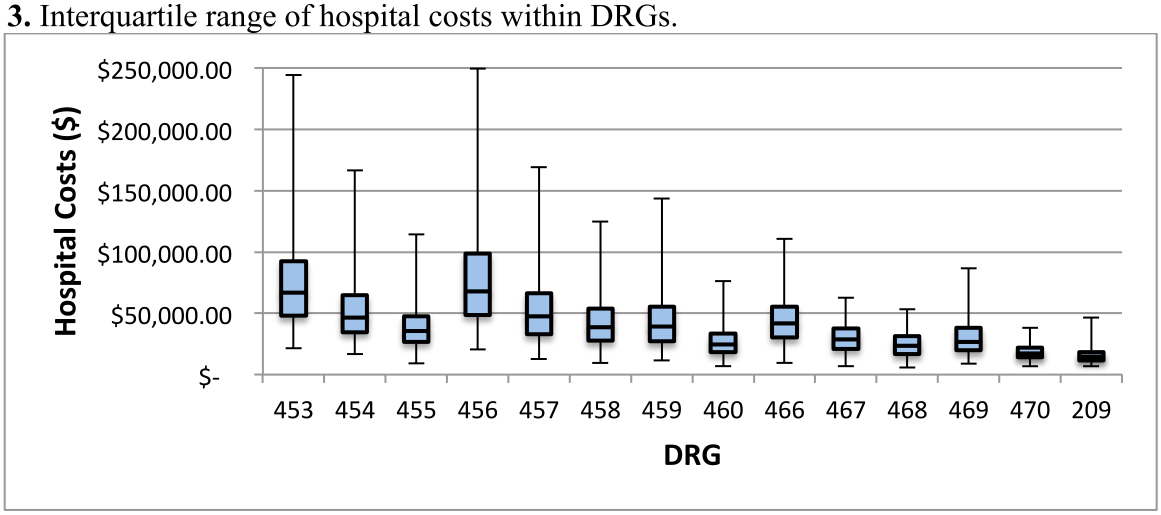
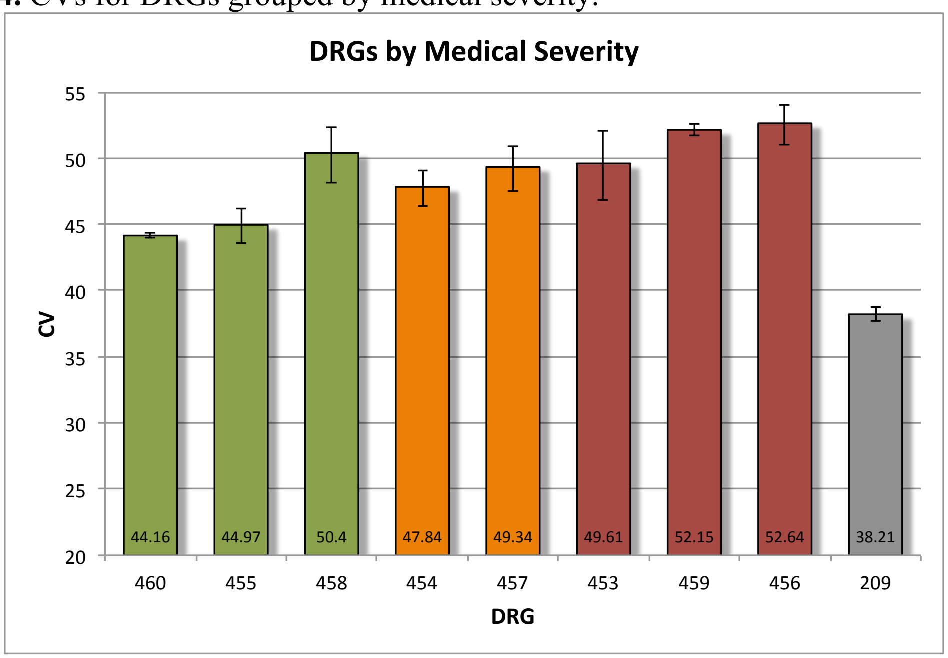


Figure 3 demonstrates the interquartile range (Q1-Q3) for hospital costs (in U.S. dollars) for each DRG. Whiskers demonstrate the 1st and 99th percentile. The median cost for each DRG is included for reference and is represented by the black horizontal line within each box. Procedural categories are "Spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels" (DRGs 458, 457, and 456); "combined anterior/posterior spinal fusion" (DRGs 455, 454, and 453); and "spinal fusion except cervical" (DRGs 459 and 460).

CVs for DRGs are grouped according to medical severity category in Figure 4. Within each medical severity category, the most procedurally complex category "Spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels" also has the highest CV. CVs for this procedural category are 50.40, 49.34, and 52.64 in the "without MCC or CC," "CC," and "MCC" severity categories, respectively. The second most complex category, "combined anterior/posterior spinal fusion," had the second highest CV in the "without CC/MCC" and "CC" groups (44.97 and 47.84, respectively). Meanwhile the "spinal fusion except cervical" procedural category (which does not currently have a CC group) had the lowest CV of the DRGs "without CC/MCC" (44.16). These results suggest that CV increases with increasing surgical complexity.

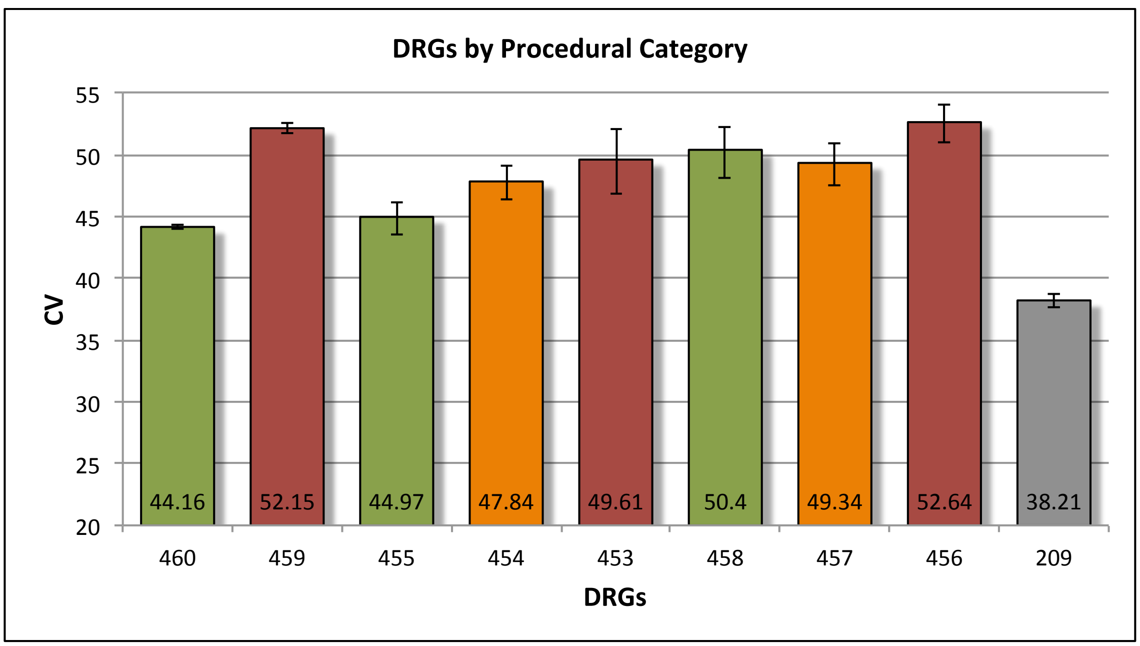
Figure 4. CVs for DRGs grouped by medical severity.



*DRGs grouped by medical severity. "Without complications or comorbidities/major complications or comorbidities" highlighted in green, "complications and comorbidities" highlighted in yellow, and "major complications or comorbidities" highlighted in red. Procedural categories are "Spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels" (DRGs 458, 457, and 456); "combined anterior/posterior spinal fusion" (DRGs 455, 454, and 453); and "spinal fusion except cervical" (DRGs 459 and 460).

DRGs are grouped by procedural category in Figure 5. Within each procedural category ("spinal fusion except cervical with spinal curv/malignancy/infection or 9+ fusion levels," "combined anterior/posterior spinal fusion," and "spinal fusion except cervical") the CVs are highest amongst those DRGs with "MCC" designation, followed by those with "CC" designation, followed by those "without CC or MCC." The only exception to this trend is for DRG 458 ("without CC/MCC"), which has a slightly higher CV than the corresponding DRG (457) in the "CC" group. However, the CVs for these DRGs are almost identical (DRG 458 = 50.40, DRG 457 = 49.34). In addition, their 95% confidence intervals overlap (DRG 458 95%CI [48.12-52.31] and DRG 457 95% CI [47.53-50.92]) making it difficult to distinguish these two groups.

Figure 5. CV for DRGs by procedural category.



*DRGs grouped by procedural category. "without MCC or CC" highlighted in green, "CC" highlighted in yellow, and "MCC" highlighted in red. Procedural categories are "Spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels" (DRGs 458, 457, and 456); "combined anterior/posterior spinal fusion" (DRGs 455, 454, and 453); and "spinal fusion except cervical" (DRGs 459 and 460). Note that "spinal fusion except cervical" group does not currently contain a CC severity category in the CMS coding manual.

DISCUSSION

In this study of the NIS, cost variation within each spinal fusion DRG was found to be unacceptably high when compared to cost variation within the established benchmark group, DRG 209 ($p<0.0001$). Within each procedural category ("spinal fusion except cervical with spinal curv/malignancy/infection or 9+ fusion levels," "combined anterior/posterior spinal fusion," and "spinal fusion except cervical") the CV is highest amongst those DRGs with major complications or comorbidities (MCC), followed by those with simple complications and comorbidities (CC), followed by those without CC or MCC. This phenomenon suggests that patient-specific factors make an important contribution to cost variation within DRGs and should continue to be taken into consideration in future DRG reimbursements. It is worth noting that the exception to this trend is for DRGs 458 in the "without CC/MCC" group, which has an almost identical CV to the corresponding DRG 457 in the "CC" group (DRG 458 = 50.40, 95%CI [48.12-52.31]; DRG 457 = 49.34 [47.53-50.92]). Interestingly, this exception occurs in the most procedurally complex category, suggesting that when procedural complexity is high, it may make a larger contribution to the overall cost variation, thus dampening the effects of patient-specific complications and comorbidities on cost variation relative to procedurally less complex DRGs such as 460 and 459. Consistent with this explanation, the largest difference in CVs for any two DRGs within a procedural category is between DRGs 459 (with MCC) and DRG 460 (without MCC or CC), both of which represent the least procedurally complex category. In this category, the patient-specific factors such as complications and comorbidities likely play a larger role in the observed cost variation within each DRG.

Independent of patient-specific factors, procedural factors also appear to influence the degree of cost variation within groups. Of the DRGs with the MCC medical severity designation

(DRGs 453, 456, and 459), the most procedurally complex category ("spinal fusion except cervical with spinal curv/malignancy/infection or 9+ fusion levels") also had the highest degree of cost variation (49.6, 52.64, and 52.15, for DRGs 453, 456, and 459 respectively). This phenomenon persists in the CC groups (DRGs 454 and 457) and in the without MCC or CC groups (455, 458, and 460). Interestingly, the second most complex category, "combined anterior/posterior spinal fusion," had the second highest CV in the without CC/MCC group, but had the lowest CV of those DRGs with the MCC designation. Meanwhile the "spinal fusion except cervical" procedural category (which does not currently have a CC group) had the second highest CV of those DRGs with the MCC designation and the lowest CV of the DRGs without CC/MCC. The wide difference in the CVs between DRGs in the "spinal fusion except cervical" procedural category may be explained by medical severity (MCC vs. without MCC), but could also be explained by the fact that even seemingly simple procedures in this category, such as a 1-level fusion, can have a wide range of outcomes based on procedural factors that are not captured in current DRGs. For example, in a study of patients undergoing transforaminal lumbar interbody fusion (TLIF) for single-level lumbar degeneration, hospital length of stay varied from 5.8 days for patients who received a minimally invasive TLIF to 10.2 days for those who received a traditional open TLIF (51). The increased length of stay associated with a traditional open TLIF would significantly increase total hospital costs. However, the current DRGs for single-level posterior fusion (DRG 459 or 460) make no distinction between these surgical approaches.

Our results suggest that both patient and procedural factors play an important role in cost variation within DRGs. However, this thesis specifically focuses on the way in which procedural factors contribute to cost variation. In fact, we assume that since patient medical severity designations (“MCC,” “CC,” and “without MCC/CC”) are used across all DRGs, regardless of procedure, the effects of these designations on cost variation should remain constant across DRGs. However, it is possible that current medical severity designations do not adequately capture important patient-specific factors that also contribute to cost variation within DRGs. More research is need to determine whether refining these medical severity designations may result in more homogenous payment groups across all DRGs.

A recent article on variation in Medicare payments for episodes of spinal surgery showed that a significant proportion of the difference in reimbursement between the highest reimbursement quintile and lowest reimbursement quintile was due to differences in procedural choices, including the use of fusion (52). In the study, Medicare claims data were used to investigate differences in Medicare reimbursement for patients that received spinal decompression or fusion, identified using ICD-9 codes. The differences in procedural choices highlighted in this article likely stem from the fact that many spinal conditions have multiple treatments with reported success (1-5). In fact, a recent study of spinal surgery in the United States showed wide variation in treatment recommendations when surgeons were asked to evaluate 5 standardized cases of common spinal disorders, suggesting that patients may be offered a range of surgical options for a given set of symptoms (53). Other studies have demonstrated significant variation in rates of spinal surgery across different regions within the Untied States, which is further thought to reflect a lack of consensus on indications for surgery (9,11,12). This lack of consensus may be because, in contrast to more standardized procedures such as hip and knee replacement, treatment of spinal pathology is highly individualized based on the degree of deformity, severity of degeneration, and presence or absence of neurologic symptoms. The result is a high degree of procedural variation that likely makes a significant contribution to the cost variation observed within DRGs in this thesis.

In summary, the results in this chapter demonstrate that current cost variation within spinal fusion DRGs is unacceptably high. Both patient medical severity (MCC, CC, or without MCC/CC) and procedural factors (approach, invasiveness, complexity) appear to contribute to the high degree of cost variation within groups. As previously demonstrated in TJA DRG 209, this variation may be leading to differences between hospital costs and reimbursement that places undue burden on some hospitals and potentially compromises access to care for some patients (18). This chapter highlights the need for future research to identify potential changes to current spinal fusion DRGs that attempt to group patients into more homogenous units of resource use such that a single fixed payment will more adequately cover the costs of hospitalization for these patients. There are several possible ways this might be achieved. Although patient-specific factors clearly contribute to cost variation, it may be more difficult to adjust or re-categorize current DRGs based on these factors because they are intrinsic and unique to each patient. However, procedural factors such as surgical approach, invasiveness, and complexity are likely sources of cost variation within spinal fusion DRGs that may be more easily adjusted for.

It could be argued that procedural factors are already accounted for in current DRGs, given that they are divided into procedural categories (spinal fusion except cervical, combined anterior/posterior fusion, and spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels). However, this study demonstrates that these current categories do not define homogenous units of resource use, suggesting that revision is necessary. As demonstrated by the example of traditional open versus minimally invasive single-level fusion, much of the variation in surgical methods for standardized procedures are not captured in current DRGs. Unfortunately, unlike TJA DRG 209, which could be conveniently re-categorized into separate DRGs based on clearly defined “primary” and “revision” procedural factors, it may not be feasible to re-categorize spinal fusion DRGs in a similar fashion given the multitude of different procedures available in this field. In future studies, it may be more appropriate to explore other procedural measures, such as the degree of surgical invasiveness, which may more accurately define homogenous payment groups. Doing so will allow for more equitable hospital reimbursement and improved patient access to care. The following chapter explores surgical invasiveness as one potential measure that may be used to achieve this goal.

CHAPTER 3

Defining New Spinal Fusion Payment Groups

In the previous chapter we quantified cost variation in current spinal fusion DRGs and determined that this variation was unacceptable when compared to the previous established benchmark of TJA. However, unlike TJA, in which a clear procedural dichotomy between "primary" and "revision" procedures allowed for *a priori* re-categorization of patients within DRG 209 into new payment groups, spinal fusion surgery is more procedurally complex, making *a priori* re-categorization much more difficult. Therefore, in this chapter we propose a modification to a measure of surgical invasiveness in spinal fusion surgery that is then used to define new payment groups according to surgical complexity.

Historically, outcomes associated with surgery such as intraoperative blood loss, duration of surgery, length of hospitalization, or adverse events have been used as measures of invasiveness (53-55). However, these outcomes do not account for procedural factors that may more directly measure the intrinsic "invasiveness" of the procedure itself. In an effort to better account for these factors in a measure of "invasiveness," Mirza et al. recently developed and validated the "Spine Surgery Invasiveness Index" (56) It assigns one point for each vertebral level that is decompressed, fused, or instrumented, anteriorly and posteriorly during a procedure and is scored using the surgeon's dictated operative report. This measure was validated in spinal surgery patients using multivariable regression modeling to demonstrate that index scores correlated well with traditional measures of invasiveness including intra-operative blood loss (adjusted $r^2 = 0.54$, $p < 0.0001$) and operative time (adjusted $r^2 = .53$, $p < 0.0001$) (56). Not surprisingly, inter-rater reliability was also found to be high between scorers, with an intraclass correlation coefficient of 0.99 for the index as a whole (57).

Quantifying surgical invasiveness may offer a means of homogenizing current DRGs according to surgical complexity. More homogenous groups would allow Medicare reimbursement to more accurately reflect hospital costs, ensuring more equitable access for all patients. We expect to find that new spinal fusion payment groups, defined according to procedural complexity, will have decreased cost variation compared to existing spinal fusion payment groups.

METHODS

Developing a Measure of Surgical Invasiveness

The "Spine Surgery Invasiveness Index" developed by Mirza et al. has previously been shown to be a valid measure of invasiveness in spinal surgery patients and is scored using the attending physician's operative report (56-57). Therefore, we aimed to develop a measure using ICD-9 codes that captured the same procedural factors used in the existing index.

The "Spine Surgery Invasiveness Index" is based on three categories of interventions commonly performed in spinal procedures: decompression, fusion, and instrumentation. These categories are combined with surgical approach (anterior or posterior) to arrive at a total of six procedural categories (anterior decompression, anterior fusion, anterior instrumentation, posterior decompression, posterior fusion, posterior instrumentation). Each category is then assigned a score of one for each vertebral level at which it is performed (47). The procedural definitions for each category in the "Spine Surgery Invasiveness Index" are as follows:

1. Anterior decompression score (ad): the number vertebrae requiring partial or complete excision of the vertebral body (regardless of surgical approach or location of skin incision), or the disc caudal to that vertebra if the disc is excised from an anterior approach.

2. Anterior fusion score (af): the number vertebrae that have graft material attached to or replacing the vertebral body, regardless of the surgical approach.
3. Anterior instrumentation score (ai): the number of vertebrae that have screws, plate, cage, or structural graft attached to the vertebral body or replacing the vertebral body, regardless of the surgical approach.
4. Posterior decompression score (pd): the number of vertebrae requiring laminectomy or foraminotomy at the foramens caudal to their pedicles, and/or discectomy at the disc caudal to the vertebral body if the disc is excised from a posterior approach.
5. Posterior fusion score (pf): the number of vertebrae that have graft material on their lamina, facets, or transverse processes.
6. Posterior instrumentation score (pi): the number of vertebrae that have screws, hooks, or wires attached to their pedicles, facets, lamina, or transverse processes (47).

While this instrument has been shown to be reliable and valid for measuring invasiveness in spine surgery, it has several drawbacks. First, scoring is not only time-intensive, but requires accurate interpretation of the patient's operative report in order to essentially "code" each procedural factor into the overall score. Furthermore, use of this score for research or quality purposes is cumbersome because it requires access to patient-level data. Such patient-level data is absent from hospital and administrative databases that are commonly used to examine national trends and quality in spinal surgery.

In contrast to operative reports, ICD-9 procedure codes are available for most patient discharges at both the hospital and national level and are commonly presented in de-identified datasets. While not as granular as operative reports, ICD-9 procedure codes for spinal fusion surgery capture many of the same procedural details used in the current invasiveness index. For example, one of the cases presented in the validation study of the "Spine Surgery Invasiveness Index" was a "L4–S1 posterolateral fusion with pedicle screws, and no decompression." This same procedure would be coded using ICD-9 codes 81.07 (lumbar and lumbosacral fusion, lateral transverse process technique, including instrumentation) plus 81.62 (fusion or re-fusion of 2-3 vertebrae). In this example, the type of fusion (lumbosacral), approach (lateral transverse process technique is a type of posterior approach), instrumentation, and number of levels fused are captured by these two ICD-9 codes. We therefore aimed to develop a measure of surgical invasiveness using ICD-9 codes that would capture the same factors currently captured in the "Spine Surgery Invasiveness," thereby allowing us to categorize patients into groups of differing surgical complexity.

For the proposed "ICD-9 Invasiveness Index for Spinal Fusion Surgery," we began by meeting with a coding expert and spine surgeon to help assemble a list of all ICD-9 procedure codes related to spinal surgery. This resulted in a total of 91 items that could potentially be used in our measure. Item reduction began by assessing the face validity of the 91 ICD-9 procedure codes gathered as described above. An orthopedic spine surgeon assessed the face validity of each of the 91 items and then attempted to fit each item into one of the 6 existing procedural categories based on the definitions above. Items that did not fall under any of these categories, as judged by expert opinion, were discarded. At the end of the item reduction phase, there were 23 items remaining, encompassing surgical approach (anterior/posterior), decompression, fusion, instrumentation, and number of operated vertebrae. These ICD-9 codes are listed and categorized in Table 5. It is important to note that instrumentation is coded under the same ICD-9 codes as fusion except in cases of interbody cages (placed for anterior fusion), which are coded separately. This means that codes 81.02-81.08 and 81.30-81.38 serve a dual purpose of measuring both fusion and instrumentation in the final measure.

Table 5. List of ICD-9 Codes Used in the "ICD-9 Invasiveness Index for Spinal Fusion Surgery."

ICD-9 Code	Description
81.00	spinal fusion not otherwise specified
81.02	Other cervical fusion Anterior Technique
81.03	Other Cervical fusion Posterior technique
81.04	Dorsal and dorsolumbar fusion, anterior technique (thoracic or thoracolumbar region)
81.05	Dorsal and dorsolumbar fusion, posterior technique(thoracic or thoracolumbar region)
81.06	Lumbar and lumbosacral fusion, anterior technique
81.07	Lumbar and lumbosacral fusion, lateral transverse process technique
81.08	Lumbar and lumbosacral fusion, posterior technique
81.30	refusion of spine not otherwise specified
81.32	Other cervical re-fusion Anterior Technique
81.33	Other Cervical re-fusion Posterior technique
81.34	Dorsal and dorsolumbar re-fusion, anterior technique (thoracic or thoracolumbar region)
81.35	Dorsal and dorsolumbar re-fusion, posterior technique(thoracic or thoracolumbar region)
81.36	Lumbar and lumbosacral re-fusion, anterior technique
81.37	Lumbar and lumbosacral re-fusion, lateral transverse process technique
81.38	Lumbar and lumbosacral re-fusion, posterior technique
84.51	Insertion of interbody spinal fusion device
81.62	Fusion or refusion of 2-3 vertebrae
81.63	Fusion or refusion of 4-8 vertebrae
81.64	Fusion or refusion of 9 or more vertebrae
03.09	Other exploration and decompression of spinal canal (Laminectomy, Laminotomy, Expansile Laminoplasty, Exploration of Spinal Nerve Root, Foraminotomy)
80.51	Excision of intervertebral disc (Discectomy, Removal of herniated nucleus pulposus at cervical, throacic, or lumbar)
80.99	Corpectomy

Code numbers (left) and description (right) of the 23 ICD-9 codes used to assemble the final measure of surgical invasiveness for the "ICD-9 Invasiveness Index for Spinal Fusion Surgery." Codes 81.00-81.08 and 81.30-81.38 describe the region of the spine (cervical, thoracic, or lumbar) as well as approach (anterior, posterior, lateral transverse process technique) for primary fusions and re-fusions. It is important to note that 81.00-81.08 and 81.30-81.38 do not distinguish between fusion with instrumentation and fusion without instrumentation. Therefore, these codes are used to assign both fusion and instrumentation scores in the "ICD-9 Invasiveness Index for Spinal Fusion Surgery." 84.51 describes insertion of an interbody cage. 81.62-81.64 describe the number of vertebrae fused or instrumented. 03.09, 80.51, and 80.99 are codes for decompression. A new instance of 03.09, 80.51, and 80.99 are listed in the patient chart for each level decompressed.

The ICD-9 Invasiveness Index for Spinal Fusion Surgery

Using these 23 items, we created a scoring algorithm that closely resembled the algorithm used by Mirza et al. For codes 81.00-81.08 and 81.30-81.38 (fusion/instrumentation), 84.51(insertion of interbody cage), 03.09 (laminectomy, laminotomy, foraminotomy), 80.51 (discectomy), and 80.99 (corpectomy) we assigned a score of one point for each time the code appeared in the chart for the patient's operative episode. Fusion/instrumentation codes 81.02-81.08 and 81.30-81.38 are further modified by the number of levels operated (81.62 = 2-3 levels, 81.63 = 4-8 levels, 81.64 = 9 or more levels). Since exact levels are not coded individually, we conservatively chose to use the lowest number of levels for each code (i.e. 81.62 =2, 81.63 = 4 and 81.64 = 9) as the "number of levels operated." We then multiplied the fusion/instrumentation code scores by the number of levels score to arrive at the equivalent of the anterior fusion, anterior instrumentation, posterior fusion, and posterior instrumentation scores previously defined by Mirza et al. In a similar fashion to the original "Spine Surgery Invasiveness Index," any posterior fusion that included a code 84.51 for interbody cage was changed to an anterior fusion prior to multiplying by the number of levels operated.

Decompression codes (03.09, 80.51, and 80.99) are coded into the patient chart for each vertebral level decompressed. For example, in a 5 level laminectomy there would be 5 instances of ICD-9 03.09 in the patient chart. Therefore no multiplication was necessary for decompression according to number of levels. There is no distinction in approach for decompression codes. However, we assigned procedures coded with 03.09 and 80.51 to posterior decompression since laminectomies, laminotomies, foraminotomies and diskectomies are almost always performed from a posterior approach. Similarly, corpectomy is most commonly performed from an anterior approach; therefore we scored procedures coded with 80.99 as anterior decompressions.

As a preliminary test of criterion validity of this new measure, we include total scores for 6 example cases taken directly from the validation study of the original "Spine Surgery Invasiveness Index." We presented these cases to a coding expert, who assigned ICD-9 codes to each case description to ensure accuracy. We then scored each case according to the "ICD-9 Invasiveness Index for Spinal Fusion Surgery" and compared these scores to the scores presented in the validation study by Mirza et al. using linear regression (56).

Quantifying Cost Variation within New DRGs According to Invasiveness

The goal of developing this measure of invasiveness was to use it to re-categorize patients according to procedural complexity into payment groups with lower cost variation than observed in existing DRGs. To accomplish this, we first applied the "ICD-9 Invasiveness Index for Spinal Fusion Surgery" outlined above to the 2011 NIS data. This resulted in a total invasiveness score for each observation in the data. A frequency table was then constructed using SPSS and is presented below in Figure 6. The frequency table and histogram of invasiveness scores demonstrate several potential cut-points that can be used to group patients according to varying degrees of surgical invasiveness. Originally, there appeared to be 5 groups with cut-points at invasiveness scores of 3, 7, 15, and 25. However, when patients were grouped according to these cut-points, we noticed that the costs for groups 1, 2, and 3 seemed to cluster around a single mean. Therefore, in the final analysis, we combined these groups into a single group. The final cut-points were group 1≤15, 15< group 2 ≤25, and group 3>25. Using these cut points we defined 9 new DRGs by categorizing each new invasiveness group by the same "MCC", "CC," and "without MCC or CC" designations used in current DRGs. This was achieved by carrying the medical severity designation for each case over from its existing DRG assignment. New DRGs are listed with description in table 6.

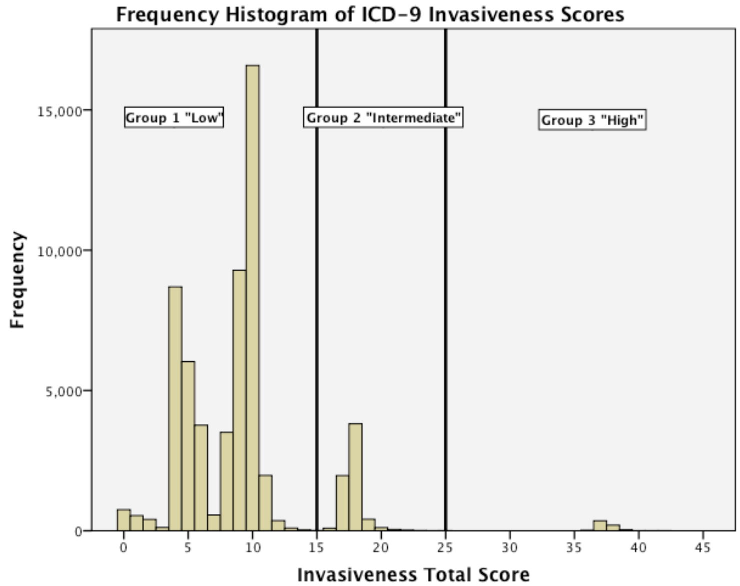
Table 6. New DRGs by ICD-9 Invasiveness scores

DRG	Description
1 MCC	Low Invasiveness with MCC
1 CC	Low Invasiveness with CC
1 without MCC/CC	Low Invasiveness without MCC/CC
2 MCC	Intermediate Invasiveness with MCC
2 CC	Intermediate Invasiveness with CC
2 without MCC/CC	Intermediate Invasiveness without MCC/CC
3 MCC	High Invasiveness with MCC
3 CC	High Invasiveness with CC
3 without MCC/CC	High Invasiveness without MCC/CC

New DRGs for patients in the 2011 NIS grouped according to procedural complexity. Groups 1, 2, and 3 represent low, intermediate, and high procedural complexity respectively. Medical severity designations were carried over for each patient from current MS-DRG classification.

Again, the primary outcome for this study was the coefficient of variation (CV) for each of these new DRGs. CVs were calculated and compared to the benchmark of TJA DRG 209 using the methods described in chapter 2 to determine the acceptability of their respective cost variations. We then examined differences in mean cost for each new invasiveness group within current spinal fusion DRGs. This was done in a fashion similar to the study by Bozic et al. that examined cost differences between different procedural groups within DRG 209 for TJA. Each current spinal fusion DRG was split into 3 categories of procedural complexity and differences in costs between each category within DRGs were tested using ANOVA.

Figure 6. Frequency histogram of ICD-9 Invasiveness scores.



Frequency histogram of total ICD-9 Invasiveness Index scores for patients in the 2011 NIS. Final cut points are marked with vertical lines at scores of 15 and 25. The low invasiveness group ≤15, 15< intermediate invasiveness group ≤25, and the high invasiveness group > 25.

RESULTS

Patient characteristics are presented by new DRG in Table 7. The DRG with the greatest number of total discharges was DRG 1 without MCC/CC, which included an estimated 200,924. Meanwhile DRG 3 without MCC/CC (high invasiveness without MCC/CC) had the fewest number of patients with just over 700. The mean age ranged from 30.2 years old in DRG 3 without MCC/CC to 62.5 years old in DRG 1 MCC. Again, the low mean age and high percentage of patients in the 1-17 year-old age category in DRG 3 without MCC/CC is consistent with the high number of pediatric scoliosis cases in this group. Similar to the patient characteristics seen in current DRGs, the majority of patients in new DRGs were female and white. The proportion of patients for which Medicare was the primary expected payer ranged from 11.6% in DRG 3 without MCC/CC to 57.06% in DRG 1 MCC, with private insurance paying for the majority of the remaining cases across all groups.

Table 7. Patient demographics by new DRG category.

DRG	1 MCC	1 CC	1 without MCC/CC	2 MCC	2 CC	2 Without MCC/CC	3 MCC	3 CC	3 Without MCC/CC
Estimated Total Discharges	11009	13259	200924	4051	7942	15454	717	1376	703
Mean Age (SD)	62.5 (15.3)	57.6 (16.1)	57.0 (14.8)	54.0 (22.6)	53.0 (21.7)	52.7 (21.3)	53.1 (27.8)	51.2 (21.8)	30.2 (22.2)
Age Category									
(1) 1-17	1.87	3.63	0.95	17.82	16.26	16.49	17.68	16.28	55.76
(2) 18-44	9.59	13.47	19.14	6.94	9.33	8.96	7.93	12.29	14.55
(3) 45-64	37	44.9	46.02	33.3	36.75	38.2	31.71	37.87	16.36
(4) 65-84	48.57	36.5	32.84	40.45	36.91	35.29	42.68	33.22	13.33
(5)85 or more	2.97	1.5	1.04	1.39	0.64	0.9	0	0.33	0
Gender (F, %)	49.71	59.25	55.75	53.47	61.94	57.44	63.41	67.44	67.27
Race %:unavailable for MN, ND, OH, and WV for 1988-2011 of NIS									
(1) White	79.29	83.71	82.22	78.82	80.93	80.71	84.52	82.29	79.1
(2) Black	9.34	6.34	7.22	9.14	7.6	8.1	5.16	6.25	8.96
(3) Hispanic	6.7	5.96	7.08	6.48	6.05	7.63	4.52	6.25	8.21
(4) Asian/PI	1.41	1.7	0.95	1.85	2.02	1.45	1.94	2.08	0.75
(5) Native American	0.45	0.28	0.28	0.69	0.35	0.18	0	0.35	0
(6) Other	2.81	2.01	2.24	3.01	3.05	1.93	3.87	2.78	2.99
Expected Payer (Not available for ME)									
(1) Medicare	57.06	42.62	38.01	45.35	40.77	37.98	47.56	38.54	11.66
(2) Medicaid	6.56	5.69	5.48	11.98	8.02	7.77	7.93	6.64	20.25
(3) Private including HMO	28.32	42.11	44.78	34.65	43.25	44.96	40.24	49.17	59.51
(4) Self-pay	1.91	0.83	0.94	2.03	1.34	0.82	1.22	1	0
(5) no charge	0.23	0.16	0.14	0.53	0.11	0.14	0	0.33	1.84
(6) other	5.92	8.59	10.65	5.45	6.51	8.34	3.05	4.32	6.75
Location of Patients Residence (Unavailable for MA)									
(1) Central Metro >= 1million population	31.81	32.48	25.9	33.05	29.14	26.9	38.75	30.54	29.27
(2) Fringe of Metro >= 1million Population	24.96	23.29	24.94	23.43	24.51	23.83	24.38	32.21	23.17
(3) Counties in Metro areas of 250K-999,999 pop	14.65	14.77	17.99	16.41	16.61	19.99	14.38	15.77	15.24
(4) Counties in Metro areas of 50,000-249,999 pop	9.58	9.94	10.77	9.4	11.71	11.35	11.25	10.4	10.37
(5) Micropolitan	10.38	10.71	12.26	9.72	9.8	10.67	6.25	5.03	14.63
(6) No metro or micro counties	8.62	8.81	8.15	7.99	8.22	7.27	5	6.04	7.32

Patient demographic factors presented by new DRG groupings. Groups 1, 2, and 3 represent low, intermediate, and high procedural complexity respectively. Medical severity designations were carried over for each patient from current MS-DRG classification.

Preliminary Validation of the ICD-9 Invasiveness Index

Tables 8 and 9 outline the scoring for the 6 example cases presented in the Mirza et al. article. Total ICD-9 Invasiveness scores showed a very high and significant correlation with Spine Surgery Invasiveness Index Scores ($r^2 = .998$, $p<0.0001$).

Table 8. Invasiveness score by "Spine Surgery Invasiveness Index"

Case Number	Description	SI_AF	SI_PF	SI_AI	SI_PI	SI_AD	SI_PD	SI_total
1	L5–S1 posterior discectomy	0	0	0	0	0	1	1
2	C5–C6 anterior cervical discectomy and fusion and plating	2	0	2	0	1	0	5
3	L4–S1 posterolateral fusion with pedicle screws and no decompression	0	3	0	3	0	0	6
4	L4–L5 posterolateral fusion with L4 and L5 laminectomy and structural graft or cages in the L4–L5 disc space	2	2	2	2	0	2	10
5	T1-Ilium posterior fusion; pedicle screws bilaterally at T1,T4, T8, L1, L2, L3, L4, L5, S1, and illium; laminectomy L1 to S1, and posterior interbody fusion with cages at L1–L2, L2–L3, L3–L4, L4–L5 and L5–S1.	6	19	6	10	0	6	47
6	Debridement of posterior lumbar wound infection after L4–L5 posterior interbody fusion.	0	0	0	0	0	0	0

Six example cases with scoring according to Spine Surgery Invasiveness Index taken directly from validation study by Mirza et al. (47) "SI" designates Spine Surgery Invasiveness Index. AI = anterior fusion, PF = posterior fusion, AI = anterior instrumentation, PI = posterior instrumentation, AD = anterior decompression, PD = posterior decompression, SI_total = total invasiveness score.

Table 9. Invasiveness score by "ICD-9 Invasiveness Index for Spinal Fusion Surgery"

Case Number	Description	ICD9_AF	ICD_PF	ICD9_AI	ICD9_PI	ICD9_AD	ICD9_PD	ICD9_Total
1	L5–S1 posterior discectomy	0	0	0	0	0	1	1
2	C5–C6 anterior cervical discectomy and fusion and plating	2	0	2	0	0	1	5
3	L4–S1 posterolateral fusion with pedicle screws and no decompression	0	2	0	2	0	0	4
4	L4–L5 posterolateral fusion with L4 and L5 laminectomy and structural graft or cages in the L4–L5 disc space	2	2	3	2	0	0	9
5	T1-Ilium posterior fusion; pedicle screws bilaterally at T1,T4, T8, L1, L2, L3, L4, L5, S1, and illium; laminectomy L1 to S1, and posterior interbody fusion with cages at L1–L2, L2–L3, L3–L4, L4–L5 and L5–S1.	9	9	10	9	0	6	43
6	Debridement of posterior lumbar wound infection after L4–L5 posterior interbody fusion.	0	0	0	0	0	0	0

Six example cases taken directly from the validation study by Mirza et al. (47) with scoring according to ICD-9 Invasiveness Index for Spinal Fusion Surgery. "ICD9" designates ICD-9 Invasiveness Index for Spinal Fusion Surgery. AI = anterior fusion, PF = posterior fusion, AI = anterior instrumentation, PI = posterior instrumentation, AD = anterior decompression, PD = posterior decompression, ICD9_total = total invasiveness score.

Primary Outcome

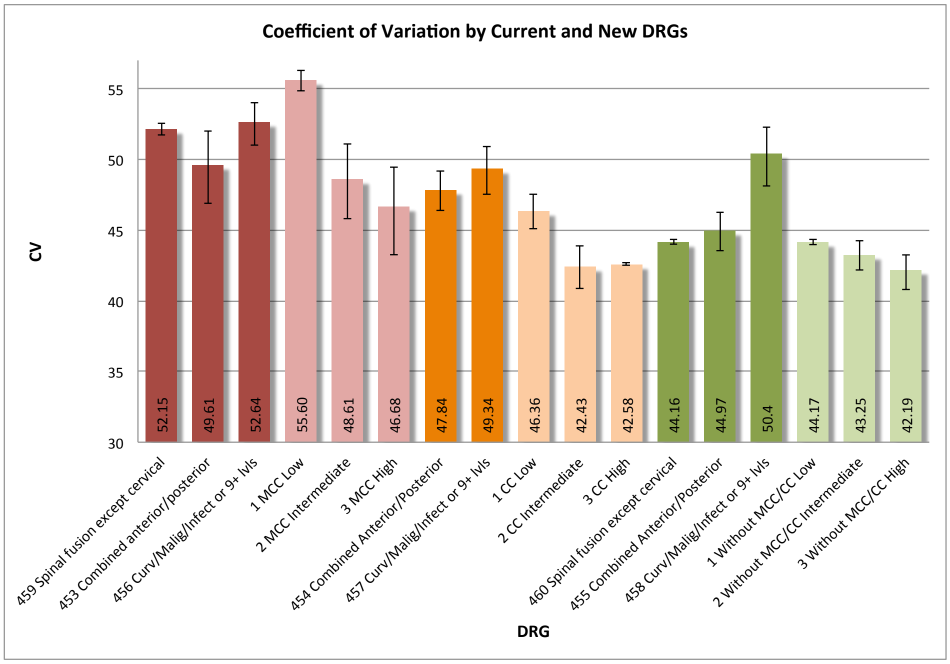
CVs for new DRG payment groups varied from 42.19 for DRG 3 without MCC/CC (high invasiveness without MCC or CC) to 55.60 for DRG 1 MCC (Low invasiveness with MCC). Within each invasiveness category, the CV was highest for patients with the "MCC" designation, followed by those with the "CC" designation, and finally those "without MCC or CC." These values remained higher than the CV for the established benchmark, DRG 209 (CV=38.21, 95%

CI 37.66-38.74).

Compared to cost variation within current DRGs, results for new DRGs were mixed. Compared to DRG 459 (spinal fusion except cervical with MCC), which had a CV of 52.15 (95% CI 51.72-52.54), DRG 1 MCC had a higher degree of cost variation, with a CV of 55.60 (95% CI 54.84-56.27). Meanwhile, the 95% confidence intervals for CV for DRG 453 overlapped with its corresponding new DRG 2 MCC, suggesting no difference in the CV for these groups. The same was true for DRG 460 and DRG 1 without MCC/CC. In contrast, CVs for DRG 456, 454, 457, 455, and 458 were improved in DRGs 3 MCC, 2 CC, 3 CC, 2 without MCC/CC, and 3 without MCC/CC respectively (See Figure 7 and table 10).

When surgical invasiveness subgroups were applied to each current spinal fusion DRG, there were significant differences in mean costs between subgroups. For example, within DRG 453, the low invasiveness group (group 1) had a mean cost of $67,753 compared to $79,524 for the intermediate invasiveness group (group 2) and $110,291 for high invasiveness group (group 3). These correspond to differences of $11,771 between low and intermediate groups and $30,767 between intermediate and high groups within DRG 453. This trend continues within each of the current DRGs when subdivided by invasiveness score. All differences between been costs within each DRG were statistically significant by ANOVA (see figure 8, table 11).

Figure 7. CV for new and current DRGs



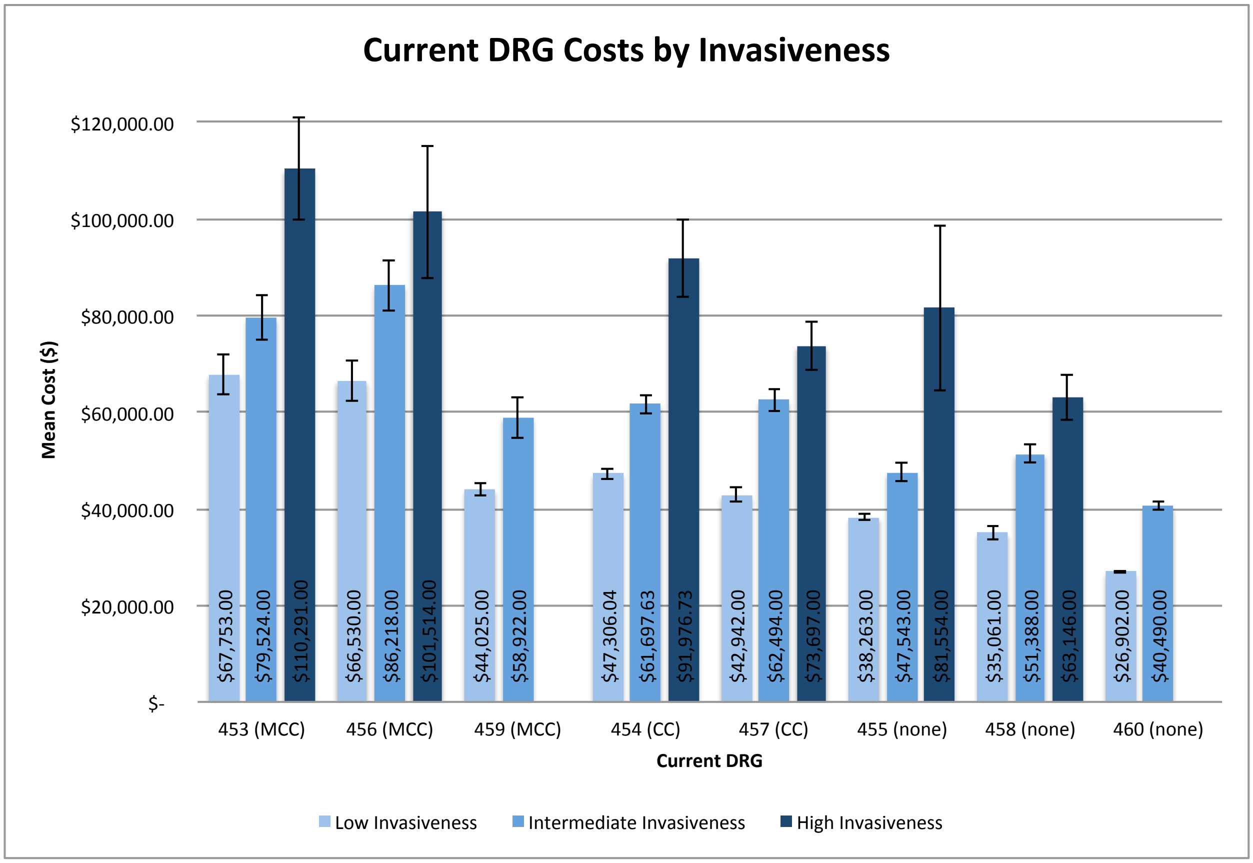
The CV (coefficient of variation) for both current spinal fusion DRGs (453-460) and new DRGs proposed in this thesis (DRGs 1-3 with MCC, CC, and without MCC/CC). Error bars represent 95% CI for the CV.

Table 10. CV for current and new DRGs

DRG	CV†† (95% CI)	
	CV	(95% CI)
459	52.15	(51.72-52.55)
1 MCC	55.60	(54.85-56.27)
453	49.61	(46.85-52.05)
2 MCC	48.61	(45.81-51.05)
456	52.64	(51.01-54.05)
3 MCC	46.68	(43.26-49.44)
1 CC	46.36	(45.10-47.50)
454	47.84	(46.39-49.13)
2 CC	42.43	(40.84-43.85)
457	49.34	(47.53-50.92)
3 CC	42.58	(42.69-42.49)
460	44.16	(44.00-44.31)
1 Without CC/MCC	44.17	(43.95-44.38)
455	44.97	(43.55-46.23)
2 Without MCC/CC	43.25	(42.18-44.24)
458	50.40	(48.12-52.31)
3 without MCC/CC	42.19	(40.80-43.25)

CV (coefficient of variation for current DRGs (453-460) and new DRGs proposed in this thesis (DRGs 1-3 with MCC, CC, and without MCC/CC). Current and new DRGs are grouped according to corresponding medical severity and procedural complexity to allow for easier comparison.

Figure 8. Mean costs for invasiveness groups within current DRGs.



This figure shows the mean costs for low, intermediate, and high invasiveness groups when these groupings are applied within each of the current spinal fusion DRGs. Error bars represent 95% CI. Mean costs are significantly different for different invasiveness groups within each DRG. Within each DRG, mean costs also increase with increasing invasiveness.

Figure 9 graphs mean costs for both current and new DRGs grouped by medical severity. For current DRGs, there is no significant difference in mean costs for the 2 most procedurally complex categories ("combined anterior/posterior spinal fusion" and "spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels") within medical severity categories. For example, there is no difference between mean costs for DRGs 453 and 456 (both with the "MCC" designation) as evidenced by overlapping 95% confidence intervals for these two groups. The same trend is observed for DRG 455 and 458 as well as for 454 and 457. However, for current DRGs the mean cost is significantly different between the least procedurally complex DRGs ("spinal fusion except cervical") compared to the more procedurally complex groups ("combined anterior/posterior spinal fusion" and "spinal fusion except cervical with spinal curvature/malignancy/infection or 9+ fusion levels").

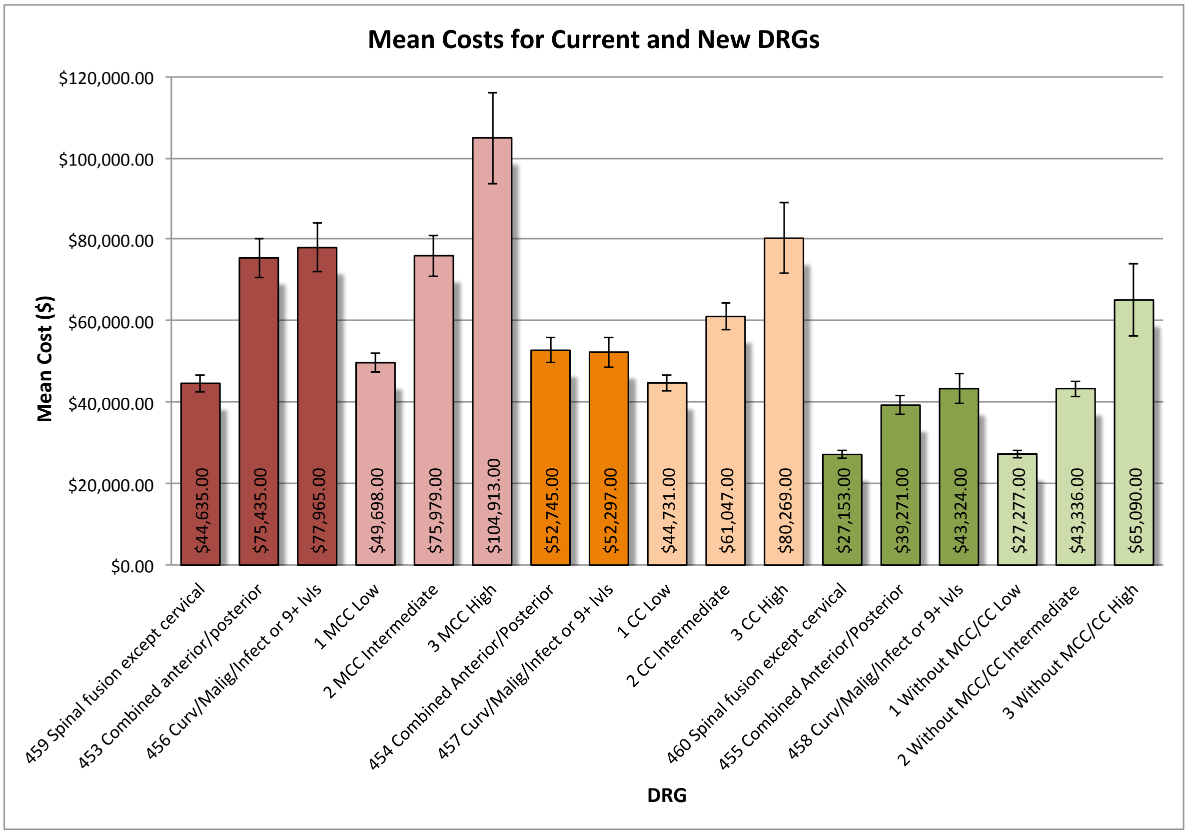
In contrast, within each medical severity category, there are significant differences in mean cost between low, intermediate, and high invasiveness groups. For example, within the MCC category, DRGs 1 MCC, 2 MCC, and 3 MCC have mean costs of $49,698, $75,979, and $104,913 respectively. These represent differences in mean cost of $26,281 between low and intermediate invasiveness groups, and $28,934 between intermediate and high invasiveness groups.

Table 11. Mean costs for low, intermediate, and high invasiveness groups within current DRGs.

453	Mean Cost ($)	95% CI	P-value*
1	$67,753	$63,693-$71,813	
2	$79,524	$74,998-$84,050	<.0001
3	$110,291	$99,691-$120,891	
454	**Mean**	**95% CI**	**P-value***
1	$47,306	$46,160-$48,452	
2	$61,698	$59,740-$63,655	<.0001
3	$91,977	$83,874-$100,080	
455	**Mean**	**95% CI**	**P-value***
1	$38,263	$37,522-$39,005	
2	$47,543	$45,672-$49,415	<.0001
3	$81,554	$64,504-$98,604	
456	**Mean**	**95% CI**	**P-value***
1	$66,530	$62,344-$70,716	
2	$86,218	$81,027-$91,408	<.0001
3	$101,514	$87,754-$115,273	
457	**Mean**	**95% CI**	**P-value***
1	$42,942	$41,434-$44,450	
2	$62,494	$60,233-$64,755	<.0001
3	$73,697	$68,756-$78,639	
458	**Mean**	**95% CI**	**P-value***
1	$35,061	$33,653-$36,469	
2	$51,388	$49,589-$53,188	<.0001
3	$63,146	$58,369-$67,922	
459	**Mean**	**95% CI**	**P-value****
1	$44,025	$42,731-$45,319	
2	$58,922	$54,675-$63,168	<.0001
460	**Mean**	**95% CI**	**P-value****
1	$26,902	$26,768-$27,037	
2	$40,490	$39,667-$41,314	<.0001

*P-value calculated using ANOVA
**P-Value Calculated using independent samples t-test

Figure 9. Mean costs for current and new DRGs



Mean costs for current and new DRGs grouped according to medical severity. “MCC” in red, “CC” in yellow, and “without MCC/CC” in green. Error bars represent 95% CI. Mean cost values are included in each bar.

DISCUSSION

In this chapter, we have proposed alternative spinal fusion payment groups, defined according to procedural complexity, that attempt to decrease the cost variation within current spinal fusion DRGs. Our findings demonstrate that while one pair of DRGs (459/1MCC) experienced an increase in cost variation, two pairs of DRGs (453/2MCC and 460/1without MCC or CC) maintained their current degree of cost variation, and five pairs of DRGs (456/3 MCC; 454/2 CC; 457/3 CC; 455/2 without MCC/CC; and 458/3 without MCC/CC) realized an improvement in cost variation. However, all groups remained above the established benchmark for acceptable variation of 38.21 set by DRG 209. This could be explained in several ways. As mentioned in chapter 2, it is possible that current medical severity categories (MCC, CC, without MCC/CC) do not adequately capture patient-specific factors that may also be contributing to excessive variation observed within both current and new DRGs. It is also possible that our proposed measure of surgical invasiveness does not fully explain the procedural factors that contribute to cost variation within current payment groups. However, categorizing patients by surgical complexity appears to improve cost variation for the majority of payment groups, suggesting that procedural factors not only make an important contribution to cost variation within spinal fusion payment groups, but that this variation is mutable to a certain degree.

Results of our secondary analysis demonstrate that grouping patients according to invasiveness within each of the current spinal fusion DRGs results in procedurally distinct groups with significantly different mean costs. These results are consistent with the previous study by Bozic et al. in TJA that examined two procedurally distinct groups within DRG 209 (primary and revision). They found that within DRG 209, mean hospital costs for primary TJA were $24,170 while mean hospital cost for revision TJA was $31,341, a difference of $7,171 ($p<0.0001$) (17). Adjusting for inflation using the consumer price index, this difference would translate to $8,259 in 2011. Meanwhile the smallest difference in mean cost observed between two groups within a spinal fusion DRG in 2011 was $9,280 between low and intermediate invasiveness groups for DRG 455. The largest difference in mean cost observed between two groups within a spinal fusion DRG in 2011 was $34,011 between intermediate and high invasiveness groups in DRG 455. According to publicly available data from the CMS, the average Medicare reimbursement for DRG 455 was $28,722 in 2011, $10,548 less than the average costs calculated in this thesis (23). In contrast, mean costs for low, intermediate, and high invasiveness groups within DRG 455 were $38,263, $47,543, and $81,554 respectively suggesting that there are patients within DRG 455 whose high degree of procedural complexity is likely resulting in significant financial losses for hospitals performing procedures for these patients.

Our findings in this chapter show that grouping patients into new DRGs according to procedural complexity results in more distinct payment groups relative to existing DRGs. This is best demonstrated in figure 10, which shows that for current MCC DRGs 459, 453, and 456, mean costs are $44,635, $77,965, and $75,435 respectively while mean costs for new MCC DRGs 1, 2, and 3 are $49,698, $75,979, and $104,913, respectively. The key point here is that DRGs 453 and 456 capture patient groups that are similar in terms of procedural complexity, resulting in mean costs that are essentially the same. Meanwhile our proposed new DRGs are more effective in separating patients into groups that are procedurally distinct, and therefore have more distinct costs means. If CMS were to adopt these DRGs, it might be able to provide more distinct and appropriate reimbursements for these groups.

Lastly, the results from the preliminary validation of the ICD-9 Invasiveness Index for Spinal Fusion Surgery are worthy of some consideration. Although we were only able to compare our measure to the six cases provided in the Spine Surgery Invasiveness Index validation article by Mirza et al., the comparison of scores for these cases was very promising with regression analysis demonstrating an R^2 value of 0.998 ($p<0.0001$) (56). If validated, this measure of surgical invasiveness could have far reaching applications due to its ease of use and availability of ICD-9 codes in administrative data. As Mirza et al. note, "Quantifying surgical invasiveness from medical records requires extensive resources. While such a complex and bulky system can be implemented in rigorous regulatory approval studies of new devices or other well-funded trials, widespread acceptance and application may require selecting subsets of risk factors and adverse outcomes that directly relate to specific patient safety concerns, *or choosing those parameters in this framework that can be ascertained reliably from brief medical records review or administrative data alone*" (57). Our measure offers the ability to obtain a measure of surgical invasiveness from administrative and medical record data using ICD-9 codes and could be broadly applied as a result.

CHAPTER 4

CONCLUSIONS

The purpose of this thesis was to investigate cost variation within spinal fusion payment groups. Cost variation within current spinal fusion DRGs was found to be unacceptably high as demonstrated by our findings that cost variation within current spinal fusion DRGs is significantly higher for all groups compared to the DRG 209 benchmark for "acceptable" variation. We found the highest degree of cost variation within DRGs with the highest procedural complexity. This trend supports the idea that, as in TJA, there may be procedurally distinct cost groups within spinal fusion DRGs that are not adequately accounted for in current DRG payments. It is also likely that the most procedurally complex patients within spinal fusion DRGs are referred to a small number of hospitals, including tertiary care hospitals and academic medical centers, that have the facilities and specialists capable of treating these more complex cases. While these referrals are entirely appropriate from a patient care perspective, these hospitals may bear the financial burden of caring for patients at the higher end of the cost distribution within spinal fusion DRGs. As in TJA, this variation may be leading to discrepancies between hospital costs and Medicare reimbursement that threatens patient access to care as hospitals attempt to limit ongoing financial losses.

While several studies have previously examined variation in costs (11,12,52) and rates (11,12,22,58) of spinal fusion surgery by geographic region and diagnosis, only one other study to our knowledge has examined cost variation within DRGs (22). This study by Boakye et al. looked at simulated 30-day, 60-day, and 90-day episodes of care for patients undergoing spinal surgery with the goal of characterizing costs that might be expected under a bundled payment system. When patients were grouped according to existing spinal surgery DRGs, they noted large differences between the minimum and maximum costs within each DRG. In addition, the interquartile range of costs within DRGs was also found to be large. However, as mentioned in chapter 2 of this thesis, describing cost variation in these absolute terms is difficult to interpret. First of all, absolute measures of spread such as the standard deviation or IQR are inherently tied to the absolute value of the mean or median cost. Groups with higher mean or median costs might reasonably have a higher standard deviation or IQR. For example, a group with a mean cost of $100 might have a standard deviation of $20 while another group with a median cost of $10,000 might have a standard deviation of $500. While the absolute value of the standard deviation is larger for the second group, when we compare the standard deviation of each group relative to its mean, we find that the CV for the first group is 20 whereas the CV for the second group is only 5. In contrast to the study by Boakye et al. this thesis is the first study to our knowledge that examines cost variation within DRGs using the CV, a standardized, unit-less measure of spread that allows for accurate comparison of within-group cost variation. Furthermore, it is the first to investigate cost variation within spinal fusion DRGs in the context of an established benchmark for cost variation. Using this benchmark, we have determined that the cost variation in spinal fusion surgery is likely to be both statistically and clinically significant.

Following this initial study, we proposed that the cost variation within current DRGs might be improved if patients were more accurately grouped according to surgical complexity. In order to create these groups, we modified an existing measure of surgical invasiveness and used it to re-categorize patients into 9 new DRGs according to total invasiveness score (56). This thesis is the first study to our knowledge to propose a measure of surgical invasiveness based on ICD-9 procedure codes. In addition, it is the first study to propose new DRGs in which patients are grouped according to invasiveness. Although somewhat mixed, our findings demonstrate a decrease in cost variation for a majority of new DRGs. In addition, results from our secondary analysis demonstrate that, as in TJA, grouping patients according to invasiveness within each of the current spinal fusion DRGs results in procedurally distinct groups with significantly different mean costs. Furthermore, when comparing mean cost differences across DRGs, our new DRGs are also more effective in separating patients into procedurally distinct cost groups compared to the cost groups defined by current DRGs. The result is a set of new DRGs with decreased cost variation *within* groups and increased cost variation *between* groups, ultimately resulting in an improvement cost resolution compared to current DRGs. In terms of affecting these changes, CMS encourages comments and input from the public as well as from the medical and research communities on ways to improve current DRG payment groups (20). With further investigation and support from professional orthopedic associations, avenues are available to recommend concrete changes to spinal fusion DRGs that would allow for more accurate Medicare reimbursements and ensure more equitable access for all patients.

Strengths and Limitations

This research has several strengths and limitations worth noting. First, although we have identified a high degree of cost variation within current spinal fusion DRGs, we are limited in our ability to extrapolate the direct effects of this variation on patient access to spinal fusion surgery. However, previous studies have demonstrated that cost variation within DRGs can lead to both undue financial burden on hospitals as well as patient access concerns that are large enough to necessitate modification to DRG coding (17-20). Furthermore, we do not adjust wage index, geographic location, teaching status, or other characteristics in our analysis that may also contribute to cost variation within DRGs. We have attempted to deal with this by selecting a reference group “DRG 209” that should, in theory, also be subject to the same effects of these factors on cost variation. Given that hospital cost data is reported as total charges in the NIS, we also chose CCRs to attempt to control for differences in hospital inflation of charges. However, the CCRs are still estimates of direct hospital costs and are therefore associated with some degree of error that cannot be measured or accounted for in this thesis. Another limitation of this thesis was the use of a previously un-validated measure of invasiveness. Although we did our best to match the scoring to the previously validated “Spine Surgery Invasiveness Index,” we will not be able to establish the scientific validity of this measure until a larger validation study is performed.

Despite these limitations, our research also has several strengths. First, and perhaps most importantly, our use of national data makes our findings generalizable on a national scale. This will ultimately strengthen any effort to convince CMS to change DRG coding for spinal fusion surgery in the future. Also, given the large number of patients available in our sample, this thesis is highly powered to detect differences between groups. Also our use of a benchmark group allows us assess the economic and policy impact of our findings based on precedent.

Personal Knowledge Gained

This process has offered me many important learning opportunities. First of all, it has given me insight into both the history and the complexities of healthcare provider reimbursement in the United States. Secondly, from a methodological perspective, I have gained valuable experience in application of complex statistical methods through the use of both SPSS and SAS, both of which I know I will use again in the future. Furthermore, I have an understanding of stratified sampling and weighting schemes as well as experience working with national data with millions of patients. I have also gained a deeper appreciation for the scientific method, especially the importance of asking a well-developed research question. Although I still have much to learn, this thesis was an important and invaluable first step towards my development as both a researcher and an evidence-based clinician.

Future Directions

This thesis is only the beginning of the work that must be done to better characterize cost variation within spinal fusion payment. As mentioned earlier, further research is needed to determine the acceptability of current medical severity designations that capture patient-specific factors contributing to cost variation within payment groups. Research in this area would carry broad implications for all diagnoses under the IPPS and may also help further elucidate the degree to which procedural factors contribute to cost variation. Further studies are also needed to more directly link disparities in patient access to care to discrepancies between Medicare reimbursement and hospital costs. If this link could be more directly shown, it would likely galvanize efforts for change. In terms of proposing improvements to existing DRGs, validation of the ICD-9 Invasiveness Index for Spinal Fusion Surgery in a larger patient population would help support its use for categorizing patients into homogenous payment groups and allow for more widespread application of this index in both research and clinical practice. Lastly, now that these trends in cost variation have been established on a national level, studies at a more local or institutional level may both corroborate the findings of this thesis as well as give further insight into which specific factors contribute most to cost differences within payment groups. This knowledge could then be used to further homogenize current DRGs payment groups and ensure more equitable access for all patients.

Conclusions

This thesis has demonstrated that cost variation within current spinal fusion DRGs is unacceptable. Reorganizing patients into new DRGs according to surgical complexity resulted in a decrease of within-group cost variation as well as an increase in between-group differences for mean cost. Further examination of cost variation within spinal fusion payment groups is warranted to determine the direct impact of this high degree of cost variation on patient access to care. Future studies should also seek to determine ways to further homogenize DRGs in an effort to ensure fair reimbursement for hospitals and equitable access for all patients.

REFERENCES

1. Aebi M. Correction of degenerative scoliosis of the lumbar spine. Clin Orthop 1988;232:80–6.
2. Herkowitz HN, Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis. A prospective study comparing decompression with decompressionand intertransverse process arthrodesis. J Bone Joint Surg Am 1991;73A: 802–8.
3. Herkowitz HN. Degenerative lumbar spondylolisthesis. Spine 1995;20:1084–90.
4. Christensen FB, Hansen ES, Laursen M, et al. Long-term functional outcome of pedicle screw instrumentation as a support for posterolateral spinal fusion. Spine 2002;27:1269–77.
5. Turner JA, Ersek M, Herron L, et al. Patient outcomes after lumbar spinal fusions: A comprehensive literature synthesis. JAMA 1992;268:907–11.
6. Charstil, J, Patel, AA. Complications associated with posterior and transforaminal lumbar interbody fusion. J Am Acad Orthop Surg. 2012 May;20(5):283-91
7. Liu, JK, Laufer, I, Bilsky, MH. Update on Management of Vertebral Column Tumors. CNS Oncology. 2014 Mar;3(2):137-47
8. Kotwicki T, et. al. Optimal management of idiopathic scoliosis in adolescence. Adolescent Health Med Ther. 2013 Jul 23;4:59-73.
9. Rajaee, S., Bae, H., Kanim, L. & Delamarter, R. Spinal Fusion in the United States: Analysis of Trends From 1998 to 2008. Spine 37, 67 (2012).
10. Baaj, A., Downes, K., Vaccaro, A., Uribe, J. & Vale, F. Trends in the treatment of lumbar spine fractures in the United States: a socioeconomics perspective. Journal of Neurosurgery: Spine 15, 367370 (2011).
11. Weinstein, J., Lurie, J., Olson, P., Bronner, K. & Fisher, E. United States' trends and regional variations in lumbar spine surgery: 1992-2003.Spine 31, 2707–14 (2006).
12. Deyo, R. A. et al. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. JAMA : the journal of the American Medical Association 303, 1259–65 (2010).
13. Cowan, J. A. et al. Changes in the utilization of spinal fusion in the United States.Neurosurgery 59, 15–20; discussion 15–20 (2006).
14. Cherkin, D. C., Deyo, R. A., Loeser, J. D., Bush, T. & Waddell, G. An international comparison of back surgery rates. Spine 19,1201–6 (1994).
15. Centers for Medicare and Medicaid Services. "Acute Inpatient PPS." https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html?redirect=/acuteinpatientpps/. Accessed April 2015.
16. W C Hsiao, H M Sapolsky, D L Dunn and S L Weiner. "Lessons of the New Jersey DRG payment system." Health Affairs, 5, no.2 (1986):32-45.
17. Bozic K. et al. Hospital resource utilization for primary and revision total hip arthroplasty. JBJS: Journal of Bone and Joint Surgery 87, 570-576 (2005).
18. Barrack, R. L. The evolving cost spectrum of revision hip arthroplasty. Orthopedics 22, 865–6 (1999).
19. Bozic, K. CMS changes ICD-9 and DRG codes for revision TJA. http://www2.aaos.org/bulletin/jun05/fline.asp American Academy of Orthopaedic Surgeons Bulletin June 2005.
20. Department of Health and Human Services. "Medicare Program; Changes to the Hospital Inpatient Prospective Payment Systems and Fiscal Year 2006 Rates." Federal Register Vol.

70 No. 155; 47303-47305: Friday August 12, 2005 https://www.cms.gov/Regulations-and-Guidance/Regulations-and-Policies/QuarterlyProviderUpdates/downloads/cms1500f.pdf. Accessed April 2015.
21. Department of Health and Human Services. "Medicare Program; Changes to the Hospital Inpatient Prospective Payment Systems and Fiscal Year 2008 Rates." Federal Register Vol. 72 No. 85; 24729-24731: Thursday, May 3, 2008 http://www.gpo.gov/fdsys/pkg/FR-2007-05-03/pdf/07-1920.pdf. Accessed April 2015.
22. Boakye, M. *et al.* Spinal Surgery: Variations in Health Care Costs and Implications for Episode-Based Bundled Payments. *Spine* **39,**1235 (2014).
23. Centers for Medicare & Medicaid Services. "100% MEDPAR Inpatient Hospital National Data for Fiscal Year 2011 Short Stay Inpatient Diagnosis Related Groups." http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/Downloads/DRG11.pdf. Accessed April 2015.
24. The World Bank. "Health expenditure, total (% of GDP)." World Health Organization Global Health Expenditure database. http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS. Accessed April 2015.
25. The World Bank. "Health expenditure per capita (current US$)." World Health Organization Global Health Expenditure database. http://data.worldbank.org/indicator/SH.XPD.PCAP. Accessed April 2015.
26. The World Bank. "Military expenditure (% of GDP)." Stockholm International Peace Research Institute (SIPRI), Yearbook: Armaments, Disarmament and International Security. http://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS. Accessed April 2015.
27. The World Bank. "Government expenditure on education, total (% of GDP)." UNESCO Institute for Statistics. http://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS. Accessed April 2015.
28. Patient Protection and Affordable Care Act, 42 U.S.C. § 18001 (2010).
29. Center for Medicare and Medicaid Services. "National Health Expenditure Projections 2012-2022" http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/downloads/proj2012.pdf. Accessed April 2015.
30. Anthem Blue Cross. "Provider Reimbursement." http://www.anthem.com/wps/portal/ahpprovider?content_path=provider/co/f5/s1/t0/pw_000557.htm&state=nv&label=Provider+Reimbursement#P12_313. Accessed April 2015.
31. Uwe E. Reinhardt. The Pricing Of U.S. Hospital Services: Chaos Behind A Veil Of Secrecy. http://content.healthaffairs.org/content/25/1/57.full.html. Health Affairs, 25, no.1 (2006):57-69. Accessed April 2015.
32. American Hospital Association, Hospital Statistics 2005 (Chicago: AHA, October 2005), Table 3.
33. Mayes, Rick. Medicare Prospective Payment and the Shaping of U.S. Health Care. Introduction; The Johns Hopkins University Press, 2006.
34. Gray, Bradford H. The Profit Motive and Patient Care: The Changing Accountability of Doctors and Hospitals. Chapter 3; Cambridge, Mass.: Harvard U, 1991.
35. W C Hsiao, H M Sapolsky, D L Dunn and S L Weiner. Lessons of the New Jersey DRG payment system. Health Affairs, 5, no.2 (1986):32-45.
36. Dowling, William L. Prospective reimbursement of hospitals. Inquiry. 1974 Sep; 11(3):163-80.

37. "Acute Inpatient PPS" CMS.gov. http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html n.p., n.d., Accessed April 2015.
38. Brady, Tim, et al. Medicare Hospital Prospective Payment System: How DRG Rates Are Calculated and Updated. Centers for Medicare and Medicaid Services, Office of Inspector General. August 2001. https://oig.hhs.gov/oei/reports/oei-09-00-00200.pdf. Accessed April 2015.
39. Andersen, R. M. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 36, 1–10 (1995).
40. "MS DRGv28 Definitions Manual: Appendix C" CMS.gov http://www.cms.gov/icd10manual/fullcode_cms/p0031.html n.p., n.d. Accessed April 2015.
41. Centers for Medicare and Medicaid Services FY 2014 Final Rule Tables. Table 5 http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY-2014-IPPS-Final-Rule-Home-Page-Items/FY-2014-IPPS-Final-Rule-CMS-1599-F-Tables.html Accessed April 2015.
42. Agency for Healthcare Research and Quality. HCUP databases. Healthcare Cost and Utilization Project databases Web site. 2011 http://www.hcup-us.ahrq.gov/nisoverview.jsp. Accessed April 2015.
43. Introduction to the HCUP Nationwide Inpatient Sample 2011. http://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2011.jsp Accessed April 2015.
44. Wegener, S., Riley, L., Skolasky, R., Maggard, A. & Thorpe, R. United States Hospital Admissions for Lumbar Spinal Stenosis: Racial and Ethnic Differences, 2000 Through 2009. Spine 38, 2272 (2013).
45. Changes in the NIS Sampling and Weighting Strategy for 1998. http://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2011.pdf pg. iii. Accessed March 2015.
46. Rihn, J. A., Currier, B. L., Phillips, F. M., Glassman, S. D. & Albert, T. J. Defining the value of spine care. *J Am Acad Orthop Surg* **21,**419–26 (2013).
47. Alvin, M. D. *et al.* Variations in cost calculations in spine surgery cost-effectiveness research. *Neurosurg Focus* **36,** E1 (2014).
48. HCUP-NIS Overview Website. Cost-to-Charge Ratio Files. http://www.hcup-us.ahrq.gov/db/state/costtocharge.jsp. Accessed April 2015.
49. SAS/STAT Software Examples. Estimating the Standard Deviation of a Variable in a Finite Population. http://support.sas.com/rnd/app/examples/stat/SurveyStdDev/new_example/index.html. Accessed March 2015
50. HCUP-NIS Method Series. Calculating Nationwide Inpatient Sample Variances. Report #2003-02. Revised Dec. 18, 2014. http://www.hcup-us.ahrq.gov/reports/methods/2003_02.pdf Accessed March 2015.
51. Lo, W.L. L. *et al.* Comparing miniopen and minimally invasive transforaminal interbody fusion in single-level lumbar degeneration.*Biomed Res Int* **2015,** 168384 (2015).
52. Andrew J. Schoenfeld et al. Variations in Medicare payments for episodes of spine surgery. The Spine Journal 14 (2014) 2793–2798.
53. Irwin, Z. *et al.* Variation in surgical decision making for degenerative spinal disorders. Part I: lumbar spine. *Spine* **30,**2208–13 (2005).
54. Turner JA, Ersek M, Herron L, et al. Patient outcomes after lumbar spinal fusions: A comprehensive literature synthesis. JAMA 1992;268:907–11

55. Weinstein, J., Lurie, J., Olson, P., Bronner, K. & Fisher, E. United States' trends and regional variations in lumbar spine surgery: 1992-2003. *Spine***31,** 2707–14 (2006).
56. Mirza, S. *et al.* Development of an Index to Characterize the 'Invasiveness' of Spine Surgery: Validation by Comparison to Blood Loss and Operative Time. *Spine* **33,** 2651 (2008).
57. Mirza, S. K. *et al.* Towards standardized measurement of adverse events in spine surgery: conceptual model and pilot evaluation. *BMC Musculoskelet Disord* **7,** 53 (2006).
58. Bederman et al. Who's in the driver's seat? The influence of patient and physician enthusiasm on regional variation in degenerative lumbar spinal surgery: a population-based study. *Spine* 36, 481-489 (2011).