Trends And Geographic Variations In Major Surgery For Degenerative Diseases Of The Hip, Knee, And Spine

Is there a roadmap for change?

by James N. Weinstein, Kristen K. Bronner, Tamara Shawver Morgan, and John E. Wennberg

ABSTRACT: Although Medicare rates for surgery to treat degenerative diseases of the hip, knee, and spine are highly variable among hospital referral regions (HRRs), the relative risk for surgery within a region is constant from year to year—a large majority of the variation in surgery in 2000–01 is “explained” by the variation in rates in 1992–93. The within-region constancy in rates for highly variable procedures (the “surgical signature”) is illustrated for South Florida HRRs. Involving the patient in choice of treatments (shared decision making) and outcomes research are promising strategies for reducing unwarranted regional variation and local constancy in surgery risk.

Musculoskeletal disease is a major source of disability, costing the U.S. economy an estimated $215 billion in health care services and lost economic productivity in 1995.1 Degenerative diseases of the hip, knee, and spine are among the most common and most costly of musculoskeletal diseases. During the past decade or so, innovations in pharmaceuticals, surgical techniques, and biomedical devices have greatly increased the treatment options available to patients with these diseases. Since the choice of treatment typically involves trade-offs among competing risks and benefits, patients need to be informed participants in the process. However, because patients commonly delegate decision making to physicians, patients’ preferences can be misinterpreted, which can lead to decisions that more closely reflect providers’ opinions than patients’ opinions. Moreover, because clinical trials and long-term cohort studies are rarely performed for these conditions, clinical decisions are often made with much scientific uncertainty concerning the outcomes.

This uncertainty has provided a useful framework for interpreting the contribution of clinical decision making to geographic variations in the incidence of surgery. Evaluating the quality of scientific evidence that supports surgical theories has established a rough association between the degree of scientific uncer-
tainty and the degree of variation in the incidence of surgery among regions. The professional-uncertainty hypothesis has also helped explain why physician practice style could be an important determinant of demand; in the absence of professional consensus based on outcomes, individual or small groups of physicians can hold onto idiosyncratic clinical rules of thumb defining who needs surgery. In a given region, local physicians tend to apply their rules of practice consistently, which results in the “surgical signature” phenomenon: rates for specific surgical procedures that are idiosyncratic to a region, sometimes differing dramatically among neighboring regions. Studies have also shown that over time, specific procedures tend to maintain their relative variability and communities maintain their surgical signatures, at least in the absence of interventions designed to reduce scientific uncertainty or increase patient involvement in the choice of treatments.2

In this study we examine the patterns of utilization among 306 hospital referral regions (HRRs) for major orthopedic procedures for patients with degenerative diseases of the knee (total knee replacement), hip (total hip replacement), and spine. We measure the degree of variability for each procedure for 2000–01 and compare it with the variability of hospitalization for hip fracture repair, which typically shows little variation. We illustrate the surgical signatures for these procedures by comparing profiles of rates among HRRs in Florida. We examine trends in use rates over a ten-year period and investigate the stability of the surgical signatures by correlating rates in 1992–93 with rates in 2000–01. We also examine the contribution of income, population density, and supply of surgeons to patterns of use. Finally, we examine the trends and pattern of use for two surgical approaches for degenerative disease of the spine: spine surgery with and without fusion. We conclude with a clinical interpretation of patterns of variation in surgery rates for these conditions and consider the policy steps needed to break the cycle of unwarranted variation.

Study Data And Methods

Data. Data on use of and reimbursement for hospital costs are from the Centers for Medicare and Medicaid Services (CMS) annual Medicare Provider Analysis and Review (MEDPAR) file, which contains records for inpatient surgery enrollees participating in traditional fee-for-service (FFS) Medicare. For the period 1992–2001, patients with hip fractures, total knee and hip replacements, and back surgery (with and without fusion) were identified.3 The annual incidence of surgery has been calculated for Medicare enrollees living in the 306 HRRs using the mid-year population file to create the denominator for rates. HRRs (developed as part of the Dartmouth Atlas of Health Care project) were formed through a two-step process.4 First, ZIP codes were aggregated into hospital service areas (HSAs) according to frequency of hospital use. Second, HSAs were aggregated into HRRs based on the frequency of use of bypass surgery and neurosurgery. The denominator for rates is the count of enrollees in Medicare Part A (traditional Medicare) residing in each HRR on 1 July of each year.

Methods. Since surgical procedures, like other forms of medical care, vary according to age, sex, and race, we removed the confounding effect of these variables using the indirect method of adjustment.5 We measured variability in incidence of hospitalization for hip fracture and rates for orthopedic surgery among regions using the mean, interquartile, and 95th percentile ranges, and the systematic component of variation (SCV). The SCV subtracts the random component of variance from the estimate of total variance, thus providing valid comparisons of relative variability among procedures with different prevailing rates.6 Using ordinary least squares (OLS) regression, we explored consistencies in the patterns of practice over time and the roles of income, population density, and supply of orthopedic and (for back surgery) neurosurgeons. We selected contiguous regions in South Florida for surgical profiling.
Results

Patterns of variation. Exhibit 1 provides a graphic representation as well as the SCV statistic describing the degree of variability in rates among the 306 HRRs for each surgical procedure compared with hip fracture hospitalization rates in 2000–01. Hip fracture showed relatively little variation. Knee replacement and hip replacement were approximately four and five times more variable, as measured by the SCV. Rates in the highest regions were more than five times greater than in the lowest regions, and the interquartile ratio (the ratio between the rates for regions ranked 75th and 25th) was 1.31 for knee and 1.45 for hip replacement. Rates for back surgery were about seven times more variable than rates of hospitalization for hip fracture. The procedure-specific degree of variation in rates among regions showed little change over the ten-year period for hip and spine; some decline was evident for knee. The SCVs for 1992–93 for knee and hip replacement were 78 and 70, respectively; for 2000–01 they were 55 and 67, respectively; for spine surgery it was 90 for 1992–93 and 93 for 2000–01.

Surgical signatures among South Florida HRRs. Exhibit 2 profiles the rates for 2000–01 for back surgery and for knee and hip replacement among eight contiguous HRRs in South Florida, four on its west coast and four on its east coast. The incidence of surgery for each procedure is presented as its ratio to the national average. Regions are organized according to location on the west or east coast and ordered from south to north. The profiles illustrate the sharp transition that can occur in the risk for surgical intervention across the boundary of contiguous HRRs. For example, enrollees in the Bradenton HRR experienced a 75 percent greater rate of spine surgery than their neighbors to the north in Tampa. Interesting juxtapositions of surgical risk were also apparent on the east coast. For example, the risk for undergoing hip replacement among residents of Fort Lauderdale was nearly twice that for residents of Miami.

The sharp transitions in risk of surgery among local communities are associated with equally sharp differences in per capita costs for specific procedures. For example, in 2000–01, per capita Medicare reimbursements for

![Exhibit 1](image-url)

**Exhibit 1**
Pattern Of Variation And Systematic Coefficient Of Variation (SCV) For Hospitalization For Hip Fracture, Knee And Hip Replacement, And Back Surgery Among 306 Hospital Referral Regions (HRRs), 2000–01

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hip fracture</td>
<td>13.8</td>
<td>13.8</td>
</tr>
<tr>
<td>Knee replacement</td>
<td>55.0</td>
<td>55.0</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>67.2</td>
<td>67.2</td>
</tr>
<tr>
<td>Back surgery</td>
<td>93.6</td>
<td>93.6</td>
</tr>
</tbody>
</table>

SOURCE: Authors’ analysis of Medicare claims data.
NOTES: The systematic coefficient of variation (SCV) is shown in parentheses; see text for an explanation. The distribution in rates is ordered in terms of relative variability, increasing from left to right. A log scale, centered on the median rate (the horizontal line), is used for clarity. The upward boundary of the box is the rate for the 77th ranked region (25th percentile); the lower boundary is the rate for the 22nd ranked region (75th percentile). The upper and lower boundaries of the whisker are set at the 5th and the 95th percentiles, which means that sixteen regions have rates higher and sixteen have rates lower than indicated by the whisker boundary lines.
the three procedures for inpatient care (Part A) in Fort Myers were 1.66 times greater than those for Miami residents ($155 versus $95), differences explained almost entirely by the greater per capita rates of surgery in Fort Myers. However, variations in discretionary surgery were not positively correlated with overall Medicare spending. For example, in 2001, total (age-, sex-, and race-adjusted) per enrollee Medicare spending in Miami was 1.65 times greater than in Fort Myers ($10,113 versus $6,136). Indeed, among the 306 HRRs, the rates for these three procedures in 2000–01 were inversely correlated with overall Medicare spending for 2001 ($r = –.49; p < .0001). The surgical signatures of a region tend to persist over longer periods of time, even though, as discussed below, the rates for these procedures increased during the 1990s. Exhibit 3 profiles the rates for back surgery and knee and hip replacement over the ten-year period 1992–2001 for three South Florida regions. Rates are expressed as the ratio to the U.S. average in the corresponding period of time and are adjusted for differences in age, sex, and race. The exhibit also lists the number of operations in excess or in deficit of the number predicted by the U.S. national rate over the ten-year period—again, adjusted for differences in age, sex, and race. For Medicare enrollees living in Fort Lauderdale, the risk for hip replacement was consistently greater than the national average, while the risk for knee replacement was consistently below it, and, for back surgery, slightly above it (Exhibit 3). The rates for all three procedures were consistently lower for residents of Miami; by contrast, residents of Fort Myers experienced consistently higher rates for all three procedures in each period. This contrast between patterns of practice is striking. Over the ten-year period, Fort Myers experienced some 4,800 more back surgeries, 3,100 more knee replacements, and 1,500 more hip replacements than predicted by the Miami rate.

### Change and constancy over time
The patterns of practice seen in South Florida typify the national experience. Among enrollees in traditional Medicare, the U.S. average rate trended upward for each procedure during the period from 1992–93 to 2000–01. The greatest increase was for spine surgery, which increased 53 percent—from 2.8 to 4.3 per 1,000. The rate for total knee replacement increased 40 percent—from 4.1 to 5.7 per 1,000. Total hip replacement exhibited a smaller increase, rising from 2.1 to 2.9, or 34 percent. However, even though rates were increasing in most regions, HRRs that were high in 1992–93 tended to be high in 2000–01, and vice versa. In other words, there was little evidence of regression to the mean, as indicated by the high correlation between HRR rates in 1992–93 and in 2000–01. Among the 306 regions, for total knee
replacement, the univariate R$^2$ correlation between 1992–93 and 2000–01 rates was .75; for total hip replacement it was .81; for spine surgery it was .51.

We also examined the association between 2000–01 surgery rates, the supply of orthopedic surgeons and neurosurgeons, and median HRR income and population density (a surrogate for distance to care). In each of the three models, higher income and greater population density were statistically associated with lower rates; greater physician supply was associated with higher rates for the hip replacement model (orthopedic surgeons) and the back surgery model (neurosurgeons). However, these variables provided virtually no additional explanatory power over that associated with the 1992–93 surgery rate (R$^2$ increased less than 0.03).7

### Back surgery with and without fusion.

Over the period 1992–2001, spine surgery with fusion became much more popular, rising 137 percent, from 0.6 procedures per 1,000 to 1.4 procedures per 1,000; by contrast, back surgery without fusion rose only 32 percent. In 1992–93, surgery with fusion represented only 17 percent of spine surgery; by 2000–01, it accounted for 36 percent. The rates among regions, however, were extremely variable. The SCV statistic for the distribution in rates among the 306 regions was 183, thirteen times greater than for hip fracture and two times greater than for spine surgery without fusion. Among the 306 HRRs, we noted a positive correlation between these two forms of back surgery ($R^2 = .30$). There was wide disparity in use of spine surgery with fusion among the eight South Florida HRRs. In 2000–01 the surgeons serving the populations of Bradenton and Orlando performed this operation 2.4 and 1.6 times more frequently than the U.S. average, respectively, while rates for Miami were 60 percent of the national average.

### Discussion

#### Hip fracture.

For hospitalizations for hip fracture, there is little controversy or ambiguity concerning the effectiveness of treatment or patient preference. The condition is painful, debilitating, and life-threatening, with 30 percent mortality at one year for FFS Medicare enrollees. The condition is more or less uniformly diagnosed, and virtually all patients are hospitalized. The demand for hospitalization is thus closely determined by the incidence of hip fracture. By contrast, the rates

### Exhibit 3

**Rates Of Orthopedic Procedures, By Year, Among Three South Florida Hospital Referral Regions (HRRs) Relative To The U.S. Average And The Number Of Cases Above Or Below Expected Based On The U.S. National Rate, 1992–2001**

<table>
<thead>
<tr>
<th>HRR/procedure</th>
<th>92–93</th>
<th>94–95</th>
<th>96–97</th>
<th>98–99</th>
<th>00–01</th>
<th>Number of cases above or below U.S. average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ft. Myers (150,155)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back surgery</td>
<td>2.05</td>
<td>2.07</td>
<td>1.76</td>
<td>1.81</td>
<td>1.67</td>
<td>+4,846</td>
</tr>
<tr>
<td>Knee replacement</td>
<td>1.53</td>
<td>1.38</td>
<td>1.25</td>
<td>1.31</td>
<td>1.48</td>
<td>+3,127</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>1.39</td>
<td>1.40</td>
<td>1.23</td>
<td>1.35</td>
<td>1.45</td>
<td>+1,484</td>
</tr>
<tr>
<td>Fort Lauderdale (275,553)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back surgery</td>
<td>1.08</td>
<td>1.11</td>
<td>1.11</td>
<td>1.02</td>
<td>1.02</td>
<td>+658</td>
</tr>
<tr>
<td>Knee replacement</td>
<td>0.78</td>
<td>0.83</td>
<td>0.83</td>
<td>0.81</td>
<td>0.88</td>
<td>−2,626</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>1.15</td>
<td>1.11</td>
<td>1.18</td>
<td>1.15</td>
<td>1.23</td>
<td>+1,272</td>
</tr>
<tr>
<td>Miami (172,787)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back surgery</td>
<td>0.55</td>
<td>0.57</td>
<td>0.63</td>
<td>0.72</td>
<td>0.69</td>
<td>−2,410</td>
</tr>
<tr>
<td>Knee replacement</td>
<td>0.61</td>
<td>0.64</td>
<td>0.66</td>
<td>0.59</td>
<td>0.63</td>
<td>−3,694</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>0.61</td>
<td>0.64</td>
<td>0.66</td>
<td>0.59</td>
<td>0.63</td>
<td>−1,659</td>
</tr>
</tbody>
</table>

**SOURCE:** Authors’ analysis of Medicare claims data.

**NOTE:** The enrollee population in 2000 is shown in parentheses by each city’s name.

---

**Trends**

**HEALTH AFFAIRS ~ Web Exclusive**

**VAR-85**
among regions for total knee and hip replacement and for back surgery are much more variable, and the degree of variation tends to be characteristic of individual procedures. Median income and population density tended to be inversely associated with surgery rates, and the supply of surgeons was essentially unrelated to rates for knee surgery and explained only a small proportion of the variation in back and hip replacement surgery among HRRs in 2000–01. What mattered most in predicting the risk in 2000–01 was the risk in 1992–93. In other words, regions’ surgical signatures are remarkably stable over time.

■ **Role of patient preference.** Although systematic differences in patients’ preferences could be evoked to explain the patterns of variation, the striking differences in rates among neighboring regions (defined according to the providers they most often use) suggest that this hypothesis lacks face validity. It seems highly improbable, for example, that Medicare retirees living in Fort Myers prefer back surgery 2.4 times more often than residents of Miami or that retirees living in Fort Lauderdale prefer knee replacement over hip replacement by a ratio of 1.4. What matters is the HRR where patients live (and therefore the clinical opinions of the physicians from whom the patients receive their care).

Gillian Hawker and her colleagues provide direct evidence that professional opinion concerning a patient’s need for knee replacement can differ from the patient’s own preferences. In an effort to estimate the prevalence of surgical need, they conducted physical examinations of a random sample of Ontario residents and determined that 4.5 percent of female respondents (742 of 16,521) were “potential candidates” for arthroplasty. However, after these same patients were interviewed, only 14 percent of the potential candidates actually preferred to have the operation; the other 86 percent preferred the more conservative medical option.8

Experimental evidence concerning the divergence of physician opinion and patient preference is provided by the growing number of clinical trials of patient decision aids (PtDAs) that compare shared decision making to a control group. Elizabeth Phelan and her colleagues, in one such clinical trial, showed that patients’ preferences for spine surgery differed from those of the control group: The frequency of use of surgery for spinal stenosis increased in the control group, while for herniated disc it decreased.9 The changes in demand seemed to be in concordance with the limited information available on the outcomes of medical versus surgical management: Observational studies indicate that most patients with herniated discs treated expectantly (nonsurgical, watchful waiting) get better over time, while those with spinal stenosis tend to stay the same or get worse.10

■ **Role of scientific uncertainty.** Scientific uncertainty also contributes to variability in clinical decision making. Major surgery is often conducted without an adequate scientific basis for making a reasonably accurate estimate of the likely outcomes. This is clearly the case for some degenerative conditions of the back and less so for the hip and knee.11 The evidence base guiding the use of the most variable orthopedic procedure, back surgery with fusion, is particularly weak, even though it has enjoyed the most rapid increase in use among all such procedures during the past decade or so. During this period, there has been an explosion of surgical and commercial interest in widely varying methods of instrumented fusion in Europe and the United States, but the scientific evaluation of outcomes has not kept up with the changes in operative techniques.12 Indeed, the few clinical trials that have been performed do not show improved clinical outcomes compared with surgery without fusion.13 Given the paucity of clinical trials, it is not possible to draw conclusions concerning the role of
instrumented fusion for a given spinal condition, much less to evaluate the relative efficacy or effectiveness of any particular device.

Breaking the cycle of unwarranted variation. Left alone, practice variations do not go away. Intervention is needed at the level of the doctor-patient relationship to reduce the role of medical opinion and enhance the role of the patient in choice of treatments. Intervention is also needed to improve the scientific basis for clinical decision making through outcomes research.

The role of shared decision making and patient decision aids. The cycle of unwarranted variation attributable to failure to understand patient preference can be interrupted by replacing delegated decision making by shared decision making—a process of interacting with patients to arrive at informed, value-based decisions when more than one treatment option exists. Clinical trials show that PtDAs are effective in promoting shared decision making. Moreover, they are cost-effective and often lead to reduced use of more invasive treatments. However, to achieve wide implementation, barriers to their use must be overcome.

One barrier is the lack of instruments for objectively evaluating the quality of the decision process: Is shared decision making taking place? Are patients making informed choices among treatment options that reflect their own individual values? However, decision aids need to be available to patients in “just in time” fashion. This means the redesign of clinical processes around the requirements for shared decision making, which will require information, favorable economic incentives, and clinical leadership. We are encouraged that Section 646 of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003 directs the CMS to work with providers to redesign the FFS reimbursement system to reward providers who implement shared decision making.

We believe that the process of change can be accelerated through the feedback of information that draws attention to variation. But to succeed, feedback needs to occur within a framework of advocacy and a call for professional action. The sponsorship by the American Academy of Orthopedic Surgeons of a musculoskeletal edition of the Dartmouth Atlas of Health Care and the academy’s advocacy of shared decision making is an important example. It also requires advocacy and action on the part of payers. We thus recommend that in addition to the 646 demonstration project, the CMS pilot-test the idea that feedback of utilization data “in real time” by Medicare’s Quality Improvement Organizations (QIOs) would accelerate the adoption of shared decision making. The development of standardized measures of decision quality would increase the CMS’s ability to monitor how well this process is being implemented.

Improving the scientific basis of clinical decision making. Clinical studies of surgical interventions can be designed to evaluate the outcomes of alternative treatments and, at the same time, address fundamental questions concerning the influence of patient preference on outcomes. The Spine Patient Outcomes Research Trial (SPORT) is an example of such a trial. Funded by the National Institute for Arthritis and Musculoskeletal and Skin Diseases (NIAMS: U01-AR45444-01A1), this eleven-state, multicenter trial is both a patient preference observational trial and a randomized trial of patients with the three most common conditions for which back surgery is performed: herniated disc, spinal stenosis, and degenerative spondylolisthesis with stenosis. Eligibility criteria are set broadly to include most patients offered surgery for these conditions at the eleven sites.

Decision aids provide patients with a full understanding of what is reasonably well understood about expected outcomes of care as well as the current limits of scientific knowledge. Patients are then offered enrollment in a clinical trial; those with strong treatment preferences who do not want to enter the randomized trial are asked to enroll in a cohort study. The enrollment process thus involves a shift from a clinical trial ethic that depends on physicians’ recognition of scientific uncertainty to one that seeks enrollment based on informed patients who understand that clinical science
has proved neither operative nor nonoperative treatments to be better. Shifting the locus of authority from physician to patient has had a major impact on enrollment. Physician-dependent surgical trials that depend on physician equipoise are difficult to organize and conduct because, understandably, surgeons tend to be strongly committed to surgical theories. Using the patient-centered standard, SPORT shows that about 38 percent of those who see the video accept randomization, and the enrollment goal of 1,170 randomized patients has nearly been achieved.

SPORT provides a notable example of how everyday clinical practice can be adapted into a “laboratory” to evaluate the common practices of medicine and improve their scientific base. The importance of systematic evaluation of medical and surgical innovations is underscored by the emergence of many new drugs and devices, including drug-eluting stents and new biological materials, all hypothesized to improve the quality of life, yet which often increase the cost of care. These innovations need careful, timely evaluation.

But expansion of the research agenda to encompass the early evaluation of new technologies as well as new theories on how to use existing technologies will need a supportive federal science policy. The mobilization of talent and focus of interest required to meet the larger task of improving the scientific basis of everyday practice will require the active participation of the National Institutes of Health (NIH) and the nation's academic medical centers. NIAMS' sponsorship of SPORT is a welcome step in the right direction. But the evaluation agenda extends beyond the reach of a single NIH institute. An opportunity to take action may rest in the “roadmap” initiative of the NIH director: the effort on the part of Elias Zerhouni to reach across the individual agencies to identify major gaps and opportunities in research that no single institute could tackle.10 One of these initiatives is the reengineering of the clinical research enterprise, which, we believe, is inevitably about patient preference and practice variation. But for such research to be successful, clinicians and patients must buy in. The reengineering required for evaluation research must therefore move beyond the confines of the laboratory or even the wards of single institutions to involve patients in everyday practice.

The authors acknowledge funding from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (U01-AR45444-01A1 and P60-AR048094-01A1) and Office of Research on Women’s Health, National Institutes of Health; National Institute of Occupational Safety and Health, Centers for Disease Control and Prevention; and the Robert Wood Johnson Foundation.

NOTES


See K.R. Sepucha, F.J. Fowler Jr., and A.G. Mulley Jr., “Policy Support for Patient-Centered Care: The Need for Measurable Improvements in Decision Quality,” Health Affairs, 7 October 2004, content.healthaffairs.org/cgi/content/abstract/hlthaff.var.54.


