Evaluating Clinical Practice Guidelines Based on Their Association with Return to Work in Administrative Claims Data

Eric T. Roberts, Eva H. DuGoff, Sara E. Heins, David I. Swedler, Renan C. Castillo, Dorianne R. Feldman, Stephen T. Wegener, Vladimir Canudas-Romo, and Gerard F. Anderson

Objective. To examine the association between non-adherence to clinical practice guidelines (CPGs) and time to return to work (RTW) for patients with workplace injuries.

Data Sources/Study Setting. Secondary analysis of medical billing and disability data for 148,199 for shoulder and back injuries from a workers' compensation insurer.

Study Design. Cox proportional hazard regression is used to estimate the association between time to RTW and receipt of guideline-discordant care. We test the robustness of our findings to an omitted confounding variable.

Data Collection. Collected by the insurer from the time an injury was reported, through recovery or last follow-up.

Principal Findings. Receiving guideline-discordant care was associated with slower RTW for only some guidelines. Early receipt of care, and getting less than the recommended amount of care, were correlated with faster RTW. Excessive physical therapy, bracing, and injections were associated with slower RTW.

Conclusions. There is not a consistent relationship between performance on CPGs and RTW. The association between performance on CPG and RTW is difficult to measure in observational data, because analysts cannot control for omitted variables that affect a patient's treatment and outcomes. CPGs supported by observational studies or randomized trials may have a more certain relationship to health outcomes.

Key Words. Administrative data uses, observational data, clinical practice guidelines, biostatistical methods

The Institute of Medicine defines clinical practice guidelines (CPGs) as “recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options” (Institute of Medicine 2011). It is generally thought

that medical care consistent with CPGs is better and improves health outcomes. Health insurers, including Medicare, are testing programs that tie payments to performance on measures of care quality (Ryan 2009; Kruse et al. 2012; CMS 2012).

Despite growing interest in using CPGs, there is limited evidence that guideline-concordant care actually improves patient outcomes. For certain severe conditions, guideline adherence has been studied in relation to hard endpoints like mortality (Cheng et al. 2009; Ryan 2009; Boland et al. 2012; Jha et al. 2012). Few studies have assessed the relationship between guideline adherence and return to health (Dickey et al. 2006; Fritz, Cleland, and Brennan 2007; Lugtenberg, Burgers, and Westert 2009; Stiles et al. 2009). Given the administrative burden on patients and the cost to researchers, it is not surprising that these studies use relatively small samples. To obtain larger samples, some CPGs are evaluated using secondary data sources, such as claims data (Jha et al. 2012; Webb et al. 2012). However, these datasets typically provide limited information about a patient's long-term health.

Workers' compensation insurers provide a unique source of data to track the effects of adherence to practice guidelines on long-term health outcomes. These insurers document medical care expenses and wage replacement costs related to a workplace injury. They also track the time from a worker's injury to the date he or she returns to work. From the worker's perspective, going back to work signals a return to normal functioning. Returning to work may also bring about further improvements in psychological health (Rueda et al. 2012). This makes return to work (RTW) an important indicator of an employee's recovery from an injury.

In the United States, workplace-related injuries and health issues are typically covered by workers' compensation insurance (National Academy of

Address correspondence to Eric T. Roberts, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Room 307-A, Boston, MA 02115-5899; e-mail: roberts@hcp.med.harvard.edu. Eva H. DuGoff, Ph.D., M.P.P., is with the Department of Population Health Sciences, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI. Sara E. Heins, B.A. and Gerard F. Anderson, Ph.D., are with the Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. David I. Swedler, M.P.H., Ph.D., is with the University of Illinois at Chicago School of Public Health, Chicago, IL. Renan C. Castillo, Ph.D., is with the METRC Coordinating Center, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. Dorianne R. Feldman, M.D., M.S., is with Johns Hopkins Hospital, Baltimore, MD. Stephen T. Wegener, M.A., Ph.D., is with the Johns Hopkins University School of Medicine, Baltimore, MD. Vladimir Canudas-Romo, Ph.D., is with the Department of Population, Family, and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Social Insurance 2011). About 4 million incidents of workplace injury or illness were reported in 2011 (U.S. Department of Labor 2012). Workers' compensation insurance is regulated at the state level. In general, patients face no medical cost sharing. While some states permit insurers in this market to engage in care management, their ability to enforce care guidelines or otherwise limit care is statutorily limited (Tanabe 2011). Consequently, patient care is highly determinant on clinical judgment and expertise.

The few existing studies that have examined CPGs and health outcomes for injured workers have found that guideline adherence is associated with faster RTW (Webster, Verma, and Gatchel 2007; Volinn, Fargo, and Fine 2009; Webster and Cifuentes 2010; Webster et al. 2011). Because CPGs are specific to particular clinical conditions, these studies focus on a narrow set of conditions and have limited timeframes. For example, Webster et al. (2011) assessed the relationship between RTW and physical therapy practice for patients with meniscal knee injuries who received surgical treatment. We are not aware of any study that has evaluated RTW for multiple guidelines.

This study explores the association between receiving care outside of CPGs and RTW for individuals with back and shoulder injuries, which are two common work-related injuries. We focus on guideline-discordant care to identify processes of care that may have unintended, adverse effects on patient outcomes, which we measure using the RTW variable. Guideline-discordant care was determined by reviewing occupational health guidelines and published studies, along with the input of expert clinical panels. We examine 13 CPGs within 12 subconditions of shoulder injuries and 15 CPGs for 10 subconditions of back injuries. Time to RTW is the outcome of interest.

We examine which guidelines have the strongest clinical and empirical relationship to worker recovery, based on their evidence base and association with RTW, which we estimate using a large database of workers' compensation cases. In keeping with the Institute of Medicine's recommendation to evaluate CPGs in light of potential biases, we consider the strength of association between receiving guideline-discordant care and RTW when accounting for unobserved sources of bias (Institute of Medicine 2011). Specifically, we examine whether differences in RTW could be driven by unobserved differences between patients instead of differential rates of guideline adherence that are due solely to provider judgement. Based on these findings, we consider whether payers and policy makers can meaningfully use guidelines to improve processes of care, and whether it is feasible to use standard administrative databases to evaluate and reward providers' adherence to guidelines.

Clinical Practice Guidelines Framework

CPGs are intended to guide medical decision making. The Institute of Medicine's 2011 report, "Clinical Practice Guidelines We Can Trust," defined trustworthy CPGs as recommendations based on a transparent review of evidence that accounts for different research and stakeholder perspectives, and that disclose biases and methodological weaknesses (Institute of Medicine 2011). However, the IOM did not recommend that CPGs should demonstrate that they actually improve health outcomes. To examine this component alongside the IOM's other recommendations, our study evaluates each CPG based on three criteria: (1) the depth of pre-existing evidence on the guideline, including observational or randomized studies supporting the guideline; (2) the strength and consistency of the estimated relationship between receiving guideline-discordant care and RTW for different subgroups of injured patients; and (3) whether our findings are robust once we consider the potential for possible omitted variable bias.

METHODS

Design

This is a retrospective, observational cohort analysis, using data from a large U.S. workers' compensation insurer.

Data Sources

We use the insurer's administrative, medical billing, care management, and disability records. The administrative data include demographic information such as age, marital status, occupational, and employment characteristics. The billing data include all procedures and diagnoses reported on final paid medical bills. The disability file describes each employee's work status following injury, including dates off and on work, and whether the worker returned with full or modified duties.

Study Population

We identified shoulder and back injuries resulting in lost work time that occurred between January 1, 2000, and December 31, 2009, in the United States. We use data collected from 2000 to 2010 to measure the care these workers received, and if and when the individual returned to work.

It is possible for a person to have multiple injuries from incidents that occurred at different times. Each unique incident is assigned a different case identifier. Therefore, the unit of analysis is a workers' compensation case. The main analytic sample consists of cases with any lost work time, encompassing 46,742 shoulder and 101,457 back cases. This represents 46,032 and 99,490 unique persons with shoulder and back injuries, respectively.

Injury Categorization

Cases were categorized into injury groups based on medical billing record diagnosis codes. We first categorized cases by broad anatomical region (shoulder and back) using all available diagnosis codes. Back injuries included the thoracic and lumbar spine. We then grouped cases into 12 subgroups for shoulder injuries and 10 subgroups for back injuries, to obtain more homogenous groups. The first 6 months of diagnosis codes on medical billing records were used to identify the appropriate subgroup (Table 1). (The ICD-9 codes used to classify cases are included in Appendix Table SA-2.) Hereafter, we refer to these subgroups as back and shoulder injury subconditions. This classification approach has been used previously in these data (Heins et al. 2013).

Selection of Clinical Practice Guidelines

The selection and development of guidelines was intended to cover patterns of medical care commonly used among the back and shoulder claims that we analyze. Development of the guidelines proceeded as follows. First, a physical medicine and rehabilitation physician on the study team (not employed by the insurer) identified a large set of CPGs for the treatment of occupational back and shoulder injuries. All guidelines published by the American College of Occupation and Environmental Medicine (ACOEM) and Work Loss Data Institute's Office of Disability Guidelines (American College of Occupational and Environmental Medicine 2008; Work Loss Data Institute 2011) were consulted in this initial review. When there was no guideline corresponding to a common process of care, when an existing guideline did not apply to a particular back or injury subcondition, or when the guideline, as written, could not be evaluated with the available data, the physician either proposed a new guideline or modified an existing guideline based on clinical judgment. This general approach is consistent with other formative studies of clinical care processes (McGlynn et al. 2003).

Table 1: Number of Cases and Distribution of Injury Type (Subgroup) by Shoulder and Back Injury

<i>Shoulder Injuries</i>		<i>Sample Size (Col %)</i>	<i>Back Injuries</i>		<i>Sample Size (Col %)</i>
1	Bursitis/tendonitis	15,523 (33.2)	1	Degenerative back conditions with myelopathy	99 (0.1)
2	Frozen shoulder	1,547 (3.3)	2	Degenerative back conditions without myelopathy	19,438 (19.2)
3	Arthritis	1,938 (4.1)	3	Postlaminectomy syndromes	153 (0.2)
4	Complete rotator cuff tear	2,478 (5.3)	4	Thoracic and lumbar disc disorders	5,614 (5.5)
5	Dislocation	2,537 (5.4)	5	Thoracic or lumbosacral neuritis or radiculitis	12,482 (12.3)
6	Fracture	2,200 (4.7)	6	Sprain/strain	48,826 (48.1)
7	Labral lesion	1,364 (2.9)	7	Complex sprain/strain	2,290 (2.3)
8	Calcifying tendinitis	1,033 (2.2)	8	Generalized back pain only	8,346 (8.2)
9	Sprain/strain	12,674 (27.1)	9	Complicated cases	3,035 (3.0)
10	Complex sprain/strain	1,951 (4.2)	10	Degenerative back conditions with myelopathy (no decompression/fusion surgeries, diagnosis unclear)	1,174 (1.2)
11	Generalized shoulder pain only	2,391 (5.1)			
12	Complicated cases	1,106 (2.4)			

Appendix Tables SA-1.1 and SA-1.2 show sample sizes for each CPG-subcondition pair.

Second, the physician met with the study’s principal investigator and a data analyst to operationalize the guidelines based on the data provided by the insurer. A hospital billing specialist was engaged, when needed, to identify billing codes for procedures included in each guideline. These first two steps preliminarily identified approximately 50 back and shoulder guidelines for further review.

Third, two clinical panels were convened to review this set of 50 guidelines. The panels consisted of the physical medicine and rehabilitation physician who conducted the original review, a surgeon (specializing in shoulders for the shoulder panel, and backs in the back panel), an internist, a clinical psychologist, a radiologist, and a physical therapist. The panels edited and selected guidelines from the initial set of guidelines and suggested additional guidelines. Ultimately, the shoulder panel reached consensus that 23 CPGs were good clinical practice. The back injury panel identified 24 CPGs as good clinical practice. The researchers finalized the set of guidelines to evaluate

based on the recommendations of the panels. The guidelines neither were developed nor modified based on any evaluation of worker outcomes, and the insurer did not direct the development or evaluation of the CPGs. Table 3 provides an overview of the CPGs evaluated in this study. Appendix Table SA-3 provides more detail on the sources consulted to develop each guideline, identifies the diagnosis and procedure codes for each guideline, and discusses relevant limitations.

For this study, we analyze 13 CPGs for shoulder and 15 CPGs for back injuries. We focus on the relationship between the receipt of guideline-discordant care and RTW. We selected these guidelines from the list identified by the panel based on two criteria. First, sample sizes needed to be sufficient to conduct statistical analyses. Second, where the panel recommended several similar CPGs related to the same process of care, we present only one guideline in the results. We verified that the conclusions were similar across the closely related guidelines that are not reported.

Outcome Measure

The primary outcome measure was the number of days from injury to the last reported RTW date for a claim.

Severity of Injury

We use four measures to assess severity and complexity: Abbreviated Injury Scale (AIS), number of chronic conditions, counts of hospitalizations in the first 30 days, and number of prior workers' compensation cases. An anatomic region-specific AIS, which is a 0–6 scale increasing in injury severity, was calculated based on the 27 most frequently occurring diagnoses for each case (a score of 0 represents an injury of insufficient severity to be measured on the scale; Baker et al. 1974). We use the upper extremity AIS for shoulder cases and the spine AIS for back cases. We included an indicator for whether a patient had 0, 1, or 2 or more chronic conditions, calculated using Clinical Classification System and Chronic Condition Indicator software provided by the Agency for Healthcare Quality and Research (Hwang et al. 2001). Hospitalization within the first 30 days of injury was determined using dates of service, injury date, and place of service. Lastly, we created a dichotomous indicator for whether a person had any prior workers' compensation cases covered by this insurer during the study period.

Case Characteristics

We categorized age into 5-year groups, grouping a small number of cases with unreported or implausible ages into separate categories. Marital status was grouped into married, not married, or unknown. Average weekly wages, prior to the worker's injury, were grouped into quartiles. Industry classification was grouped into eight broad industry groups based on the North American Industry Classification System (U.S. Census Bureau 2012). Workers' state of residence at the time of injury was grouped into four regions.

Statistical Analysis

We assessed whether receipt of care outside of guidelines was associated with time to RTW for CPGs and specific subgroups of back and shoulder injuries. For each CPG and injury subcondition pair, we estimated a multivariate Cox proportional hazards model of the form:

$$\lambda(t|D, X) = \lambda_0(t)e^{\gamma D + \beta' X}$$

where $\lambda_0(t)$ is the baseline probability of returning to work at time t , D is a dichotomous indicator of non-compliant care for a particular CPG, and X is a vector of control variables consisting of the baseline injury severity and case characteristics described above. The coefficient of interest is γ , which represents the proportional effect of receiving guideline-discordant care on the probability of returning to work at any time t . Using the opioid guideline as an example, if γ is negative (positive), the model predicts an $\exp(\gamma)$ factor reduction (increase) in the probability of RTW at any time t following injury, if a patient receives more than 3 months of opioids. An assumption of the model is that this proportional hazard is constant over time.

We performed four robustness tests. First, we compared results from the proportional hazard regressions with a non-parametric Kaplan Meier survival analysis, which assessed the difference in probabilities of RTW by guideline compliance status. Second, we estimated multivariate logistic regressions of work status at 90 and 180 days following injury. Third, we re-estimated the proportional hazard models, including only cases with at least 7 days of lost work time. This enabled us to check whether the associations found in our base models persisted among more severely injured workers.

Lastly, we assessed the robustness of each CPG analysis to the inclusion of a simulated variable that is related to both CPG compliance and RTW—that is, a confounder. The rationale for this analysis is that claims data

frequently omit, or provide limited information about, factors that may affect a patient's treatment and RTW. Omission of such a variable from the regressions could bias our results. Moreover, it may not be feasible to hold a physician accountable for providing care, or for a health outcome, caused by clinical characteristics that are not captured in administrative data, and which an analyst cannot control for.

Methodologically, our approach builds from a sensitivity analysis technique developed by Greenland (1996) and Liu, Kuramoto, and Stuart (2013). In this procedure, we first estimate propensity scores for receiving guideline-discordant care, as a function of all control variables included in the main regression models. Then, within each propensity score quintile, we do the following sub-analysis. First, we construct a binary variable U , which might be a measure of the patient's pain (high or low), since this is not reported in the data (McGorry, Shaw, and Lin 2011). We simulate U by letting $Pr(U|D)$ take on one of three values between 0.1 and 0.9, and by letting $Pr(Y|D,U)$ take on one of five values over the same interval. This yields 15 modeled values of U , each of which is used to construct a synthetic dataset that reflects the assumed relationship between U , D , and Y . Second, we use this synthetic dataset to estimate a logistic regression of Y on U and D . This calculation is performed once for each propensity score quintile from a particular CPG-subcondition pair.

After completing this process for each of the 15 modeled values of D , we assess whether the direction (i.e., having a positive or negative sign on the logarithmic scale) and statistical significance of D , across the propensity score quintiles, remains unchanged from our estimates in the base set of regressions, where we did not adjust for U . We consider D to be robust in the main regressions if the sign and statistical significance on D remained unchanged in all of the sensitivity analyses for the corresponding CPG-injury subcondition examined in the sensitivity analysis.

RESULTS

Table 1 presents the distribution of injury subconditions, and Table 2 describes the characteristics of our sample. The most common shoulder injury was bursitis/tendonitis, while sprains and strains accounted for nearly one-half of back injuries. The majority of claimants are male, and the plurality worked in heavy industries (e.g., construction or manufacturing) at the time of injury. A small, but not insignificant, proportion of

Table 2: Sample Characteristics for Shoulder and Back/Spine Cases

	<i>Shoulder Cases</i> <i>N (Col %)</i>	<i>Back/Spine Cases</i> <i>N (Col %)</i>
<i>N</i>	46,742	101,457
Case characteristics (at the time of injury)		
Gender		
Female	15,056 (32.2)	32,318 (31.9)
Male	31,500 (67.4)	68,667 (67.7)
Unknown	186 (0.4)	472 (0.5)
Age		
19–28	5,840 (12.5)	20,061 (19.8)
29–38	9,667 (20.7)	28,730 (28.3)
39–48	13,413 (28.7)	29,054 (28.6)
45–54	12,051 (25.8)	16,935 (16.7)
49–58	4,549 (9.7)	4,428 (4.4)
59–68	632 (2.4)	511 (1.0)
≥69	590 (1.3)	1,738 (1.7)
Unknown	590 (1.3)	1,738 (1.7)
Marital status		
Not married	28,357 (60.7)	67,859 (66.9)
Married	18,378 (39.3)	33,576 (33.1)
Unknown	7 (0.0)	22 (0.0)
Wage quartiles		
1 st	10,022 (21.4)	27,282 (26.9)
2 nd	11,218 (24.0)	27,399 (27.0)
3 rd	12,086 (25.9)	24,892 (24.6)
4 th	13,392 (28.7)	21,818 (21.5)
Industry		
Agriculture	493 (1.1)	1,212 (1.2)
Health, education, arts	5,149 (11.0)	13,641 (13.4)
Heavy industry	16,225 (34.7)	31,071 (30.6)
Office	6,333 (13.6)	13,962 (13.8)
Services	2,689 (5.8)	6,826 (6.7)
Trade	3,006 (6.4)	7,386 (7.3)
Transportation	8,063 (17.2)	17,473 (17.2)
Missing	4,784 (10.2)	9,886 (9.7)
Region		
Midwest	9,100 (19.5)	15,476 (15.2)
Northeast	12,084 (25.9)	25,483 (25.1)
South	10,513 (22.5)	23,675 (23.3)
West	15,045 (32.2)	36,823 (36.3)
Injury characteristics		
Upper extremity AIS score (mean)	1.081	N/A
Spine AIS score (mean)	N/A	0.8894
Previous workers' compensation cases (% with ≥1) [†]	15.3	13.6
Hospitalization in 30 days (% with ≥1)	1.9	1.4

continued

Table 2. *Continued*

	<i>Shoulder Cases N (Col %)</i>	<i>Back/Spine Cases N (Col %)</i>
Chronic conditions		
0	42,579 (91.1)	87,657 (86.4)
1	2,395 (5.1)	7,885 (7.8)
2+	1,768 (3.8)	5,915 (5.8)
Lawyer representation (%)	27.2	33.9
Post-injury characteristics		
Case management (%)	49.6	41.9
Claim results in at least 7 days of lost work time (%)	95.5	93.5
Return to work in ≤90 days (%)	31.8	47.6
Return to work in ≤180 days (%)	48.5	58.5

[†]Although 15.3% of shoulder and 13.6% of back cases had prior claims with this workers' compensation insurer, not all prior claims were shoulder or back cases. Consequently, the number of unique persons with back injuries ($N = 99,490$) is only 1.9% smaller than the sample size of back claims, and the number of unique persons with shoulder injuries is only 1.5% ($N = 46,032$) smaller than the sample of shoulder claims.

shoulder (15.3%) and back (13.6%) cases had a previous workers' compensation claim with this insurer. Overall, only 31.8% of shoulder cases returned to work within 90 days, but nearly half of back cases did. Among all shoulder cases, 95.5% resulted in at least 7 days of lost work, while 93.5% of back cases incurred at least 7 days of lost work time (recall that the main sample included cases with at least some lost work time).

Table 3 describes the CPGs, the subconditions to which they apply, and the evidence base supporting each guideline. We group CPGs into four domains: early use of care, inappropriate care, overuse, and underuse. Of the 17 unique CPGs in this analysis, three were published by ACOEM or by the Work Loss Data Institute's Official Disability Guidelines (ODG). Three CPGs were based on one or more randomized controlled trials or more than two observational studies. In some cases, these studies focused on a limited number of subconditions, and the panels advised us if the guideline applied to additional subconditions. Twelve CPGs were based primarily on the expert opinion of the panels. The far right column shows compliance rates for each CPG. Consistent with prior studies, we found wide variation in rates of non-adherent care, ranging from 0.9% (epidural injections for backs), to 70.3%, (maximum physical therapy for shoulders) (McGlynn et al. 2003).

Table 4 (shoulder injuries) and Table 5 (back injuries) present logarithmic-scale hazard ratios for the estimated association between guideline-discordant care and RTW. Coefficients significantly greater than zero indicate that

Table 3: Description of Clinical Practice Guidelines

Treatment Guideline	Description of Guideline-Discordant Care	Subconditions to Which Guideline Applies*		Evidence Base (References Cited in Columns)**		Rates of Guideline-Discordant Care†	
		Shoulder	Back	Observational Studies	ACOEM/ODG	Shoulder	Back
Early care‡ Minimum time to MRI	Receiving an MRI within 4–24 weeks of injury (time depends on injury subcondition)	1,2,3,8,9,11	2,3,4,5,6,8	≥1 RCT or ≥2 Observational Studies		56.6%	31.1%
	Receiving a CT scan within 4–24 weeks of injury (time depends on injury subcondition)	INS	2,3,4,5,6,8			24.8%	21.2%
X-ray versus MRI order	Receiving an MRI before or instead of an X-ray	All except 7	2,6,8			26.4%	30.7%
	Receiving a CT scan before or instead of an X-ray	INS	2,6,8			8.4%	20.6%
Minimum time to specialist referral Surgery before PT	Having a specialist visit within 4 weeks of injury	1,2,3,8,9,11	2,6,8			29.0%	24.2%
	Receiving surgery before physical therapy	1,2,3,8,9,11	2,3,4,5,6,8			45.8%	32.8%
Minimum time to surgery	Receiving surgery within 8–12 weeks of injury (time depends on injury subcondition)	1,2,3,8,9	2,3,4,5			33.8%	22.1%

continued

Table 3. Continued

Treatment Guideline	Description of Guideline-Discordant Care	Subconditions to Which Guideline Applies*		Evidence Base (References Cited in Columns)**		Rates of Guideline-Discordant Care†	
		Shoulder	Back	≥1 RCT or ≥2 Observational Studies	ACOEM/ODG	Shoulder	Back
Inappropriate care							
Overall active PT ratio	Receiving more passive than active PT	All	All			21.1%	29.6%
Increasing active to passive PT ratio	Receiving a lower proportion of active PT in 2nd half of treatment than in 1st half	All	All			28.5%	34.6%
	Receiving a shoulder brace for a particular subcondition	1,2,3,8,11	N/A	x	x	14.8%	N/A
Shoulder bracing ^s							
Shoulder steroid injections	Receiving shoulder injections for a particular subcondition	5,6,8,9	N/A	x	x	22.0%	N/A
Epidural injections	Receiving an epidural injection for a particular subcondition	N/A	1,6,8		x	N/A	0.9%
	Receiving a trigger point injection for a particular subcondition	N/A	1,3,4,5			N/A	1.9%
Trigger point injections							

continued

Table 3. Continued

Treatment Guideline	Description of Guideline-Discordant Care	Subconditions to Which Guideline Applies*		Evidence Base (References Cited in Columns)**		Rates of Guideline-Discordant Care†	
		Shoulder	Back	≥1 RCT or ≥2 Observational Studies	ACOEM/ODG	Shoulder	Back
Over-utilization‡							
Maximum time on opioids	Continuously§ receives opioids for longer than 3 months	All	All	x		18.7%	22.8%
Maximum PT	Receives more than 75% above the minimum PT requirement	All	All			70.3%	66.0%
Under-utilization							
Need imaging to diagnose	Does not receive imaging to properly diagnose the subcondition	3,4,5,6,7,8,10,12	1,2,3,4,5,7,9			7.1%	8.8%
Minimum PT	Does not receive a minimum number of PT visits	All	All		x	52.9%	66.7%

Notes. INS indicates that there was not a sufficiently large sample in any subcondition of patients with shoulder injuries to analyze the relationship of minimum time to CT scan and CT scan/X-Ray order on RTW.

Guidelines not based on an RCT, observational study, or not recommended by ODG or ACOEM were recommended based on the clinical expertise of the panels. In some cases, studies or published guidelines did not address all subconditions; guidelines for the remaining subconditions were determined by panel recommendation. Appendix Table SA-1 shows sample sizes for each CPG-subcondition pair.

*Numbering of subconditions refers to Table 1.

**Where applicable, references for each guideline are provided in Appendix Table SA-3.

†Rate of non-compliance across all subconditions to which the guideline applies.

‡Those who did not receive the indicated modality of care were excluded from the reference group.

§Excludes post-arthroplasty arthritis claimants.

¶Defined as an average of at least one prescription per month.

ACOEM, American College of Occupational and Environmental Medicine; IME, Independent Medical Evaluation; ODG, Official Disability Guidelines; PT, physical therapy.

Table 4: Hazard Ratios Multivariate Cox Proportional Hazard Regression Analysis of Time to Return to Work on CPG-Discordant Care, by *Shoulder Injury Subcondition*

<i>Domain of Inappropriate Care: Subcondition</i>	<i>CPGs Related to Early Care</i>					
	<i>Minimum Time to MRI</i>	<i>X-Ray versus MRI Order</i>	<i>Minimum Time to Specialist Referral</i>	<i>Surgery before PT</i>	<i>Minimum Time to Surgery (12 Weeks)</i>	
	<i>Faster</i>	<i>Inconclusive</i>	<i>Faster</i>	<i>Faster</i>	<i>Faster</i>	
<i>General Association of Non-Compliant Care with RTW[†]</i>						
Bursitis/tendonitis	0.412**	0.086**	0.443**	0.417**	0.788***‡	
Frozen shoulder	0.399**	−0.102	INS	INS	INS	
Arthritis	INS	0.032	0.517**	0.477**	0.813**	
Complete rotator cuff tear	−	0.043	−	−	−	
Dislocation	−	INS	−	−	−	
Fracture	−	INS	−	−	−	
Labral lesion	−	−	−	−	−	
Calcifying tendinitis	0.593**	INS	INS	INS	INS	
Sprain/strain	0.691**	0.058	0.565**	0.342**	1.217**	
Complex sprain/strain	−	0.097	−	−	−	
Generalized shoulder pain only	INS	−0.036	INS	INS	−	
Complicated cases	−	0.042	−	−	−	

continued

Table 4. Continued

Domain of Inappropriate Care: Subcondition	CPGs Related to Inappropriate Care				CPGs Related to Over-Utilization			CPGs Related to Under-Utilization	
	Overall Active PT Ratio	Increasing Active to Passive PT Ratio		Shoulder Steroid Injections	Maximum Time on Opioids (3 months)	> 75% More Than Minimum PT	Imaging Needed to Diagnose	Minimum Physical Therapy	
		Inconclusive	Inconclusive						Slower
General Association of Non-Compliant Care with RTW [†]	Inconclusive								
Bursitis/tendonitis	0.117**	-0.047	-0.257**,\$	-	-0.206**	-0.649**,\$	-	0.374**	
Frozen shoulder	-0.026	0.129	-0.172	-	INS	-0.486**	-	0.291**	
Arthritis	0.042	-0.003	-0.156*	-	-0.190	-0.584**	INS	0.297**	
Complete rotator cuff tear	0.054	-0.063	-	-	-0.216*	-1.096**	INS	0.447**	
Dislocation	-0.019	-0.111	-	-0.569**	INS	-0.792**	INS	0.606**	
Fracture	-0.219*	-0.163*	-	-0.614**	INS	-0.492**	0.266**	0.403**	
Labral lesion	INS	0.006	-	-	-0.488**	INS	INS	0.210**	
Calcifying tendinitis	0.055	-0.056	INS	-0.314**	INS	INS	INS	0.296**	
Sprain/strain	0.036	-0.054	-	-0.471***,‡	-0.575**	-0.599**,\$	-	0.499***,‡	
Complex sprain/strain	-0.027	-0.100	-	-	-0.188*	INS	INS	0.378**	
Generalized shoulder pain only	-0.045	-0.012	INS	-	INS	-0.145	-	0.299**	
Complicated cases	INS	-0.069	-	-	INS	INS	INS	0.108	

Notes. INS indicates that there was not a sufficiently large sample (at least 200 observations per compliant and non-compliant group in a given CPG-injury subcondition pair) to estimate the model. Coefficients reported on the logarithmic scale. A positive coefficient indicates that guideline-discordant care is associated with faster return to work, while a negative coefficient indicates that guideline-discordant care is associated with slower return to work. All models adjust for hospitalization within 30 days, injury severity (upper extremity AIS score), previous injuries, number of chronic conditions, wages at the time of injury, industry, age at the time of injury, gender, marital status at the time of injury, and region (midwest, northeast, south, and west). (-) Indicates that the CPG was not applicable for the given subcondition.

[†]When sample sizes are sufficient.

[‡]Finding is robust to simulation of an unobserved confounder at the 95% confidence level.

[§]Finding is robust to simulation of an unobserved confounder at the 90% confidence level.

p* < .01; *p* < .05.

Table 5: Hazard Ratios from Multivariate Cox Proportional Hazard Regression Analysis of Time to Return to Work on CPG-Discordant Care, by *Back Injury Subcondition*

<i>Domain of Inappropriate Care: Subcondition</i> <i>General Association of Non-Compliant Care with RTW[†]</i>	<i>CPGs Related to Early Care</i>							
	<i>Minimum Time to MRI</i>		<i>Minimum Time to CT Scan</i>		<i>X-Ray versus MRI Order</i>		<i>X-Ray versus CT Order</i>	
	<i>Faster</i>	<i>Minimum Time to MRI</i>	<i>Faster</i>	<i>Minimum Time to CT Scan</i>	<i>Inconclusive</i>	<i>Faster</i>	<i>Faster</i>	<i>Minimum Time to Specialist Referral</i>
Degenerative back conditions with myelopathy	—	—	—	—	—	—	—	—
Degenerative back conditions without myelopathy	0.342**	0.695**	—	—	0.031	0.150*	0.285**	0.351**
Postlaminectomy syndromes	INS	INS	INS	INS	—	—	INS	INS
Thoracic and lumbar disc disorders	0.403**	INS	INS	INS	—	—	INS	INS
Thoracic or lumbosacral neuritis or radiculitis	0.363**	0.711**	—	—	—	—	0.166	INS
Sprain/strain	0.535**	1.045**	—	—	−0.002	0.277**	INS	—
Complex sprain/strain	—	—	—	—	—	—	—	—
Generalized back pain only	0.641**	INS	INS	INS	0.054	INS	INS	0.717**
Complicated cases	—	—	—	—	—	—	—	—
Degenerative back conditions with myelopathy (no decompression/fusion surgeries, diagnosis unclear)	—	—	—	—	—	—	—	—

continued

Table 5. Continued

Domain of Inappropriate Care: Subcondition General Association of Non-Compliant Care with RTW [†]	CPGs Related to Inappropriate Care				CPGs Related to Over-Utilization		CPGs Related to Under-Utilization	
	Overall Active PT Ratio	Increasing Active to Passive PT Ratio	Epidural Injections	Trigger Point Injections	Maximum Time on Opioids (3 months)	>75% More Than Minimum PT	Imaging Needed to Diagnose	Minimum Physical Therapy
	Inconclusive	Inconclusive	Slower	Slower	Slower	Slower	Faster	Faster
Degenerative back conditions with myelopathy	INS	INS	INS	INS	INS	INS	INS	INS
Degenerative back conditions without myelopathy	0.014	-0.015	-	-	-0.445**	-0.354**	0.199**	0.165**
Postlaminectomy syndromes	INS	INS	-	INS	INS	INS	INS	INS
Thoracic and lumbar disc disorders	0.000	-0.057	-	INS	-0.530**	-0.419**	0.328**	0.239**
Thoracic or lumbosacral neuritis or radiculitis	-0.029	-0.014	-	-0.405**	-0.410**	-0.519**	0.411**	0.265**
Sprain/strain	-0.070**	-0.029	-0.923**	-	-0.750**	-0.528**	-	0.408**
Complex sprain/strain	0.005	-0.062	-	-	-0.320**	INS	INS	0.257**
Generalized back pain only	-0.071	-0.063	INS	-	-0.547**	-0.445**	-	0.369**

continued

Table 5. Continued

	CPGs Related to Inappropriate Care				CPGs Related to Over-Utilization		CPGs Related to Under-Utilization	
	Overall		Increasing Active to Passive PT Ratio		Maximum Time on Opioids (3 months)		> 75% More Than Minimum PT	
	Active PT Ratio	Inconclusive	Inconclusive	Slower	Slower	Slower	Faster	Faster
Domain of Inappropriate Care:								
Subcondition								
General Association of								
Non-Compliant Care with RTW [†]								
Complicated cases	0.065		0.010	–	–0.416**	INS	0.185*	0.224**
Degenerative back conditions with myelopathy (no decompression/fusion surgeries, diagnosis unclear)	–0.071		0.070	–	INS	–0.622**	–	0.199*

Notes. The following CPG-subcondition pairs were robust at the 20% significance level: minimum time to CT scan for sprain/strain; minimum time to surgery for degenerative back conditions without myelopathy; and >75% minimum physical therapy for sprain/strain. (–) Indicates that the CPG was not applicable for the given subcondition. INS indicates that there was not a sufficiently large sample (at least 200 observations per compliant and non-compliant group in a given CPG-injury subcondition pair) to estimate the model. Coefficients reported on the logarithmic scale. A positive coefficient indicates that guideline-discordant care is associated with faster return to work, while a negative coefficient indicates that guideline-discordant care is associated with slower return to work. All models adjust for hospitalization within 30 days, injury severity (spine AIS score), previous injuries, number of chronic conditions, wages at the time of injury, industry, age at the time of injury, gender, marital status at the time of injury, and region (midwest, northeast, south, and west).

[†]When sample sizes are sufficient.

** $p < .01$; * $p < .05$.

receiving care outside of guidelines is associated with a higher probability of RTW, while estimates significantly less than zero indicate that non-compliant care is correlated with a lower probability of RTW.

For shoulder injuries, we found that guideline-discordant care was associated with faster RTW (i.e., patients return to full health faster, on average) for six guidelines and slower RTW for four guidelines. For these 10 CPGs, the direction of effect was consistent across all applicable subconditions. Our findings were inconclusive for three guidelines, where the sign of coefficients varied by subcondition and were not always statistically different from zero. For back injuries, guideline-discordant care was associated with faster recovery for eight CPGs and slower RTW for four guidelines. Relationships were inconsistent for three guidelines. We obtained very similar results when estimating the models among claims with at least 7 days of lost work time (see Appendix Tables SA-4 and SA-5). Our results were also consistent with estimates from both the logistic regression and Kaplan-Meier survival analyses (not shown).

In general, receipt of guideline-discordant care for the early-use and under-use CPGs was correlated with faster RTW. Over-use of physical therapy and opioids, as well as the receipt of inappropriate bracing, injections, and physical therapy, was associated with slower RTW.

While there may be some clinical plausibility to these relationships, finding, for example, that receiving less than the recommended level of care is associated with faster recovery suggests that unobserved differences between the guideline-compliant and non-compliant cases may be driving these results. For example, persons receiving less physical therapy may be less severely injured, in some unobserved way, which is correlated with faster RTW. Failure to control for a confounder of treatment and RTW will bias the regression estimates. We examine this possibility by evaluating whether our estimates are robust to the inclusion of a simulated confounding variable. At the 95% confidence level, we found the model to be robust (at all assumed levels of association between the simulated confounder, guideline non-adherence, and RTW) for only three pairs of CPGs and shoulder injury subconditions: bursitis/tendonitis and time to surgery; sprain/strain and steroid injections; and sprain/strain and minimum physical therapy (marked with a ‡). Of these three, only receipt of a steroid injection was associated with slower RTW. At the 90% confidence level, an additional three CPG-injury subconditions were robust (marked with a §). No back CPG-subcondition pairs were robust at a 90% or greater confidence level. However, for somewhat lower confidence levels, for back sprains and strains, we found evidence to support the association between receipt of excessive physical therapy with slower

RTW, and early receipt of a computed tomography (CT) scan with faster RTW. Thus, for many of the guidelines we evaluate, it is not clear whether the association with RTW reflects the independent effect of guideline non-adherence that could be attributed to clinical judgement, or instead, an unobserved confounder.

DISCUSSION

Medical care that meets CPGs may not result in better patient outcomes, as measured by time to RTW. For the majority of the guidelines we considered, receiving non-recommended care was actually associated with faster RTW. For only four guidelines in each of the shoulder and back samples, we found that receiving non-recommended care was correlated with slower RTW. We often find the counter-intuitive result that the provision of too little care, and the receipt of early care, are associated with faster RTW, when we control for observable baseline characteristics. On the other hand, the receipt of some excessive and inappropriate care is associated with slower RTW.

Of the 17 unique guidelines we consider, only two came close to meeting our assessment criteria: inappropriate receipt of bracing for shoulder injuries with bursitis and tendonitis, and steroid injections for shoulder sprains and strains. The shoulder bracing and steroid injection guidelines, for some subconditions, are well-supported by observational studies (Dameron and Grubb 1981; Ozaki and Kawamura 1984; Shibata et al. 2001; Izquierdo et al. 2010) or are endorsed by ACOEM (American College of Occupational and Environmental Medicine 2008) and the shoulder injection guideline, for some subconditions, is substantiated by randomized controlled trials or observational studies, or is endorsed by ACOEM (Hollingworth, Ellis, and Hattersley 1983). These guidelines exhibited a consistent relationship with RTW, in the expected direction, across all applicable subconditions. However, we note that these guidelines were *not* evaluated, in the scientific literature, for the particular subconditions that passed the simulated confounder robustness test.

While a strong scientific basis for a guideline may be related to its statistical robustness, the simulated confounder analysis showed that even this criterion may not be sufficient to capture causal relationships. As an example, observational studies have documented the association between long-term opioid use and delayed patient recovery (Gross et al. 2009; Volinn, Fargo, and Fine 2009). Ideally, an analyst would have sufficient information to control for patients' baseline case characteristics, to isolate the effect of differences

in providers' practice styles on guideline adherence. In this case, we could reasonably consider observed CPG compliance to be a modifiable decision of health care providers. However, our omitted variable bias test shows that our data do not permit us to distinguish the effect of care based on the clinical decisions of providers from unmeasured differences between patients. Such unmeasured differences could include injury characteristics, pain, or patient-specific preferences for care. These unobserved differences may account for variation in guideline non-adherence, and can differentially affect RTW. Consequently, for most of the CPGs we evaluate, our study does not find a certain relationship between performance on CPGs and worker recovery. Additionally, it may not be administratively feasible to use claims data to evaluate providers based on rates of guideline compliance, given that it may not be possible to empirically distinguish the receipt of clinical processes from confounding variables.

The strength of a CPG comes from the base of scientific support for it *and* the availability of data to control for confounding factors, which is necessary both to develop and operationalize a guideline in a quality evaluation program. Our study suggests that guidelines based on stronger scientific evidence tended to be related to RTW in the expected direction. These guidelines also appeared to be more robust to an omitted confounding variable, although we did not rigorously test this observation on a broad sample of evidence-supported guidelines. In the absence of a randomized trial, guideline developers should carefully consider how to define the comparison group in standard claims or medical record data to minimize unmeasured confounding. Policy makers should be cautious about using such a guideline to influence care processes.

To our knowledge, this is the first study to assess the relationship of CPGs with RTW in a large sample of individuals with shoulder and back injuries. Using a large workers' compensation database, we are able to overcome some of the inherent limitations in traditional commercial health insurance databases, surveys, and registries. We are able to observe cases from 1 to as long as 11 years, providing a better measure of work status for cases with multiple periods out of work. We also evaluate the strength of the estimated association between guideline compliance and RTW, and consider the implications of our findings for guideline developers.

This study has several limitations. First, many of the guidelines evaluated here were tailored to this study setting, and it is not known how they apply to other populations of injured workers. Second, we use data from a single workers' compensation insurer. Although this may limit the generalizabil-

ity of the study to other workers' compensation programs, this insurer held a substantial share of the U.S. workers' compensation market during the years represented in this study, and our sample included cases from nearly every U.S. state.

The worker's compensation population is also substantively different from the general population, which may limit the generalizability of the guidelines to workplace-based private insurance and Medicare. In particular, our sample includes a larger proportion of blue-collar and low-wage workers than is present in the general population. Several studies have found that patients on workers' compensation have worse surgical outcomes compared to comparable patients with other insurance (Harris 2005; Atlas et al. 2010; de Moraes et al. 2012). On the other hand, selection of healthier workers into the workforce may make our sample of claims, at baseline, healthier than the general population.

Third, we consider RTW to be an indicator of functional status, although the validity of this measure as a surrogate for health has been questioned (Steenstra et al. 2011). This outcome measures the date on which a worker's physical functioning had returned sufficiently to enable the resumption of occupational duties. However, measuring RTW dates is complicated by the fact that patients may resume work with modified duties but may not consistently remain working. Resumption of part-time work may indicate an incomplete recovery. Other studies have used time to the first regular work date, among other measures (Baldwin, Johnson, and Butler 1996; Oleinick and Zaidman 2004; Bultmann et al. 2007; Steenstra et al. 2011). We tested the sensitivity of our findings to the definition of the outcome variable by estimating logistic regressions for work status at different points in time, and we obtained results that are consistent with those presented here.

Fourth, we rely on administrative records, which are used primarily for billing and case management. These data provide limited information about workers' occupations, demographic characteristics, and attributes of health that do not relate directly to the injury (such as co-morbidities). While we incorporated different measures of case severity and complexity in our models, we cannot rule out that a confounding variable is missing. Furthermore, differences in case management after the injury could differentially affect RTW. We considered controlling for whether the insurer used care management or if a lawyer was involved in adjudicating a case. We ultimately do not use these variables because of the potential that attorney involvement or case management was initiated *because* a case did not improve over time, a consequence of which would be delayed RTW.

Lastly, researchers have also noted that disability insurance may cause delays in resuming labor force participation (Maestas, Mullen, and Strand 2013). There may even be offsetting effects resulting from incentives to return some individuals to RTW more quickly. However, we do not observe these variables in our data.

CONCLUSION

We find that adherence to some but not all CPGs is associated with better patient outcomes. Patients receiving excessive and inappropriate care had longer periods out of work, while patients receiving less than the recommended amount of care returned to work more quickly. For most guidelines, we find a high degree of susceptibility to bias from an unobserved confounder. Guidelines supported by a randomized controlled trial or several observational studies may have a more certain relationship to health outcomes, even when some patient characteristics, such as injury severity, are unmeasured. Guideline developers should clearly identify the evidence foundation, and carefully define control and treatment groups, when proposing and evaluating care recommendations.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported by a grant from a large U.S. insurer. We would like to thank Elizabeth Stuart for statistical advice and two anonymous reviewers for their helpful feedback on earlier versions of this paper. Eric T. Roberts, Eva H. DuGoff, and David I. Swedler were Ph.D. students in the Department of Health Policy and Management at the Johns Hopkins Bloomberg School of Public Health when this study was conducted.

Disclosures: None.

Disclaimers: None.

REFERENCES

Agency for Healthcare Research and Quality. 2103. "Clinical Classifications Software (CCS) for ICD-9-CM" [accessed on September 1, 2012]. Available at <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>

- American College of Occupational and Environmental Medicine. 2008. *Occupational Medicine Practice Guidelines: Evaluation and Management of Common Health Problems and Functional Recovery of Workers*. Elk Grove Village, IL: OEM Press.
- Atlas, S. J., T. D. Tosteson, E. A. Blood, J. S. Skinner, G. S. Pransky, and J. N. Weinstein. 2010. "The Impact of Workers' Compensation on Outcomes of Surgical and Nonoperative Therapy for Patients with a Lumbar Disc Herniation: SPORT." *Spine (Phila Pa 1976)* 35 (1): 89–97.
- Baker, S. P., B. O'Neill, W. Haddon Jr, and W. B. Long. 1974. "The Injury Severity Score: A Method for Describing Patients with Multiple Injuries and Evaluating Emergency Care." *Journal of Trauma* 14 (3): 187–96.
- Baldwin, M. L., W. G. Johnson, and R. J. Butler. 1996. "The Error of Using Returns-to-Work to Measure the Outcomes of Health Care." *American Journal of Industrial Medicine* 29 (6): 632–41.
- Boland, G. M., G. J. Chang, A. B. Haynes, Y.-J. Chiang, R. Chagpar, Y. Xing, C.-Y. Hu, B. W. Feig, Y. N. You, and J. N. Cormier. 2012. "Association between Adherence to National Comprehensive Cancer Network Treatment Guidelines and Improved Survival in Patients with Colon Cancer." *Cancer* 119 (8): 1593–601.
- Bultmann, U., R. L. Franche, S. Hogg-Johnson, P. Cote, H. Lee, C. Severin, M. Vidmar, and N. Carnide. 2007. "Health Status, Work Limitations, and Return-to-Work Trajectories in Injured Workers with Musculoskeletal Disorders." *Quality of Life Research* 16 (7): 1167–78.
- Cheng, S. H., C. J. Wang, J. L. Lin, C. F. Horng, M. C. Lu, S. M. Asch, L. H. Hilborne, M. C. Liu, C. M. Liu, and A. T. Huang. 2009. "Adherence to Quality Indicators and Survival in Patients with Breast Cancer." *Medical Care* 47 (2): 217–25.
- CMS. 2012. *CMS Makes Changes to Improve Quality of Care During Hospital Inpatient Stays*. Baltimore, MD: Centers for Medicare & Medicaid Services.
- Dameron Jr, T. B., and S. A. Grubb. 1981. "Humeral Shaft Fractures in Adults." *Southern Medical Journal* 74 (12): 1461–7.
- Dickey, B., S.-L. T. Normand, S. Eisen, R. Hermann, P. Cleary, D. Cortes, and N. Ware. 2006. "Associations between Adherence to Guidelines for Antipsychotic Dose and Health Status, Side Effects, and Patient Care Experiences." *Medical Care* 44 (9): 827–34.
- Fritz, J. M., J. A. Cleland, and G. P. Brennan. 2007. "Does Adherence to the Guideline Recommendation for Active Treatments Improve the Quality of Care for Patients with Acute Low Back Pain Delivered by Physical Therapists?" *Medical Care* 45 (10): 973–80.
- Greenland, S. 1996. "Basic Methods for Sensitivity Analysis of Biases." *International Journal of Epidemiology* 25 (6): 1107–16.
- Gross, D. P., B. Stephens, Y. Bhambhani, M. Haykowsky, G. P. Bostick, and S. Rashiq. 2009. "Opioid Prescriptions in Canadian Workers' Compensation Claimants: Prescription Trends and Associations between Early Prescription and Future Recovery." *Spine (Phila Pa 1976)* 34: 525–31.
- Harris, I., J. Mulford, M. Solomon, J. M. van Gelder, and J. Young. 2005. "Association between Compensation Status and Outcome after Surgery: A Meta-Analysis." *JAMA* 293 (13): 1644–52.

- Heins, S., D. Feldman, E. DuGoff, S. Wegener, and R. Castillo. 2013. "Development and Evaluation of a Categorization Methodology for Occupational Back and Shoulder Injuries Using Claims Data." *Health Services and Outcomes Research Methodology* 13: 140–56.
- Hollingworth, G. R., R. M. Ellis, and T. S. Hattersley. 1983. "Comparison of Injection Techniques for Shoulder Pain: Results of a Double Blind, Randomised Study." *British Medical Journal (Clinical Research Ed)* 287 (6402): 1339–41.
- Hwang, W., W. Weller, H. Ireys, and G. Anderson. 2001. "Out-of-Pocket Medical Spending for Care of Chronic Conditions." *Health Affairs* 20 (6): 267–78.
- Institute of Medicine. 2011. *Clinical Practice Guidelines We Can Trust*. Washington, DC: National Academies Press.
- Institute of Medicine (U.S.). 2011. "Committee on Standards for Developing Trustworthy Clinical Practice Guidelines." In *Clinical Practice Guidelines We Can Trust*, edited by R. Graham, pp. 53–56. Washington, DC: National Academies Press.
- Izquierdo, R., I. Voloshin, S. Edwards, M. Q. Freehill, W. Stanwood, J. M. Wiater, W. C. Watters, M. J. Goldberg, M. Keith, C. M. Turkelson, J. L. Wies, S. Anderson, K. Boyer, L. Raymond, and P. Sluka. 2010. "Treatment of Glenohumeral Osteoarthritis." *Journal of the American Academy of Orthopaedic Surgeons* 18 (6): 375–82.
- Jha, A. K., K. E. Joynt, E. J. Orav, and A. M. Epstein. 2012. "The Long-Term Effect of Premier Pay for Performance on Patient Outcomes." *New England Journal of Medicine* 366 (17): 1606–15.
- Kruse, G. B., D. Polsky, E. A. Stuart, and R. M. Werner. 2012. "The Impact of Hospital Pay-For-Performance on Hospital and Medicare Costs." *Health Services Research* 47 (6): 2118–36.
- Liu, W., S. J. Kuramoto, and E. A. Stuart. 2013. "An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research." *Prevention Sciences* 14 (6): 570–80.
- Lugtenberg, M., J. S. Burgers, and G. P. Westert. 2009. "Effects of Evidence-Based Clinical Practice Guidelines on Quality of Care: A Systematic Review." *Qual Saf Health Care* 18 (5): 385–92.
- Maestas, N., K. J. Mullen, and A. Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–829.
- McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26): 2635–45.
- McGorry, R. W., W. S. Shaw, and J. H. Lin. 2011. "Correlations between Pain and Function in a Longitudinal Low Back Pain Cohort." *Disability and Rehabilitation* 33 (11): 945–52.
- de Moraes, V. Y., K. Godin, M. J. Tamaoki, F. Faloppa, M. Bhandari, and J. C. Belloti. 2012. "Workers' Compensation Status: Does It Affect Orthopaedic Surgery Outcomes? A Meta-Analysis." *PLoS ONE* 7 (12): e50251.
- National Academy of Social Insurance. 2011. *Workers' Compensation: Benefits, Coverage, and Costs, 2009*. p. 115. Washington, DC: National Academy of Social Insurance.

- Oleinick, A., and B. Zaidman. 2004. "Methodologic Issues in the Use of Workers' Compensation Databases for the Study of Work Injuries with Days Away from Work. I. Sensitivity of Case Ascertainment." *American Journal of Industrial Medicine* 45 (3): 260–74.
- Ozaki, J., and I. Kawamura. 1984. "'Zero-Position' Functional Shoulder Orthosis." *Prosthetics and Orthotics International* 8 (3): 139–42.
- Rueda, S., L. Chambers, M. Wilson, C. Mustard, S. B. Rourke, A. Bayoumi, J. Raboud, and J. Lavis. 2012. "Association of Returning to Work with Better Health in Working-Aged Adults: A Systematic Review." *American Journal of Public Health* 102 (3): 541–56.
- Ryan, A. M. 2009. "Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost." *Health Services Research* 44 (3): 821–42.
- Shibata, Y., K. Midorikawa, G. Emoto, and M. Naito. 2001. "Clinical Evaluation of Sodium Hyaluronate for the Treatment of Patients with Rotator Cuff Tear." *Journal of Shoulder and Elbow Surgery* 10 (3): 209–16.
- Steenstra, I., E. Irvin, Q. Mahood, S. Hogg-Johnson, and M. W. Heymans. 2011. *Systematic Review of Prognostic Factors for Workers Time Away from Work Due to Acute Low-Back Pain: An Update of a Systematic Review*, pp. 1–22. Toronto, ON: Institute for Work & Health.
- Stiles, P. G., R. A. Boothroyd, K. Dhont, P. F. Beiler, and A. E. Green. 2009. "Adherence to Practice Guidelines, Clinical Outcomes, and Costs among Medicaid Enrollees with Severe Mental Illnesses." *Evaluation and the Health Professions* 32 (1): 69–89.
- Tanabe, R. P. 2011. *Workers' Compensation Medical Cost Containment: A National Inventory, 2011*. R. P. Tanabe. Cambridge, MA: Workers Compensation Research Institute.
- U.S. Census Bureau. 2012. *NAICS 2012 Definitions* [accessed July 1, 2013]. Available at <http://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2012>.
- U.S. Department of Labor. 2012. *Economic News Release: Workplace Injury and Illness Summary*. Washington, DC: Bureau of Labor Statistics.
- Volinn, E., J. D. Fargo, and P. G. Fine. 2009. "Opioid Therapy for Nonspecific Low Back Pain and the Outcome of Chronic Work Loss." *Pain* 142 (3): 194–201.
- Webb, B. J., B. S. Dangerfield, J. S. Pasha, N. Agrwal, and H. R. Vikram. 2012. "Guideline-Concordant Antibiotic Therapy and Clinical Outcomes in Healthcare-Associated Pneumonia." *Respiratory Medicine* 106 (11): 1606–12.
- Webster, B. S., and M. Cifuentes. 2010. "Relationship of Early Magnetic Resonance Imaging for Work-Related Acute Low Back Pain with Disability and Medical Utilization Outcomes." *Journal of Occupational and Environmental Medicine* 52 (9): 900–7.
- Webster, B. S., S. K. Verma, and R. J. Gatchel. 2007. "Relationship between Early Opioid Prescribing for Acute Occupational Low Back Pain and Disability Duration, Medical Costs, Subsequent Surgery and Late Opioid Use." *Spine* 32 (19): 2127–32.
- Webster, B. S., S. Verma, J. Willetts, K. Hopcia, and R. Wasiak. 2011. "Association of Disability Duration with Physical Therapy Services Provided after Meniscal Sur-

gery in a Workers' Compensation Population." *Archives of Physical Medicine and Rehabilitation* 92 (10): 1542–51.

Work Loss Data Institute. 2011. *Official Disability Guidelines*. Encinitas, CA: Work Loss Data Institute.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Additional Information on the Data, Methods Used to Define Injuries and Clinical Practice Guidelines, and a Sensitivity Analysis on the Proportional Hazards Models.