Evidence-Based Imaging Guidelines and Medicare Payment Policy

Christopher L. Sistrom and Niccie L. McKay

Objective. This study examines the relationship between evidence-based appropriateness criteria for neurologic imaging procedures and Medicare payment determinations. The primary research question is whether Medicare is more likely to pay for imaging procedures as the level of appropriateness increases.

Data Sources. The American College of Radiology Appropriateness Criteria (ACRAC) for neurological imaging, ICD-9-CM codes, CPT codes, and payment determinations by the Medicare Part B carrier for Florida and Connecticut.

Study Design. Cross-sectional study of appropriateness criteria and Medicare Part B payment policy for neurological imaging. In addition to descriptive and bivariate statistics, multivariate logistic regression on payment determination (yes or no) was performed.

Data Collection Methods. The American College of Radiology Appropriateness Criteria (ACRAC) documents specific to neurological imaging, ICD-9-CM codes, and CPT codes were used to create 2,510 medical condition/imaging procedure combinations, with associated appropriateness scores (coded as low/middle/high).

Principal Findings. As the level of appropriateness increased, more medical condition/imaging procedure combinations were payable (low = 61 percent, middle = 70 percent, and high = 74 percent). Logistic regression indicated that the odds of a medical condition/imaging procedure combination with a middle level of appropriateness being payable was 48 percent higher than for an otherwise similar combination with a low appropriateness score (95 percent CI on odds ratio = 1.19–1.84). The odds ratio for being payable between high and low levels of appropriateness was 2.25 (95 percent CI: 1.66–3.04).

Conclusions. Medicare could improve its payment determinations by taking advantage of existing clinical guidelines, appropriateness criteria, and other authoritative resources for evidence-based practice. Such an approach would give providers a financial incentive that is aligned with best-practice medicine. In particular, Medicare should review and update its payment policies to reflect current information on the appropriateness of alternative imaging procedures for the same medical condition.

Key Words. Medicare payment policy, appropriateness, evidence-based medicine, radiology, neurological imaging

Evidence-based medicine is a key component of recent efforts to improve the quality and cost-effectiveness of clinical practice. The past 20 years have seen the development and dissemination of numerous clinical guidelines, critical pathways, clinical protocols, appropriateness criteria, and the like. Although the names vary, the common theme is the use of scientific evidence and expert opinion to specify “best practices” for the evaluation and treatment of a given medical condition, with the ultimate goals of improving the quality of medical care and controlling costs (Garber 2001; Institute of Medicine 2001; Adams and Corrigan 2003). In the United States, health policy is ultimately implemented through the payment determinations of public and private insurance carriers. Thus, financial incentives must be aligned with clinical best practices to encourage the provision of cost-effective medical services. This is especially important for the Medicare program, which has tremendous influence by virtue of its own buying power and as an exemplar for private payers.

This study examines the consonance between evidence-based guidelines and Medicare coverage policies. Specifically, we analyze the relationship between established appropriateness criteria for neurologic imaging procedures and Medicare Part B payment determinations. The primary research question is whether Medicare is more likely to pay for imaging procedures as the level of appropriateness assigned by an expert panel of physicians and neuroscientists increases. Besides having comprehensive appropriateness guidelines already available, imaging is important to study because it accounts for a substantial portion of spending for physician services. Growth of imaging is of particular concern to the Medicare program, with costs for services under the Part B physician fee schedule having risen by nearly 90 percent between 1999 and 2004, from \$5.7 billion to \$10.9 billion (Hackbarth 2006). Moreover, expenditure for imaging is projected to accelerate if current trends in technology development, complexity, and volume continue (Medicare Payment Advisory Commission 2005a, b).

The appropriateness ratings come from evidence-based guidelines for neurologic imaging procedures developed by an expert committee empaneled by the American College of Radiology (ACR) beginning in 1993 (ACR 2005). These ratings, specified for various medical conditions and imaging

Address correspondence to Christopher L. Siström, M.D., M.P.H., Department of Radiology, University of Florida School of Medicine, PO Box 100374, Gainesville, FL 32610-0374. Niccìe L. McKay, Ph.D., is with the Department of Health Services Research, Management and Policy, University of Florida, Gainesville, FL.

procedures, were mapped onto the International Classification of Diseases, ICD-9-CM (2005) and the Common Procedural Terminology, CPT (American Medical Association 2005), code sets. We then added the payment determination for the CPT/ICD-9-CM code pair by the carrier who processes claims for Medicare Part B services in Florida and Connecticut. The resulting analytic data set includes 2,510 observations, each of which represents a medical condition/imaging procedure combination, along with the associated appropriateness rating and Medicare payment determination. By comparing appropriateness ratings produced by a formal physician consensus process with information about Medicare payment determinations, this study provides insight into the extent to which Medicare's policies about whether or not to pay for imaging services are consistent with evidence-based appropriateness criteria.

The results of this study have implications for the current policy interest in pay-for-performance initiatives. Recently, the Medicare Payment Advisory Commission (2005c) recommended that a portion of providers' payments be based on quality of care, with quality to be assessed by evidence-based, standardized, and accepted measures. Our study suggests that, in addition to paying more for higher quality services, Medicare also should review existing payment determinations in light of current evidence-based standards. The appropriateness criteria we analyzed, cover all diagnostic radiology procedures that might conceivably be performed for a given condition or symptom, including those having low appropriate scores. Therefore, the results of this study lend themselves to some initial recommendations regarding Medicare payment determinations for neurologic imaging procedures.

EVIDENCE-BASED IMAGING GUIDELINES

The ACR is the primary professional socioeconomic organization for radiologists in the United States. The ACR Appropriateness Criteria™ Task Force and its expert panels used well-established, validated, and standardized methods to come to consensus about the appropriateness of imaging procedures in the evaluation of more than 190 general medical conditions (ACR 2005). The ACR leadership has formally stated its commitment to continuously updating the content and scope of the guidelines (Cascade 1994, 2000), an important consideration given the rapid rate of technological change in the imaging field.

The ACR Appropriateness Criteria were developed using the RAND methodology of generating appropriateness scores for medical tests and procedures, which combines a thorough literature review and synthesis with

expert panel consensus development using a modified Delphi technique (Brown 1968; Dalkey 1969; Brook, Chassin, and Fink 1986; Phelps 1993; Brook 1994; Black 1998). The general conclusion of the large body of literature assessing the reliability and validity of appropriateness scores derived by the RAND method is that the method is reproducible and a valuable tool for generating clinical appropriateness criteria (Shekelle et al. 1998, 2001; Barnato and Garber 2003; Bindels et al. 2003). The AHRQ National Guideline Clearinghouse (NGC) now includes structured abstracts of the ACR Appropriateness Criteria in its repository.

MEDICARE COVERAGE POLICIES

Medicare Part B covers fee-for-service physician professional services. For the actual administration of payments, Medicare uses “carriers,” which are state and regional contracting organizations. Part B payment relies on claims for specific services and procedures coded by the Healthcare Common Procedure Coding System (HCPCS), with Level I consisting of CPT codes used primarily for services and procedures supplied by physicians and other health care professionals (General Accounting Office 2003). Each claim also contains an explanation by the provider of why the procedure was done, encoded using the ICD-9-CM.

Medicare Part B payment determinations (i.e., whether claims for given services are payable) are made at both the national and carrier levels. The Centers for Medicare and Medicaid Services (CMS) may make a national coverage determination (NCD), although these determinations generally are for newly developed therapies and tests. If an NCD exists, all carriers must follow the policy (General Accounting Office 2003). In the large majority of cases, however, there is no directly applicable national rule or the relevant NCD is quite general in scope. Consequently, carriers have developed local medical review policies (LMRPs), which are applicable only within their own regional jurisdiction (Foote, Halpern, and Wholey 2005). Beginning in 2006, existing LMRPs were converted into local coverage determinations (LCDs), with the major difference being the format and scope of the policy documents.

The current system of Medicare payment determination is highly decentralized, with considerable variation across carriers (Foote et al. 2004; Foote, Halpern, and Wholey 2005; Foote, Wholey, and Halpern 2006). To avoid inequities across beneficiaries and duplication of effort by carriers, the General Accounting Office (2003) recommended that the coverage process be centrally managed. Acknowledging the known variation in Medicare Part B

coverage policies across carriers, we focus on a single one. Specifically, this study uses data on Part B Medicare payment determinations by First Coast Service Options Inc., the carrier for Florida and Connecticut.

METHODS

To analyze the relationship between evidence-based imaging guidelines and Medicare payment determination, we mapped the ACR Appropriateness Criteria (2005) onto the ICD-9-CM and CPT code sets, respectively (International Classification of Diseases 2005; American Medical Association 2005). The imaging appropriateness criteria come from summary documents, each for a general clinical condition such as back pain, head trauma, or epileptic seizures. Each guideline document includes a summary of the relevant literature, along with a series of tables organized by clinical condition and variant. Table 1 is a reproduction of one of the tables in the head trauma guideline (ACR 2005). Row entries in the tables list specific imaging procedures and the associated appropriateness rating from 1 to 9 with higher numbers being more appropriate.

To keep the analysis manageable, we chose to focus on neurologic imaging, which includes 16 guidelines containing a total of 139 individual clinical condition-variant tables. Using the row entries of these tables, we constructed a data set with 2,804 observations, each of which is a medical condition/imaging procedure combination. For each observation, independent variables include: medical condition (ICD-9-CM), imaging procedure (CPT code), appropriateness rating (1–9), modality (e.g., computed tomography), part of the body being imaged (e.g., head), and relative value units (RVU) for performing and interpreting the test.

The dependent variable, Medicare Part B payment determination for a given medical condition/imaging procedure combination, was obtained from a commercial web-based service used by many imaging providers (<http://www.codecorrect.com>). The coding service uses the publicly available official Medicare Part B database of payment policies to determine the coverage category, based on the carrier, procedure, and medical condition. Coverage categories include: payable, no local medical review policy (payable), not payable, and not covered (not payable). In this study, the Medicare payment variable indicates the final determination of payable or not payable. The Supplementary Material Appendix A contains additional detail regarding construction of the data set.

Table 1: Example of a Single Variant Table Found in the ACR Appropriateness Criteria™ (American College of Radiology 2005)

<i>Clinical Condition: Closed Head Injury</i>		
<i>Variant 7: Subacute Head Injury, Late Neurologic Deterioration</i>		
<i>Radiologic Exam Procedure</i>	<i>Appropriateness Rating</i>	<i>Comments</i>
CT	8	
MR	8	
MR with gadolinium	2	
MR with MRA	2	
CT with contrast	2	
Skull plain film radiographs	2	
C-spine plain film radiographs	2	
Cerebral angiography	2	
Xenon-enhanced CT	2	
Transcranial Doppler	2	
SPECT	0	Functional imaging such as SPECT and PET may reveal abnormalities related to head injury that are not apparent on anatomic imaging studies. Further investigation is needed
	No consensus	
PET	0	Functional imaging such as SPECT and PET may reveal abnormalities related to head injury that are not apparent on anatomic imaging studies. Further investigation is needed
	No consensus	
Appropriateness Criteria Scale		
1 2 3 4 5 6 7 8 9		
1 = Least appropriate		9 = Most appropriate

Before analyzing the data, we deleted 294 observations. The majority of the deleted observations ($N=231$) were cases in which the appropriateness rating was 0, indicating that the expert panel could come to no consensus. Other observations deleted were for procedures rarely used for neurologic imaging: thermography, to detect changes in skin surface temperature ($N=12$), and magnetoencephalography, which measures magnetic fields generated by neuronal activity of the brain ($N=6$). Finally, 45 observations were deleted because the listed imaging technique was so new that it had not yet been assigned a CPT code (e.g., functional brain magnetic resonance imaging [MRI]).

The analytic data set has 2,510 observations for 139 different medical conditions and 65 imaging procedures. Modalities include ANG (catheter

angiography), CT (computed tomography), MRI, NUC (nuclear imaging), RAD (standard X-rays), SPEC (special invasive procedures, such as myelography), and US (diagnostic ultrasound, including vascular diagnostic procedures). Body part categories are Head, Neck, and Spine.

In interpreting the results of this study, it is important to keep in mind that the unit of analysis is a medical condition/imaging procedure combination from the ACR guidelines for neurologic imaging. Consequently, this data set contains no information about the extent to which any given medical condition/imaging procedure combination actually is submitted to Medicare for reimbursement or utilized in practice.

Descriptive Statistics

Table 2 presents descriptive statistics for variables used in the analysis. For Medicare payment determination, 66.2 percent of the medical condition/imaging procedure combinations were identified as payable, with the remaining 33.8 percent being not payable.

Although the appropriateness ratings in the ACR tables range from 1 (least appropriate) to 9 (most appropriate), preliminary analysis using the full range of scores showed a distribution of predominantly even numbered scores, suggesting that the actual number of meaningful categories is <9. Consequently, we chose to collapse the appropriateness ratings into three categories: 1–3 (low), 4–6 (middle), and 7–9 (high). As shown in Table 2, 47.0 percent of the medical condition/procedure combinations had a low appropriateness rating, with 35.7 and 17.3 percent having middle and high appropriateness ratings, respectively.

Table 2 also shows that, for modality, CT and MRI account for the majority of medical condition/imaging procedure combinations: 29.9 and 31.3 percent, respectively. Head procedures account for approximately three-quarters of the medical condition/imaging procedure combinations, reflecting the fact that the data set is restricted to neurologic imaging.

Although the RVU values were entered in numerical form, a frequency distribution indicated that the RVU values were strongly clustered toward certain values. For example, 67 observations had $RVU = 0.47$, but no observations had RVU values equal to 0.43, 0.44, 0.45, 0.46, 0.48, or 0.49; this pattern probably reflects the fact that RVU values are administratively assigned. Consequently, we created a categorical variable that assigned observations to one of five RVU ranges. The frequency distribution for RVU category, also included in Table 2, shows that RVU category 3, which includes

Table 2: Descriptive Statistics

<i>Variable</i>	<i>Frequency (n = 2,510)</i>	
	<i>Number</i>	<i>Percent</i>
Medicare payment determination (<i>P</i>)		
1 = payable	1,661	66.2
0 = not payable	849	33.8
Appropriateness category (APPR)		
Low	1,180	47.0
Middle	896	35.7
High	434	17.3
Modality (MOD)		
CT	751	29.9
ANG	164	6.5
MRI	785	31.3
NUC	420	16.7
RAD	210	8.4
SPEC	58	2.3
US	122	4.9
Body part (BOD)		
Head	1,848	73.6
Neck	95	3.8
Spine	567	22.6
RVU category (RVUCAT)		
RVU < 0.5	178	7.1
0.499 < RVU < 1.5	469	18.7
1.499 < RVU < 2.5	1,672	66.6
2.499 < RVU < 3.5	111	4.4
RVU > 3.499	80	3.2

RVU, relative value units.

the intermediate values (1.5–2.5), is most common—approximately 67 percent of the imaging procedures fall into this RVU category.

Statistical Analysis

The primary research question of this study is whether Medicare is more likely to pay for imaging procedures as the level of appropriateness increases. To investigate this question, we used both bivariate and multivariate analyses.

First, we cross-tabulated the data according to appropriateness category (low, middle, and high) and Medicare payment determination (payable, not payable). A Cochran–Armitage test for trend was used to determine whether the percentage of payable observations exhibited a significant trend across the three appropriateness categories.

Then we used logistic regression analysis to examine the effect of appropriateness category on the likelihood of Medicare payment, holding constant other factors expected to affect Medicare payment determination. The dependent variable is the Medicare payment determination (P): 1 = payable, 0 = not payable. The independent variable of primary interest is the appropriateness category (APPR).

Control variables included in the regression analysis to take account of other factors that may affect payment coverage are modality (MOD), body part (BOD), and RVU category (RVUCAT). For a given medical condition, payment may be different for standard X-rays or angiography, which have been in use longer, than for CT or MRI. Imaging modality may also control for differences in lobbying intensity among physician subspecialty groups and equipment vendors. The degree of consensus regarding diagnosis and treatment also may affect the probability of being payable. For example, the choice of imaging procedure—and if one should be done at all—for low back pain is less well-defined than for severe head trauma where CT scanning is universally considered to be necessary. Body part is included as a control variable to approximate for this effect. The final control variable included in the regression analysis is RVU category, because RVU is directly related to the dollar amount of payment.

The general form of the regression equation is P
 $= f[\text{APPR}; \text{MOD}, \text{BOD}, \text{RVUCAT}]$

where P is Medicare payment determination, APPR is appropriateness category, MOD is modality, BOD is body part, and RVUCAT is RVU category.

We specify the equation using a logistic linking function, because the dependent variable is dichotomous. For the dependent variable of Medicare payment determination ($P=1$: payable, or $P=0$: not payable), the linear logistic model takes the form

$$\begin{aligned} \log[p/(1-p)] = & \alpha + \beta_1 \text{ APPR : Middle} + \beta_2 \text{ APPR : High} \\ & + \beta_3 \text{ MOD : ANG} + \beta_4 \text{ MOD : MRI} \\ & + \beta_5 \text{ MOD : NUC} + \beta_6 \text{ MOD : RAD} \\ & + \beta_7 \text{ MOD : SPEC} + \beta_8 \text{ MOD : US} \\ & + \beta_9 \text{ BOD : Neck} + \beta_{10} \text{ BOD : Spine} \\ & + \beta_{11} (0.499 < \text{RVU} < 1.5) \\ & + \beta_{12} (1.499 < \text{RVU} < 2.5) \\ & + \beta_{13} (2.499 < \text{RVU} < 3.5) \\ & + \beta_{14} (\text{RVU} > 3.499) + \varepsilon \end{aligned}$$

where p is the probability of being payable, α is the intercept parameter, β is the vector of slope parameters, APPR is appropriateness category (APPR : Low being the reference category), MOD is modality (MOD : CT being the reference category), BOD is body part (BOD : Head being the reference category), and RVUCAT is RVU category (RVU < 0.5 being the reference category), and ε is the error term. We used standard maximum likelihood methods (SAS PROC LOGISTIC) to estimate the parameters. For ease of interpretation, the results are reported in terms of odds ratio estimates, along with confidence intervals.

RESULTS

Table 3 presents the appropriateness category (high, middle, and low) cross tabulated against Medicare payment determination (payable, not payable). A Cochran–Armitage trend test indicated that the percentage of payable observations exhibited a significant trend across the three appropriateness categories ($p < .0001$). It is notable that even medical condition/imaging procedure combinations assigned a low appropriateness rating were payable the majority of the time (60.5 percent). Nevertheless, the payable proportion did increase significantly as the appropriateness category increases, being 70.1 percent for combinations with a middle appropriateness rating and 73.5 percent for those in the high category. The bivariate analysis thus indicates that being in a higher appropriateness category was significantly associated with a higher probability of being payable.

Results from the logistic regression are shown in Table 4. The appropriateness category, modality, and RVU category had statistically significant

Table 3: Medicare Part B Payment Determination by Appropriateness

	<i>Receive Payment</i>		<i>Total</i>
	<i>No</i>	<i>Yes</i>	
Appropriateness			
Low	466 (39.5%)	714 (60.5%)	1,180
Middle	268 (29.9%)	628 (70.1%)	896
High	115 (26.5%)	319 (73.5%)	434
Total	849	1,661	2,510

Row percentages are in parentheses.

Cochran–Armitage Trend Test ($Z = -5.59$, $p < .0001$).

Table 4: Results of Logistic Regression on Medicare Part B Payment Determination (Payable = 1, Not Payable = 0)

<i>Variable</i>	<i>Odds Ratio</i>	<i>95% Wald Confidence Interval</i>	<i>p-value</i>
Appropriateness category (APPR) (reference = low)			<.0001
Middle	1.48	1.19–1.84	
High	2.25	1.66–3.04	
Modality (MOD) (reference = CT)			<.0001
ANG	19.09	6.94–52.48	
MRI	0.46	0.36–0.60	
NUC	0.42	0.32–0.54	
RAD	7.37	0.98–55.57	
SPEC	10.54	2.48–44.70	
US	0.14	0.07–0.29	
Body part (BOD) (reference = head)			.756
Neck	0.98	0.45–2.14	
Spine	1.11	0.85–1.45	
RVU category (RVUCAT) (reference = RVU < 0.5)			.024
0.499 < RVU < 1.5	0.38	0.04–3.85	
1.499 < RVU < 2.5	0.28	0.03–2.85	
2.499 < RVU < 3.5	0.47	0.04–5.01	
RVU > 3.499	0.48	0.04–5.37	

RVU, relative value units.

effects (at the .05 confidence level or better) on the probability of being payable, while body part did not have a statistically significant effect. The odds ratios for appropriateness category indicate that the probability of being payable increased significantly with the level of appropriateness. Specifically, the odds of a medical condition/imaging procedure combination with a middle level of appropriateness being payable was 48 percent higher than for an otherwise similar combination with a low level of appropriateness (95 percent Wald confidence limits: 1.19–1.84). For high compared with low level of appropriateness, the odds ratio was 2.25 (95 percent Wald confidence limits: 1.66–3.04).

The logistic regression also indicates that modality had a significant effect on the probability of being payable. For example, the odds of being payable relative to CT were higher in some cases (ANG, RAD, and SPEC) and lower in other cases (MRI, NUC, and US). Because the regression holds other factors, including the appropriateness category, constant, the modality odds

ratios thus indicate that the odds of being payable are higher for some modalities than for others, even for a given level of appropriateness.

Finally, even though RVU category as a whole had a statistically significant effect on the likelihood of being payable ($p = .02$), none of the odds ratios for individual RVU categories were significantly different from one in comparison with the reference category. Consequently, it appears that RVU category plays a minor role at best in influencing the likelihood of being payable, after controlling for the level of appropriateness and modality.

As shown in Table 3, even medical condition/imaging procedure combinations assigned a low appropriateness rating were payable 60.5 percent of the time. One possible explanation for this finding is that the same ICD-9-CM/CPT code pair may apply to multiple scenarios, some with low appropriateness ratings and some with middle and high ratings. This would result in an overestimate of the rate of payment in the low appropriateness category. In order to further examine this possibility, we first identified observations that shared the same ICD-9-CM/CPT code pair, then set the appropriateness rating for each one to the maximum so as obtain a lower bound for the proportion in the low category being payable. Of the 1,180 observations originally in the low appropriateness category, 192 were reassigned to the middle category and 97 moved up to the high category leaving 735 of which 385 (52 percent) were still payable. Of the 896 original middle category observations, 84 would move into the high category. After this adjustment, there were 1,051 middle appropriateness observations of which 722 (69 percent) were payable and 724 in the high category with 542 (75 percent) payable.

SUMMARY AND DISCUSSION

This study examined the extent to which Medicare payment determination is consistent with evidence-based appropriateness guidelines for neurologic imaging procedures. The analytic data set has 2,510 observations, each of which represents a specific medical condition/imaging procedure combination. The results indicate that payment was more likely to be received as the level of appropriateness increases. The odds of a medical condition/imaging procedure combination with a middle level of appropriateness being payable was 48 percent higher than for an otherwise similar combination with a low appropriateness rating, with the estimated odds ratio being 2.25 for high compared with low level of appropriateness. Although a higher level of

appropriateness was associated with a higher probability of being paid, the majority of procedures (at least 52 percent) in the low appropriateness category were also payable. This was true even after correcting for lack of specificity in the ICD-9-CM coding of medical conditions which resulted in 373 (15 percent) of the original 2,510 observations having identical ICD-9-CM/CPT codes yet different appropriateness categories.

Another interesting finding was that the type of procedure (modality) had a significant effect on the probability of being payable. For example, the odds of being payable for CT were higher than for some (ANG, RAD, and SPEC) and lower than other modalities (MRI, NUC, and US), even after controlling for appropriateness and costliness (RVU) of the test. This may reflect the relative intensity and success of lobbying efforts by imaging equipment vendors and physician groups on decisions made by Medicare Carrier Advisory Committees.

Limitations

Perhaps the most important limitation of this study is that the data set is composed of the medical condition/imaging procedure combinations contained in the ACR Appropriateness Criteria. Consequently, this study cannot determine how often a given medical condition/imaging procedure combination is actually claimed or paid for. For example, of the 1,661 medical condition/imaging procedure combinations being payable, 714 (43 percent) are in the low appropriateness category. However, it is possible that these combinations are rarely performed in practice or submitted for reimbursement. We currently are working on a companion study that will examine the extent to which neurologic imaging procedures that are actually submitted to and paid by Medicare vary by appropriateness category.

Using ICD-9-CM codes to represent clinical events in hospital abstracts and claims data have been called into question in many venues (Walsh 1992; Chute et al. 1996; Tuttle and Nelson 1996; Phillips and Hillman 2001). First, a given ICD-9-CM code may refer to multiple medical conditions. Second, using ICD-9-CM codes may be especially problematic when evaluating indications for imaging because they are often expressed in terms of symptoms and signs. While the ICD-9-CM system is extremely robust and complete in covering medical diagnoses, clinical signs and symptoms are relegated to a relatively short section. For example, although headache has only one code (784.0) under *Symptoms, Signs, and Ill Defined Conditions* in ICD-9-CM, there are more than 10 codes for specific types of headache (e.g., migraine = 346.9,

allergic = 346.2, and menopausal = 627.2). However, these diagnoses are often based on results of imaging tests (e.g., negative CT or MRI of the brain). Thus much of the noise in our data may be due to problems with assigning ICD-9-CM codes to the clinical scenarios described by the ACR expert panels for neurological imaging. Nonetheless, in the daily practice of ordering, billing, and adjudicating claims for imaging services, the ICD-9-CM and CPT systems are the standard for Medicare in the United States (Anonymous 2000).

Recently, the ACR Appropriateness Criteria have been criticized for relying on expert consensus, as well as regarding their methodology, content, format, and availability (Blackmore and Medina 2006). While there is clearly room for improvement in the criteria (Sistrom and Honeyman 2002; Sistrom 2005), many disagree with Blackmore's argument that the Delphi consensus method should be abandoned because expert-derived appropriateness scores are irredeemably invalid and unreliable. In order to develop evidence-based guidelines, at some point an individual or a committee must make substantive decisions about what clinical conditions constitute appropriate reasons for doing imaging tests. When a group is involved, the RAND methodology is a well-accepted approach for developing consensus about appropriateness scores. In response to Blackmore's article, the current chairman of the ACR Appropriateness Task Force responded positively to many of the critiques, but demurred from eliminating consensus-based appropriateness tables in subsequent versions (Bettmann 2006).

Another limitation is that the Medicare payment determinations in this study are for a single carrier, First Coast Service Options Inc., for two states, Florida and Connecticut. Other studies have found considerable variation in Medicare payment determinations across carriers (Foote et al. 2004; Foote, Halpern, and Wholey 2005). However, at a minimum the results are relevant for Florida and Connecticut, which together accounted for 8.6 percent of U.S. Medicare beneficiaries in 2005 (CMS 2007). Finally, this study looks only at appropriateness criteria for neurologic imaging procedures, which come from only one of the expert panels convened by the ACR. Because there are 17 separate expert panels, the way in which the appropriateness scale is applied is likely to differ by type of imaging procedure.

Policy Implications

On one level, the results of this study should be reassuring to policy makers in that Medicare was found to be more likely to pay for procedures that were rated as more appropriate. Thus, Medicare payment determinations do

appear to be generally consistent with evidence-based appropriateness guidelines, at least for neurologic-imaging procedures claims processed by the Medicare carrier for Florida and Connecticut. On the other hand, the results provide some cause for concern in that the majority of procedures in the low appropriateness category were payable, perhaps due to outdated procedures that have not been deleted from Medicare's coverage policies. However, even if those procedures are rarely performed, it is hard to understand why Medicare should pay at all for procedures deemed to be least appropriate among alternative imaging procedures for the same medical condition.

Our results suggest that Medicare could improve its payment determinations for imaging by taking advantage of well-established and generally accepted evidence-based guidelines. Conceptually, Medicare coverage policy should use payment to favor use of cost-effective imaging and avoid payment for inappropriate testing (Garber 2001). Historically, little evidence has been available to distinguish between more and less appropriate levels of utilization. However, as well-accepted clinical guidelines, appropriateness criteria, and other forms of evidence-based medicine proliferate, Medicare (and other payers) should take advantage of this scientific information to review existing payment policies as well as in making payment decisions about new procedures and technologies. Such an approach would give providers a financial incentive that is aligned with best-practice and cost-effective medicine.

The 2003 Medicare Prescription Drug, Improvement and Modernization Act mandates that CMS consolidate the roughly 40 carriers into 15 Medicare Administrative Contractors (MAC) by 2011 (U.S. Congress 2003). Therefore, in the medium term, comprehensive reform of the payment policies for imaging procedures should not be attempted at the local carrier level since these entities will cease to exist. However, existing NCD might reasonably be altered to reflect evidence-based appropriateness criteria and other guidelines as they would remain in force after carrier consolidation into the MAC structure. Presumably the MAC entities themselves will adapt and/or create their own ICD-9-CM/CPT payment tables and this process should be informed by the same evidence-based criteria and guidelines.

Our findings suggest two distinct strategies for updating payment policies to align with evidence-based appropriateness criteria. The first is to eliminate any specific denials (ICD-9-CM/CPT pairs) deemed to be appropriate by the relevant guidelines. This would serve to reduce the fraction of "not payable" determinations rendered for appropriate imaging procedures. The clinical logic for eliminating denials of appropriate procedures is straightforward: patients should not be denied care that is known to be

appropriate. The second strategy would concentrate on procedures deemed to have a low appropriate score for particular clinical indications, with ICD-9-CM/CPT pairs that the criteria indicate having a low appropriateness rating being classified as not payable. Because the appropriateness criteria are updated periodically, which is important given the rapid rate of technological change in imaging, payment policies referring to the criteria would have to be reviewed and updated on a regular basis. Before considering either strategy, however, Medicare Part B data should be analyzed to determine which ICD-9-CM/CPT code pairs occur frequently in claims for imaging and to compare these with existing appropriateness criteria. This would allow policy makers to assess the economic impact of proposed changes to payment policies. Comparing claims data with appropriateness criteria for imaging might also reveal substantial “gaps” in the evidence base if the existing guidelines do not cover medical condition/imaging procedure combinations that are frequently performed.

The previously discussed limitations of the ICD-9-CM coding system and critiques of the process and content of the ACR Appropriateness Criteria may limit the perceived validity of “back end” procedures (adjudicating submitted claims) that use guidelines for reducing cost and improving quality. Other proposed strategies include limiting imaging providers to those with special training and radiology benefits management using preauthorization mechanisms, although these schemes are controversial and opposed by many physicians. An alternative approach is to intervene at the “front end” when imaging procedures are being ordered, by incorporating appropriateness criteria in computerized physician order entry (CPOE) systems. This method has shown considerable promise and has been used in the Harvard–Partners system for more than 5 years (Harpole et al. 1997; Rothschild et al. 2000; Khorasani 2001); the ACR Appropriateness Criteria form the core content of the Harvard–Partners decision support logic for imaging, although some additions and modifications were required.

ACKNOWLEDGMENTS

The authors thank Jeffrey Harman and two anonymous reviewers for helpful comments. Dr. Sistrom received financial support for this research from the Department of Radiology, University of Florida College of Medicine and from the “Advanced Program in Clinical Investigation” at the University of Florida (NIH #K30 HL004109).

REFERENCES

- Adams, K., and J. M. Corrigan, Eds. 2003. *Priority Areas for National Action: Transforming Health Care Quality*. Washington, DC: Institute of Medicine.
- American College of Radiology (ACR). 2005. "American College of Radiology Appropriateness Criteria™" [accessed on August 11, 2005]. Available at <http://www.acr.org/ac>
- American Medical Association. 2005. "Common Procedure Terminology Codes" [accessed on August 11, 2005]. Available at http://catalog.ama-assn.org/Catalog/cpt/cpt_home.jsp
- Anonymous. 2000. "CPT and ICD-9 Code Systems Approved as the National Standard." *ACR Bulletin* 56: 4.
- Barnato, A. E., and A. M. Garber. 2003. "Performance of the RAND Appropriateness Criteria." *Medical Decision Making* 23: 177–9.
- Bettmann, M. A. 2006. "The ACR Appropriateness Criteria: View from the Committee Chair." *Journal of the American College of Radiology* 3: 510–2.
- Bindels, R., A. Hasman, J. W. van Wersch, P. Pop, and R. A. Winkens. 2003. "The Reliability of Assessing the Appropriateness of Requested Diagnostic Tests." *Medical Decision Making* 23: 31–7.
- Black, N. 1998. "Assessing the Appropriateness of Medical Care." *New England Journal of Medicine* 39: 1479–80.
- Blackmore, C. C., and L. S. Medina. 2006. "Evidence-Based Radiology and the ACR Appropriateness Criteria." *Journal of the American College of Radiology* 3: 505–9.
- Brook, R. H. 1994. "The RAND/UCLA Appropriateness Method." In *Methodology Perspectives. AHCPR Pub. NO. 95-0009*, edited by K. A. McCormick et al., pp. 59–70. Rockville, MD: Public Health Service, U.S. Department of Health and Human Services.
- Brook, R. H., M. R. Chassin, A. Fink, D. H. Solomon, J. Kosecoff, and R. E. Park. 1986. "A Method for the Detailed Assessment of the Appropriateness of Medical Technologies." *International Journal of Technology Assessment in Health Care* 2: 53–63.
- Brown, B. B. 1968. *The Delphi Method: A Methodology Used for the Elicitation of Opinion of Experts*. Santa Monica, CA: The RAND Corporation.
- Cascade, P. N. 1994. "Setting Appropriateness Guidelines for Radiology." *Radiology* 192: 50A–4A.
- . 2000. "The American College of Radiology ACR Appropriateness Criteria Project." *Radiology* 214 (suppl): 3–46.
- Centers for Medicare and Medicaid Services (CMS). 2007. "Medicare Enrollment—All Beneficiaries: As of July 2005" [accessed on January 13, 2007]. Available at <http://www.cms.hhs.gov/MedicareEnRpts>
- Chute, C. G., S. P. Cohn, K. E. Campbell, D. E. Oliver, and J. R. Campbell. 1996. "The Content Coverage of Clinical Classifications. For the Computer-Based Patient Record Institute's Work Group on Codes & Structures." *Journal of the American Medical Informatics Association* 3: 224–33.

- Dalkey, N. C. 1969. *The Delphi Method: An Experimental Study of Group Opinion*. Santa Monica, CA: The RAND Corporation.
- Foote, S. B., R. Halpern, and D. R. Wholey. 2005. "Variation in Medicare's Local Coverage Policies: Content Analysis of Local Medical Review Policies." *American Journal of Managed Care* 11: 181-7.
- Foote, S. B., D. R. Wholey, and R. Halpern. 2006. "Rules for Medical Markets: The Impact of Medicare Contractors on Coverage Policies." *Health Services Research* 41 (3): 721-42.
- Foote, S. B., D. R. Wholey, T. Rockwood, and R. Halpern. 2004. "Resolving the Tug-of-War between Medicare's National and Local Coverage." *Health Affairs* 23: 108-23.
- Garber, A. M. 2001. "Evidence-Based Coverage Policy." *Health Affairs* 20: 62-82.
- General Accounting Office. 2003. *Medicare: Divided Authority for Policies on Coverage of Procedures and Devices Results in Inequities*. Pub. No. GAO-03-175. Washington, DC: General Accounting Office.
- Hackbarth, G. M. 2006. "MedPAC Recommendations on Imaging Services: Testimony before the Subcommittee on Health, Committee on Energy and Commerce." U.S. House of Representatives, July 18.
- Harpole, L. H., R. Khorasani, J. Fiskio, G. J. Kuperman, and D. W. Bates. 1997. "Automated Evidence-Based Critiquing of Orders for Abdominal Radiographs: Impact on Utilization and Appropriateness." *Journal of the American Medical Informatics Association* 4: 511-21.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- International Classification of Diseases, Ninth Revision, Clinical Modification. 2005. "ICD-9-CM Codes" [accessed on August 11, 2005]. Available at <http://www.cms.hhs.gov/ICD9ProviderDiagnosticCodes/>
- Khorasani, R. 2001. "Computerized Physician Order Entry and Decision Support: Improving the Quality of Care." *Radiographics* 21: 1015-18.
- Medicare Payment Advisory Commission. 2005a. *A Data Book: Healthcare Spending and the Medicare Program*. Washington, DC: Medicare Payment Advisory Commission.
- . 2005b. *Recommendations on Imaging Services*. Washington, DC: Medicare Payment Advisory Commission.
- . 2005c. *Medicare Payment Policy*. Washington, DC: Medicare Payment Advisory Commission.
- Phelps, C. E. 1993. "The Methodologic Foundations of Studies of the Appropriateness of Medical Care." *New England Journal of Medicine* 329: 1241-5.
- Phillips, C. D., and B. J. Hillman. 2001. "Coding and Reimbursement Issues for the Radiologist." *Radiology* 220: 7-11.
- Rothschild, J. M., R. Khorasani, and D. W. Bates. 2000. "Guidelines and Decision Support Help Improve Image Utilization. [In Process Citation]." *Diagnostic Imaging* 22: 95-7, 99, 101.
- Shekelle, P.G., J. P. Kahan, S. J. Bernstein, L. L. Leape, C. J. Kamberg, and R. E. Park. 1998. "The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures." *New England Journal of Medicine* 338 (26): 1888-95.

- Shekelle, P. G., R. E. Park, J. P. Kahan, L. L. Leape, C. J. Camberg, and S. J. Bernstein. 2001. "Sensitivity and Specificity of the RAND/UCLA Appropriateness Method to Identify the Overuse and Underuse of Coronary Revascularization and Hysterectomy." *Journal of Clinical Epidemiology* 54: 1004–10.
- Sistrom, C. L. 2005. "The American College of Radiology Appropriateness Criteria: Translation to Practice and Research." *Journal of the American College of Radiology* 2: 61–7.
- Sistrom, C. L., and J. C. Honeyman. 2002. "Relational Data Model for the American College of Radiology Appropriateness Criteria." *Journal of Digital Imaging* 15: 216–25.
- Tuttle, M. S., and S. J. Nelson. 1996. "Editorial: A Poor Precedent." *Methods of Information in Medicine* 35: 211–7.
- U.S. Congress. 2003. Medicare Prescription Drug, Improvement and Modernization Act of 2003 § 911(a)(1).
- Walsh, E. 1992. "The Coder's Dilemma. Expediency vs. Accuracy." *ADM Radiology* 11: 40–1.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:
Appendix A. Construction of Data Set.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1475-6773.2007.00802.x>
(this link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.