© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12597
RESEARCH ARTICLE

Low-Value Service Use in Provider Organizations

Aaron L. Schwartz, Alan M. Zaslavsky, Bruce E. Landon, Michael E. Chernew, and J. Michael McWilliams

Objective. To assess whether provider organizations exhibit distinct profiles of low-value service provision.

Data Sources. 2007–2011 Medicare fee-for-service claims and enrollment data.

Study Design. Use of 31 services that provide minimal clinical benefit was measured for 4,039,733 beneficiaries served by 3,137 provider organizations. Variation across organizations, persistence within organizations over time, and correlations in use of different types of low-value services within organizations were estimated via multilevel modeling, with adjustment for beneficiary sociodemographic and clinical characteristics.

Principal Findings. Organizations provided 45.6 low-value services per 100 beneficiaries on average, with considerable variation across organizations (90th/10th percentile ratio, 1.78; 95 percent CI, 1.72–1.84), including substantial between-organization variation within hospital referral regions (90th/10th percentile ratio, 1.66; 95 percent CI, 1.60–1.71). Low-value service use within organizations was highly correlated over time (r , 0.98; 95 percent CI, 0.97–0.99) and positively correlated between 13 of 15 pairs of service categories (average r , 0.26; 95 percent CI, 0.24–0.28), with the greatest correlation between low-value imaging and low-value cardiovascular testing and procedures (r , 0.54).

Conclusions. Use of low-value services in provider organizations exhibited substantial variation, high persistence, and modest consistency across service types. These findings are consistent with organizations shaping the practice patterns of affiliated physicians.

Key Words. Low value, efficiency, quality measurement, Medicare

At the leading edge of payment reform are new payment models that rely on financial incentives at the level of provider organizations to improve health care value. For example, both private and public insurers have launched accountable care organization (ACO) initiatives, which place total spending for a provider organization's patients under a global budget with concomitant incentives to improve quality of care (McWilliams et al. 2014b, 2015, 2016; Song et al. 2014). More broadly, the Department of Health and Human

Services recently set a goal of tying 50 percent of payments in traditional Medicare to alternative value-based payment models by the end of 2018 (Burwell 2015).

The long-term impact of these models will depend on organizations' ability to shape the behavior of affiliated physicians (Rebitzer and Votruba 2014). Although there are many recent examples of studies demonstrating heterogeneity in physician behavior (Abaluck et al. 2014; Currie, MacLeod, and Van Parys 2015), there has been less focus on the role of provider organizations. From small group practices to large integrated delivery systems, provider organizations may influence the practice patterns of affiliated physicians in several ways (Landon 1998): by setting the form of physician compensation (Robinson 2001), making purchasing decisions with regard to biomedical and information technology (Javitt, Rebitzer, and Reisman 2008; Bright et al. 2012), adopting care delivery models like the patient-centered medical home (Jackson et al. 2013), selectively attracting physicians with specific training or interests (Asch et al. 2009; Asch, Epstein, and Nicholson 2007; Doyle, Ewer, and Wagner 2010), or standardizing clinical practices through guidelines, performance assessment, and retraining. Even in the absence of such active management techniques, physicians may adopt new practices as a consequence of learning from colleagues within an organization (Landon et al. 2012).

We used 31 claims-based measures of low-value service use to determine whether provider organizations exhibit distinct patterns of low-value service use. Specifically, we measured three properties of low-value service use in the fee-for-service Medicare population: variation across organizations, persistence of service use within an organization over time, and consistency of organizational behavior across different types of low-value services. The presence of organizational variation, persistence, and consistency would suggest organizational influences already at work that would support the rationale for payment models with organization-level financial incentives.

Address correspondence to J. Michael McWilliams, M.D., Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115; e-mail: mcwilliams@hcp.med.harvard.edu. Aaron L. Schwartz, Ph.D., Alan M. Zaslavsky, Ph.D., Bruce E. Landon, M.D., M.B.A., M.Sc., and Michael E. Chernew, Ph.D., are with the Department of Health Care Policy, Harvard Medical School, Boston, MA. Bruce E. Landon, M.D., M.B.A., M.Sc., is also with the Division of General Internal Medicine and Primary Care, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA. J. Michael McWilliams, M.D., Ph.D., is also with the Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

METHODS

An ideal assessment of organizational influences on the provision of low-value services would entail randomizing patients to provider organizations and observing the resulting patterns of low-value service use (variation, persistence, and consistency). In the absence of random assignment, it is difficult to isolate organizational effects from patient demand, especially in the setting of ambulatory patient encounters. Instrumental variables approaches used to study variation in inpatient providers (Doyle et al. 2015) are challenging to apply in the ambulatory setting, where there is greater scope for patient choice of provider; prior work has demonstrated limitations of using distance to outpatient providers to predict provider choice for nonemergent care (Keating et al. 2016). Our approach is therefore to measure patterns of low-value service use within and across provider organizations, adjusted for case mix, and to perform additional statistical tests of alternate explanations for practice patterns besides organizational influence on provider behavior.

Study Population of Patients and Organizations

We analyzed 2007–2011 claims and enrollment information for a random 20 percent sample of fee-for-service Medicare beneficiaries. In each year of the study period, we excluded beneficiaries who were not continuously enrolled in Medicare Parts A and B (while alive) during that year and the prior year. Beneficiaries were also excluded from the study for any years in which they did not receive evaluation and management (E&M) services that we used to attribute patients to provider organizations (Appendix).

Each provider organization was defined by a single taxpayer identification number (TIN). TINs, which are present in Medicare claims for professional services, identify a range of organizational units, from individual physicians, to small group practice sites, to large multisite provider groups, or even integrated delivery systems (Welch et al. 2013). Following the patient attribution rules for the Medicare Shared Savings Program and previously described methods (McWilliams et al. 2015), we attributed each beneficiary in each year to the TIN that accounted for the most allowed charges for a set of outpatient E&M services (CPT codes 99201–99215, 99241–99245, G0402, G0438, and G0439) during the year. We restricted our sample to beneficiaries assigned to medium-sized or large practices/organizations, defined as TINs with 1,000 or more attributed patients over the 5-year study period. For our

analysis of persistence of low-value service use, which we assessed from 2010 to 2011, we further restricted the sample to beneficiaries attributed to organizations with more than 200 attributed patients in 2010 and in 2011. These attribution rules associated each organization with an exclusive set of patients whose care it potentially influenced, allowing us to quantify use of low-value services per patient-year of care provided.

Because Medicare beneficiaries often receive care from multiple practices and thus may receive much of their care from providers outside of an attributed TIN (Pham et al. 2007; McWilliams et al. 2014a), correlations between use of different types of low-value services could be low within groups of patients attributed to TINs even if physicians in each TIN exhibited similar practice patterns. This is because the low-value treatment decisions of providers within an organization are unlikely to be highly correlated with those of providers outside of the organization, particularly if different organizations influence the practice patterns of their constituent providers differently. Therefore, for our analysis of the consistency of organizational behavior across different types of low-value services, we also examined a set of large provider organizations—the 250 groups of providers that entered the Medicare ACO programs in 2012 or 2013. Because ACOs are defined as collections of TINs and must serve at least 5,000 Medicare beneficiaries per year (Code of Federal Register. Title 42. Part 425.110), they are larger and encompass more specialties, on average, than the TINs we analyzed. Consequently, we expected an ACO to provide a greater proportion of services for its attributed beneficiaries, on average, than a single TIN. We focused on ACOs as a convenient set of large organizations because, unlike other organizations with multiple constituent TINs appearing in claims, we could observe the TINs composing each ACO. ACOs therefore provided an opportunity to assess organizational consistency with less attenuation due to care fragmentation. Because we were interested in ACOs as a conveniently well-defined set of large organizations and not in the short-term effects of ACO contracts, we examined data before ACO contracts started in 2012. Although most ACOs were preformed organizations prior to the start of contracts, any lack of clinical integration prior to the contracts would only bias correlations in the provision of different types of low-value services toward zero. For the ACO analysis, beneficiaries were attributed to an ACO if they accumulated more allowed charges for the outpatient E&M services of interest at that ACO than at any other ACO or non-ACO TIN. See Appendix for details regarding beneficiary attribution to provider organizations.

Study Variables

Low-Value Services. Our primary outcome was the overall rate of low-value service use, defined as the count of 31 low-value services per beneficiary per year by each organization (Schwartz et al. 2014). As described in a previous study of 26 of these services (Schwartz et al. 2014), they were chosen because they provide minimal average clinical benefit in specific clinical scenarios. The services were selected from evidence-based lists published in the American Board of Internal Medicine Foundation's Choosing Wisely initiative (Choosing Wisely 2014), the US Preventive Services Task Force "D" recommendations (U.S. Preventive Services Task Force 2014), and the Canadian Agency for Drugs and Technologies in Health technology assessments (Canadian Agency for Drugs and Technologies in Health), or from the peer-reviewed medical literature (Elshaug et al. 2009). Services were chosen for measure construction if their appropriate use could be distinguished from likely inappropriate overuse with reasonable accuracy using Medicare claims and enrollment data. Because there can be scope for discretion in how to define a low-value service (Schwartz et al. 2014), we employed more specific definitions of low-value services, accepting some loss of sensitivity to reduce the likelihood of misclassifying a service as low value when it had high value in its clinical setting. We grouped services into six clinical categories (Table 1) for the purpose of assessing the consistency of organizations' use of different types of low-value services.

Table 1 presents the operational definitions used to detect each type of low-value service. These definitions incorporate relevant patient demographic and clinical characteristics such as age, sex, and current or past diagnoses. Some measure definitions also rely on the timing of a service (e.g., imaging preceding a surgical operation) or the service setting (e.g., nonemergent). Service occurrences meeting these definitions were detected on the basis of clinical data in claims such as *Current Procedural Terminology (CPT)* codes and *International Classification of Diseases, Ninth Revision (ICD-9)* codes, as well as information from annual beneficiary summary files, such as age and presence of chronic conditions in the Chronic Condition Data Warehouse (CCW) (Centers for Medicare and Medicaid Services 2015). We employed claims data from as early as January 1 of the year before a service occurred in order to evaluate whether the service met the measure definition. Details regarding service detection, including all codes used in detection algorithms, are presented in the Appendix (Table 1).

Because many low-value services do not apply to all patients, we defined qualifying criteria for each service. For example, to qualify for preoperative

Table 1: Measures of Low-Value Services

<i>Clinical Category</i>	<i>Measure</i>	<i>Source</i>	<i>Operational Definition</i>	<i>Qualifying Beneficiaries</i>
Cancer screening	Cancer screening for patients with CKD receiving dialysis	CW (Holley 2007)	Screening for cancer of the breast, cervix, colon, or prostate for patients over age 75 with chronic kidney disease receiving dialysis services*	Patients with CKD [†] receiving dialysis [‡]
	Cervical cancer screening for women age 65 and over	CW USPSTF (Vesco et al. 2011)	Screening Papanicolaou test for women over age 65 with no personal history of cancer or dysplasia noted in claim or in prior claims, and no diagnoses of other female genital cancers, abnormal Papanicolaou findings, or human papillomavirus positivity in prior claims [§]	Women over 65
	Colorectal cancer screening for adults over age 85	USPSTF (Whitlock et al. 2008)	Colorectal cancer screening (colonoscopy, sigmoidoscopy, barium enema, or fecal occult blood testing) for patients age 86 or over with no history of colon cancer	Patients over 75
	PSA testing for men age 75 and over	USPSTF (Lin et al. 2008)	PSA testing for patients age 85 and over with no history of prostate cancer	Men over 85

Continued

Table 1. Continued

Clinical Category	Measure	Source	Operational Definition	Qualifying Beneficiaries
Diagnostic and preventive testing	Bone mineral density testing at frequent intervals	Literature (Hillier et al. 2007; Bell et al. 2009)	Bone mineral density test within 2 years of a prior bone mineral density test for patients with an established osteoporosis diagnosis	Patients with osteoporosis [†]
	Homocysteine testing in cardiovascular disease	Literature (Martí-Carvajal et al. 2013)	Homocysteine testing with no diagnoses of folate or B12 deficiencies in the claim and no folate or B12 testing in prior claims	All patients
	Hypercoagulability testing for patients with deep vein thrombosis	CW (Baglin et al. 2010)	Lab tests for hypercoagulable states within 30 days following diagnosis of lower extremity deep vein thrombosis or pulmonary embolism; no prior evidence of recurrent thrombosis, defined by diagnosis of DVT or pulmonary embolism more than 90 days prior to the testing claim	Patients with deep vein thrombosis [‡]

Continued

Table 1. Continued

Clinical Category	Measure	Source	Operational Definition	Qualifying Beneficiaries
Diagnostic and preventive testing	PTH measurement for patients with stage 1–3 CKD	NICE (Levin et al. 2007; Palmer et al. 2009)	PTH measurement for patients with chronic kidney disease and no dialysis services before PTH testing or within 30 days following testing, as well as no hypercalcemia diagnosis during the year	Patients with CKD [†] not receiving dialysis [‡]
	Total or free T3 level testing for patients with hypothyroidism	CW (Garber et al. 2012)	Total or free T3 measurement in a patient with a hypothyroidism diagnosis during the year	Patients with hypothyroidism [‡]
	1,25-dihydroxyvitamin D testing in the absence of hypercalcemia or decreased kidney function	CW (Holick 2007)	Calcitriol testing for patients without hypercalcemia, secondary hyperparathyroidism of renal origin, or conditions related to non-PTH-mediated hypercalcemia noted in claim (sarcoidosis, TB, selected neoplasms), and without a history of chronic kidney disease; no diagnosis of hypercalcemia in the past 30 days	All patients

Continued

Table 1. *Continued*

<i>Clinical Category</i>	<i>Measure</i>	<i>Source</i>	<i>Operational Definition</i>	<i>Qualifying Beneficiaries</i>
Preoperative testing	Preoperative chest radiography	CW/CADTH (Mohammed et al. 2011; Joo et al. n.d.)	Chest X-ray not associated with inpatient or emergency care [†] and occurring within 30 days prior to a low or intermediate risk noncardiothoracic surgical procedure ^{**}	Patients undergoing selected surgeries [‡]
	Preoperative echocardiography	CW (Douglas et al. 2011)	Echocardiogram not associated with inpatient or emergency care and occurring within 30 days prior to a low or intermediate risk noncardiothoracic surgical procedure ^{**}	Patients undergoing selected surgeries [‡]
	Preoperative PFT	CW (Qaseem et al. 2006)	PFT not associated with inpatient or emergency care and occurring within 30 days prior to a low or intermediate risk surgical procedure ^{††}	Patients undergoing selected surgeries [‡]
	Routine preoperative stress tests	CW (Fleisher et al. 2007)	Stress electrocardiogram, echocardiogram, nuclear medicine imaging, cardiac MRI or CT angiography, not associated with inpatient or emergency care and occurring within 30 days prior to a low or intermediate risk surgical procedure ^{**}	Patients undergoing selected surgeries [‡]

Continued

Table 1. Continued

<i>Clinical Category</i>	<i>Measure</i>	<i>Source</i>	<i>Operational Definition</i>	<i>Qualifying Beneficiaries</i>
Imaging	CT of the sinuses for uncomplicated acute rhinosinusitis	CW (Cornelius et al. 2013)	Maxillofacial CT study with a diagnosis of sinusitis and no complications of sinusitis, ** immune deficiencies, nasal polyps, or head/face trauma noted in claim and no sinusitis diagnosis between 30 and 365 days prior to imaging	All patients
	Head imaging in the evaluation of syncope	CW (Moya et al. 2009)	CT or MR imaging of the head with a diagnosis of syncope and no diagnoses in claim warranting imaging ^{§§}	Patients with syncope diagnosis [†]
	Head imaging for uncomplicated headache	CW (Jordan et al. 2009)	Brain CT or MR imaging with non-post-traumatic, nonthunderclap headache diagnosis, and no diagnoses in claim warranting imaging ^{¶¶}	All patients
	EEG for headaches	CW (Gronseth and Greenberg 1995)	EEG with headache diagnosis in claim, and no epilepsy or convulsions noted in current or prior claims	All patients

Continued

Table 1. *Continued*

<i>Clinical Category</i>	<i>Measure</i>	<i>Source</i>	<i>Operational Definition</i>	<i>Qualifying Beneficiaries</i>
Imaging	Back imaging for patients with nonspecific low back pain	CW, NICE (Chou et al. 2009)	Back imaging with a diagnosis of lower back pain occurring within 6 weeks of initial back pain diagnosis and with no indication of radiculopathy or other diagnoses in claim warranting imaging****	All patients
	Screening for carotid artery disease in asymptomatic adults	CW, USPSTF (Wolff et al. 2007)	Carotid imaging not associated with inpatient or emergency care for patients without a history of stroke or TIA, and without a diagnosis of stroke, TIA, or focal neurological symptoms in claim	All patients
	Screening for carotid artery disease for syncope	CW (Moya et al. 2009)	Carotid imaging with syncope diagnosis for patients without a history of stroke or TIA, and without a diagnosis of stroke, TIA, or focal neurological symptoms in claim	Patients with syncope diagnosis†
	Imaging for diagnosis of plantar fasciitis	CW (Buchbinder 2004)	Radiographic or MR imaging with diagnosis of plantar fasciitis occurring within 2 weeks of initial foot pain diagnosis	Patients with fasciitis diagnosis†

Continued

Table 1. Continued

<i>Clinical Category</i>	<i>Measure</i>	<i>Source</i>	<i>Operational Definition</i>	<i>Qualifying Beneficiaries</i>
Cardiovascular testing and procedures	Stress testing for stable coronary disease	CW (Boden et al. 2007; Hendel et al. 2009)	Stress testing not associated with inpatient or emergency care ^{†††} for patients with an established diagnosis of acute myocardial infarction (≥ 6 months before testing)	IHD patients [†]
	Percutaneous coronary intervention with balloon angioplasty or stent placement for stable coronary disease	Literature (Boden et al. 2007; Lin, Dudley, and Redberg 2007)	Coronary stent placement or balloon angioplasty, not associated with an ER visit, ^{†††} for patients with an established diagnosis of acute myocardial infarction (≥ 6 months before testing)	IHD patients [†]
	Renal artery angioplasty or stenting	Literature (Wheatley et al. 2009; Cooper et al. 2014)	Renal/visceral angioplasty or stent placement with a diagnosis of renal atherosclerosis or renovascular hypertension noted in procedure claim	Patients with hypertension [†]
	Carotid endarterectomy for asymptomatic patients	CW (Wolff et al. 2007; Goldstein et al. 2011)	Carotid endarterectomy, not associated with an ER visit, ^{†††} for female patients without a history of stroke or TIA and without stroke, TIA, or focal neurological symptoms noted in claim	All patients

Continued

Table 1. *Continued*

<i>Clinical Category</i>	<i>Measure</i>	<i>Source</i>	<i>Operational Definition</i>	<i>Qualifying Beneficiaries</i>
Other invasive procedures	Inferior vena cava filters for the prevention of pulmonary embolism	Literature (PREPIC Study Group 2005; Sarosiek, Crowther, and Sloan 2013)	Any IVC filter placement	All patients
	Pulmonary Artery Catheterization in the ICU	Literature (Rajaram et al. 2013)	Pulmonary artery catheterization for monitoring purposes during an inpatient stay that involved an ICU and a nonsurgical MS-DRG; claim contains no diagnoses indicating pulmonary hypertension, cardiac tamponade, or preoperative assessment	Patients who were hospitalized with a nonsurgical MS-DRG [†]
	Vertebroplasty or kyphoplasty for osteoporotic vertebral fractures	Literature (Buchbinder et al. 2009; Kallmes et al. 2009; Boonen et al. 2011)	Vertebroplasty/kyphoplasty for vertebral fracture, with no bone cancers, myeloma, or hemangioma noted in procedure claim	Patients with osteoporosis [†]
	Arthroscopic surgery for knee osteoarthritis	NICE (Laupattarakasem et al. 2008)	Arthroscopic debridement/chondroplasty of the knee with diagnosis of osteoarthritis or chondromalacia in the procedure claim and no meniscal tears noted in procedure claim	All patients

Continued

Table 1. Continued

Clinical Category	Measure	Source	Operational Definition	Qualifying Beneficiaries
	Spinal injection for low back pain	Literature (Staal et al. 2008; Pinto et al. 2012)	Outpatient epidural, facet, or trigger point injections for lower back pain, excluding etanercept; no radiculopathy diagnoses in the claim	All patients

*The age cutoff is included because transplantation is uncommon in this patient population.
†Defined by the presence of CCW first indication date prior to December 31st of the year.
‡Defined by presence of relevant diagnosis or service codes during the year.
§Prior claims refer throughout the table to claims for services before the day of the measured service and during or after the prior calendar year.
¶Inpatient-associated is defined here as occurring during or within 30 days after an inpatient stay; ED-associated, during or 1 day after an ED visit.
**Includes breast procedures, colectomy, cholecystectomy, transurethral resection of the prostate, hysterectomy, orthopedic surgical procedures other than hip and knee replacement, corneal transplant, cataract removal, retinal detachment, hernia repair, lithotripsy, arthroscopy, and cholecystectomy.
††Includes procedures listed immediately above as well as coronary artery bypass graft, aneurysm repair, thromboendarterectomy, percutaneous transluminal coronary angioplasty, and pacemaker insertion.
‡‡Includes inflammation of eyelid or orbit, orbital cellulitis, and visual problems.
§§Exclusion diagnoses include epilepsy, stroke/TIA, history of stroke, head trauma, convulsions, altered mental status, nervous system symptoms (e.g., hemiplegia), and speech problems.
¶¶Exclusion diagnoses include those listed immediately above as well as giant cell arteritis, cancer, and history of cancer.
***Exclusion diagnoses include cancer, trauma, intravenous drug abuse, neurological impairment, endocarditis, septicemia, tuberculosis, osteomyelitis, fever, weight loss, loss of appetite, night sweats, and anemia.
†††Inpatient-associated is defined here as occurring during an inpatient stay; ED-associated, during or within 14 days after an ED visit.
CKD, chronic kidney disease; CT, computed tomography; CW, Choosing Wisely; ED, emergency department; EEG, electroencephalography; ICU, intensive care unit; IVC, inferior vena cava; MR, magnetic resonance; MS-DRG, Medicare Severity Diagnosis Related Group; NICE, National Institute for Health and Care Excellence; PFT, pulmonary function testing; PSA, prostate-specific antigen; PTH, parathyroid hormone; TIA, transient ischemic attack; USPSTF, US Preventive Services Taskforce.

testing measures, beneficiaries must have undergone surgery (Table 1). These qualifying criteria were important for accurately characterizing an organization's tendency to use low-value services, as some organizations' patients may have more frequent opportunities to receive low-value services (e.g., more indications for surgery prompting more low-value preoperative testing). We attempted to minimize the sensitivity of qualifying criteria to variation in organizations' diagnostic coding or practice practices. For example, the qualifying criterion for detection of vertebroplasty or kyphoplasty for osteoporotic vertebral fractures was a diagnosis of osteoporosis rather than a diagnosis of a vertebral compression fracture, because compression fractures may be coded or diagnosed by imaging to varying extents across organizations or may be coded only when a vertebroplasty or kyphoplasty is performed.

Covariates. For case mix adjustment, we also assessed the following patient characteristics from Medicare beneficiary summary files: age, sex, race/ethnicity, disability as the initial reason for Medicare enrollment, end-stage renal disease, and 27 chronic conditions recorded in the CCW, which draws from Medicare claims since 1999 to describe beneficiaries' accumulated burden of chronic disease. From the 2007–2011 American Community Survey Summary File (U.S. Census Bureau 2012), we also obtained the following characteristics of the population aged 65 and over in each beneficiary's ZIP code tabulation area (ZCTA): median income, proportion of residents below the federal poverty level (FPL), proportion of residents with a high school degree, and proportion of residents with a college degree.

Statistical Analysis

We estimated three quantities describing organizational patterns of low-value service use: (1) variation across organizations in rates of low-value service use per beneficiary; (2) the persistence of organizations' rates of low-value service use over time; and (3) correlations between rates of use of different categories of low-value services within organizations. Constructing each estimate involved three general steps. First, for each of the 31 low-value services, we estimated the rate of service use in each organization (i.e., the count of services received per qualifying beneficiary), adjusted for case mix. Second, these adjusted rates for each service were combined to create composite measures of each organization's overall rate of low-value service use, or composite rates within service categories, and estimates of sampling variances of the measures.

Third, we estimated the parameters of interest by fitting multilevel models to the composite measures and accompanying variance estimates, with multiple outcomes for the correlational measures of persistence and consistency (Fay and Herriot 1979; Zaslavsky 2007; O'Malley and Zaslavsky 2008). This approach follows established practices for analyzing composite measures of quality that are based on multiple quality components (Agency for Healthcare Research and Quality 2015). A complete description of these methods and the particular models we fit are included in the Appendix.

Specifically, use of each low-value service by patients attributed to each organization was adjusted for case mix via linear modeling of the count of low-value services as a function of the clinical and sociodemographic covariates listed above, a vector of year fixed effects, and a vector of organization fixed effects. The unit of observation in this first step was the beneficiary-year. Models included a term for age squared, and CCW conditions were specified as indicators for each condition as well as indicators for the condition count (2, 3, 4, etc. up to ≥ 9). Estimating a separate model for each low-value service allowed for service-specific case mix adjustment. Each model included data from only those beneficiaries who satisfied the qualifying criteria for the service. Importantly, we included all low-value services received by a patient in our analysis, including those received outside of the attributed provider organization.

Composite measures were calculated as the sum of an organization's estimated fixed effects from models for different services, weighted by the proportion of patient-level observations across all organizations satisfying a measure's qualifying criteria (Appendix). We chose to employ weights based on the average rate of eligible patients across all organizations, rather than employing different weights for each organization based on their eligible patient population. With our weighting scheme, the composite measure can be interpreted as the total number of low-value services per patient per year the organization would be expected to deliver to a standardized population of patients, plus a constant. The constant exists because the composite measures are relative and represent the difference in performance between an organization and a reference organization.

We then used multilevel models to estimate organizational variation, persistence, and consistency (Appendix). Our modeling strategy addresses three challenges to assessing variation in composite outcomes across units of observation. First, there can be substantial sampling error when observing a limited number of patients within each organization (Zaslavsky 2001; Dimick, Welch, and Birkmeyer 2004). As a result of this additional variation, the observed mean outcomes are overdispersed. In other words, they vary more

across organizations than do the long-run means that would be obtained with very large sample sizes. The multilevel model corrects the bias in estimates of variation due to overdispersion and, when the outcome is multivariate, corrects estimates of correlation as well.

Second, modeling of our composite measures is complicated by the fact that different measures apply to different patients, due to their specificity to one sex or to a particular clinical scenario. For example, women cannot qualify for low-value PSA testing. The resulting incompleteness of the data matrix could be problematic for person-level modeling of multiple outcomes. We used the aggregated or “area-level” model of Fay and Herriot (1979) to accommodate this data structure. This model first reduces the data for each organization to two approximately sufficient statistics: (1) a composite point estimate, the adjusted mean of the organization’s observed rates on the included measures among the respective eligible patients; and (2) a variance estimate for the point estimator, calculated using standard Taylor series methods for sample surveys. We then estimate models of organizational variation, consistency or persistence using the point estimators and their variance estimates (see Appendix for details and equations). Our approach provides a straightforward method to account for nonrandom missing outcomes for beneficiaries who do not qualify for particular low-value services. Because every organization serves some beneficiaries that qualify for each low-value service measure (as confirmed at the outset of our analysis), there are no missing outcomes at the organization-level. Organization-level analysis thus obviates the problem of jointly modeling multiple patient-level outcomes when patients tend to qualify for only some of the outcomes. Third, individual-level modeling of multiple outcomes can be computationally intensive. An additional advantage of our approach is avoiding the computational intensity of jointly modeling 31 outcomes in over 10 million observations in thousands of organizations.

In our analysis of organizational variation, the quantity of interest was the between-organization variation in the overall low-value service composite rate, which we present as a standard deviation. To aid interpretation, we also present a corresponding ratio of the adjusted use of low-value services in an organization at the 90th percentile to that of an organization at the 10th percentile, as predicted by the estimated standard deviation (Institute of Medicine 2013). In order to explore whether organizational variation is extensive within regions, we repeated our variation analysis controlling for hospital referral region (HRR) fixed effects.

In order to explore whether patient preferences might explain observed variation, we also repeated the variation analysis separately for two sets of

low-value services: services that are more likely to be sensitive to patient preferences, and services that are less likely to be sensitive to patient preferences (Schwartz et al. 2015). For example, patients are unlikely to request testing for hypercoagulability following a deep venous thrombosis (DVT) as they are unlikely to know that such tests exist, let alone whether they are indicated following a DVT.

Models for persistence and consistency each included vectors of random effects for different years or categories of services, with an unstructured covariance matrix to characterize correlations between these effects. For the analysis of persistence in organizational behavior, the parameter of interest was the correlation between organizations' low-value service composite rates in 2010 and 2011. For the analysis of consistency in organizational behavior, the parameters of interest were the pairwise correlations between organizations' different service category composite rates. In order to test whether organizations deviated from average local behavior consistently across service types, we performed a secondary consistency analysis that included patient region in case mix adjustment (Table A2). Because variation across organizations in case mix could contribute to greater variation and higher consistency, we repeated analyses of variation and consistency without adjustment for CCW conditions to gauge the degree to which adjustment for patients' disease burden affected estimates (Table A2). We calculated 95 percent confidence intervals for estimates via bootstrapping.

RESULTS

After exclusions, the general sample consisted of 4,039,733 beneficiaries attributed to 3,137 organizations defined by TINs (Table 2). Of these organizations, 90.1 percent appeared in claims for the study sample in every study year, and the median number of physicians primarily associated with these organizations in each year was 16 (mean, 45.5). Among TINs in our study sample, the average unadjusted rate of low-value service receipt was of 45.6 low-value services per 100 beneficiaries. This corresponds to an average of 30.4 percent of beneficiaries receiving at least one low-value service each year. The secondary ACO sample consisted of 1,432,644 beneficiaries attributed to 250 ACOs. Organizations in the ACO sample were considerably larger, with an average of 5,731 attributed beneficiaries compared to 1,288 in the general sample. On average, beneficiaries attributed to ACOs received 47.7 low-value services per 100 beneficiaries.

Table 2: Unadjusted Beneficiary and Provider Organization Characteristics

	<i>General Sample</i>	<i>Secondary ACO Sample</i>
Beneficiaries, no.	4,039,733	1,432,644
Observations, no. of beneficiary-years	10,149,111	3,580,702
Mean age, years	72.6 $\pm$ 11.7	72.4 $\pm$ 11.9
Female sex, %	58.4	58.2
Race/ethnicity, %		
White	89.4	87.3
Black	7.4	7.2
Hispanic	0.9	1.7
Other	2.2	3.8
Medicaid recipient, %	16.7	20.8
Disabled, [†] %	20.3	21.1
End-stage renal disease, %	1.0	1.1
CCW conditions [‡]		
Total no., mean	5.1 $\pm$ 2.6	5.2 $\pm$ 2.7
≥ 6 conditions	43.8	46.1
≥ 9 conditions	16.0	17.8
Low-value service measure qualified for, mean	14.9 $\pm$ 2.5	14.9 $\pm$ 2.5
ZCTA characteristics for age 65+, mean		
Median income	38,801 $\pm$ 13,533	39,706 $\pm$ 14,161
% below FPL	8.5	8.4
% with high school degree	77.2	77.3
% with college degree	20.3	21.1
Provider organizations, no.	3,137	250
Assigned beneficiaries per organization, mean no.	1,288 $\pm$ 1,447	5,731 $\pm$ 5,135
Physicians per organization, mean no.	45.5 $\pm$ 89.1	301.7 $\pm$ 369.36
Low-value services per 100 beneficiaries, mean		
Cancer screening	14.0 $\pm$ 5.6	14.0 $\pm$ 3.8
Diagnostic and preventive testing	7.8 $\pm$ 0.6	9.2 $\pm$ 4.2
Preoperative testing	2.5 $\pm$ 1.2	2.6 $\pm$ 1.0
Imaging	15.7 $\pm$ 5.2	16.3 $\pm$ 4.0
Cardiovascular testing and procedures	1.1 $\pm$ 0.6	1.2 $\pm$ 0.4
Other invasive procedures	4.4 $\pm$ 2.2	4.4 $\pm$ 1.6
Total	45.6 $\pm$ 13.8	47.7 $\pm$ 11.4

Notes. Estimates are derived from 2007 to 2011 data. All means and percentages are unadjusted. Means are presented $\pm$ standard deviations.

[†]Refers to beneficiaries for whom disability was the original reason for Medicare eligibility.

[‡]Chronic conditions include 25 conditions from the CCW: acute myocardial infarction, Alzheimer's disease, Alzheimer's disease and related disorders or senile dementia, anemia, asthma, atrial fibrillation, benign prostatic hyperplasia, breast cancer, chronic kidney disease, chronic obstructive pulmonary disease, colorectal cancer, depression, diabetes, endometrial cancer, heart failure, hip/pelvic fracture, hyperlipidemia, hypertension, hypothyroidism, ischemic heart disease, lung cancer, osteoporosis, prostate cancer, rheumatoid arthritis/osteoarthritis, stroke/transient ischemic attack.

ACO, accountable care organization; CCW, chronic conditions warehouse; FPL, federal poverty level; HCC, hierarchical condition categories; ZCTA, ZIP code tabulation area.

Figure 1 displays the distribution of adjusted estimates of low-value services across organizations. Table 3 presents adjusted estimates of variation in organizations' annual delivery of low-value services. The between-organization standard deviation in the use of low-value services was 10.0 services per 100 beneficiaries (95 percent CI, 9.4–40.5), with 57.7 and 32.2 services per 100 beneficiaries at the 90th and 10th percentile of organizations, respectively, corresponding to a 90th/10th percentile ratio of 1.78 (95 percent CI 1.72–1.84). Without adjustment for overdispersion, the between-organization standard deviation would have been estimated to be 5 percent higher (10.5 services per 100 beneficiaries). After adjusting for patients' HRR of residence, the 90th/10th percentile ratio was 1.51 (95 percent CI, 1.46–1.54). The between-organization variation without adjustment for geography was modestly higher after excluding CCW indicators from models (90th/10th percentile ratio, 1.89; 95 percent CI, 1.83–1.96). Variation in services categorized as more sensitive to patient preferences (90th/10th percentile ratio, 1.61; 95 percent CI, 1.58–1.64) was substantially less than variation in services categorized as less sensitive to patient preferences (90th/10th percentile ratio, 2.84;

Figure 1: Distribution of Adjusted Rates of Low-Value Services

Notes: Estimates are predicted modes derived from random intercept model estimation of organizations' total composite low-value service use, using 2007–2011 data. The model adjusts for overdispersion of observed performance, beneficiary sociodemographic and clinical characteristics, local area economic characteristics, year, and the number of patients qualifying for component measure denominator. [Color figure can be viewed at wileyonlinelibrary.com]

Table 3: Variation in Low-Value Service Delivery

	<i>Variation in Low-Value Services Per 100 Beneficiaries</i>			
	<i>Standard Deviation</i>	<i>95% CI</i>	<i>90th/10th Percentile Ratio</i>	<i>95% CI</i>
Baseline model	10.0	9.4–10.5	1.78	1.72–1.84
With adjustment for region	8.8	8.3–9.3	1.66	1.60–1.71
Without adjustment for chronic conditions	11.0	10.4–11.5	1.89	1.83–1.96
High preference sensitivity services only	5.2	5.0–5.3	1.61	1.58–1.64
Low preference sensitivity services only	6.3	5.8–6.8	2.84	2.60–3.13

Notes. Estimates are derived from models of organizations’ total composite low-value service use. Standard deviation estimates refer to across-organization variation, estimated via a random intercept model that includes 2007–2011 data. The baseline model adjusts for overdispersion of observed performance, beneficiary sociodemographic and clinical characteristics, local area economic characteristics, year, and the number of patients qualifying for component measure denominators. The model with regional adjustment also adjusts for patient hospital referral region. The model without adjustment for chronic conditions does not contain indicators for conditions present in the Chronic Conditions Data Warehouse.

95 percent CI, 2.60–3.13). Adjusted low-value service use was highly correlated between years (r , 0.98; 95 percent CI 0.97–0.99) among 2,682 organizations with more than 200 attributed beneficiaries in both 2010 and 2011.

Within organizations, service use was positively correlated between almost all pairs of categories of low-value services (Table 4). Adjusted correlations between categories were positive and statistically significant for 13 of 15 pairs of service categories in the general sample and 12 of 15 pairs in the ACO sample. All nonsignificant correlations involved a single category of low-value services—other invasive procedures. The mean correlation coefficient across all pairs was 0.26 in the general sample (95 percent CI, 0.24–0.28) and 0.37 in the ACO sample (95 percent CI, 0.31–0.43). The corresponding means were 0.35 and 0.51 among the pairs that did not include other invasive procedures. In both samples, the greatest correlation was between low-value cardiovascular testing and procedures and low-value imaging. Low-value imaging was most highly correlated with other categories, with a mean of 0.36 in the general sample and 0.49 in the ACO sample. In the general sample, the mean correlation coefficient between pairs of low-value service categories was slightly lower after adjustment for HRR (r , 0.21; 95 percent CI, 0.21–0.25) and minimally higher after excluding chronic conditions from the case mix adjustment (r , 0.28; 95 percent CI, 0.26–0.30) (Table A2).

Table 4: Consistency Across Low-Value Service Domains

Measure composite	Cancer Screening	Diag.	Preop.	Imaging	Cardio.	Other.
General sample						
Cancer screening	—	—	—	—	—	—
Diagnostic and preventive testing	0.30 (0.26–0.35)	—	—	—	—	—
Preoperative testing	0.33 (0.29–0.37)	0.28 (0.22–0.34)	—	—	—	—
Imaging	0.47 (0.43–0.51)	0.28 (0.23–0.32)	0.35 (0.31–0.39)	—	—	—
Cardiovascular testing and procedures	0.33 (0.29–0.37)	0.28 (0.24–0.33)	0.36 (0.32–0.41)	0.54 (0.49–0.58)	—	—
Other invasive procedures	0.11 (0.07–0.16)	0.03 (–0.02 to 0.08)	–0.04 (–0.10 to 0.02)	0.15 (0.11–0.19)	0.10 (0.05–0.16)	—
ACO sample						
Cancer screening	—	—	—	—	—	—
Diagnostic and preventive testing	0.41 (0.27–0.54)	—	—	—	—	—
Preoperative testing	0.55 (0.46–0.64)	0.47 (0.37–0.57)	—	—	—	—
Imaging	0.60 (0.51–0.70)	0.53 (0.42–0.65)	0.50 (0.40–0.60)	—	—	—
Cardiovascular testing and procedures	0.45 (0.34–0.56)	0.49 (0.39–0.59)	0.45 (0.37–0.53)	0.65 (0.54–0.76)	—	—
Other invasive procedures	0.04 (–0.13 to 0.20)	0.06 (–0.11 to 0.23)	0.01 (–0.15 to 0.17)	0.16 (0.01–0.30)	0.22 (0.08–0.36)	—

Notes: Estimates are correlation coefficients and their 95% confidence intervals. For each sample of organizations, all correlations are derived from a single model of organizations' low-value service composite measures. The models adjust for overdispersion of observed performance, beneficiary sociodemographic and clinical characteristics, local area economic characteristics, year, and the number of patients qualifying for component measures. ACO, accountable care organization.

DISCUSSION

In this national study of low-value service use among provider organizations, we observed substantial variation in service use across organizations, highly persistent service use over time, and positive associations between use of various categories of low-value services within organizations. We also observed that low-value service use varied substantially across organizations within the same HRR, consistent with prior studies of within-HRR variation in overall spending (Zhang et al. 2012). Despite the considerable variation across organizations, low-value services were used frequently even among the best performing organizations. Notably, the amount of low-value service use at organizations in the 10th percentile was greater than the difference in low-value service use between the 10th percentile and 90th percentile.

Correlations between different clinical categories of low-value service use were generally positive and significant, suggesting some consistency of low-value practices within organizations. Because these correlations were modest in strength, many measures of low-value practices may be required to comprehensively assess organizational performance with respect to provision of unnecessary care. Correlations between different categories of low-value service use were higher for ACOs than for TINs in the general sample, consistent with ACOs (as larger collections of TINs) having influence over a greater range of providers and greater proportion of opportunities for low-value services in a patient's care. However, even for ACOs, we observed little association between use of some surgical procedures and use of other low-value services. This finding suggests that some physician specialties may be less influenced either by practice patterns of colleagues in other specialties or by an organization's efforts to increase or decrease use of low-value services. This finding also likely reflects the fact that many ACOs do not include surgical specialists (McWilliams et al. 2016).

More generally, our findings are consistent with, though not definitive evidence of, provider organizations influencing the amount of low-value care that their patients receive. If organizations have historically shaped the practice patterns of their member physicians, alternative payment models that modify organization-level incentives may be more likely to reduce unnecessary care, either by leveraging organizational mechanisms that have already led to less low-value care or discouraging behavior that has led to more low-value care. In other words, our findings support the implicit assumption

underlying ACO-like payment models, namely that provider organizations can shape the value of health care provided to their patients.

There are several limitations to this study. First, it is a descriptive analysis, and the results may differ from those that would be observed if patients and physicians were randomized to different provider organizations. Distinct organizational profiles of service use may be driven in part by physicians and patients sorting among organizations because of unobserved factors like training or preferences. Whatever the cause of the variation in low-value care that we observe across organizations and the consistency of practice patterns across different types of low-value care within organizations, these findings suggest that organizations may be particularly suitable targets for incentives to change behavior. A more uniform distribution of low-value care across organizations, a lack of organizational consistency across low-value services, or a lack of organizational persistency over time would have suggested less utility from policies relying on organizations to influence value, as such results would suggest no evidence of organizational influence to date.

Specific sources of potential confounding merit discussion. Although our analyses adjusted for many sociodemographic and clinical characteristics of patients, it is possible that unobserved differences in patient characteristics or market factors may have contributed to organizational variation and consistency in low-value service use. Adjustment for chronic conditions, however, reduced the standard deviation in between-organization low-value service counts by less than 10 percent and minimally affected within-organization correlations between low-value service counts in different categories, suggesting that differences in case mix were not a major contributor to our findings. This is not surprising given that our measures of low-value care attempt to identify services that should not be provided to patients, regardless of their clinical characteristics. In addition, patient preferences are unlikely to have driven the variation in low-value services use, as organizational variation in the use of services that are less sensitive to patient preferences was greater than variation in the use of services that are more sensitive to patient preferences. This finding also suggests that provider behavior, rather than patient demand, largely drives organizational variation in the provision of low-value services. In addition, the substantial variation in low-value service use we observed within geographic regions suggests that local market factors at the HRR level did not drive much of the overall variation in low-value service use we observed.

Second, although we observed variation in organizational rates of low-value service use, we could not isolate which characteristics of organizations drove this variation. Of note, prior work has demonstrated that organizational

characteristics were associated with heterogeneity in responses to new Medicare payment contracts, with some evidence of greater spending reductions among independent primary care groups than hospital-integrated groups and greater reductions in low-value service use among Pioneer ACOs that provided more low-value care at baseline (Schwartz et al. 2015; McWilliams et al. 2016). Third, although we attempted to craft specific claims-based definitions of low-value services, there may be some misclassification of high-value services as low-value. Relatedly, because some measures of low-value services depend on diagnostic codes in claims, variation in measured service use might reflect differences in the completeness of organizations' diagnostic coding. However, many services we examined were not defined on the basis of diagnoses in claims. Fourth, the strength of correlations across services may have been weakened because organizations in our sample, especially smaller organizations, may not have delivered all of the types of services that we measured. This was the major motivation for examining consistency of low-value services in larger organizations such as ACOs. Finally, it is unclear whether variation in use of low-value services is associated with variation in other outcomes of interest, including spending, health status, and patient satisfaction.

In conclusion, our study indicates that organizations have exhibited distinct profiles in their provision of low-value services, with substantial performance variations that are persistent over time and consistent across different types of low-value services. These findings suggest an organizational imprint on health care value that supports a key premise of payment models attempting to foster accountability at an organizational level.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors are grateful to Jesse B. Dalton, M.A., for research assistance and to Pasha Hamed, M.A., for statistical programming support. This study was supported by grants from the National Institutes of Health (F30 AG044106, U01 MH103018, P01 AG032952). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding sources did not play a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript.

Disclosures: The authors of this article (J. M. M. and A. L. S.) have been compensated as consultants by the Medicare Payment Advisory Commission

to aid in their efforts to measure low-value care in Medicare for research purposes. Dr. Schwartz also reports serving as a paid consultant on broad issues of quality measurement to Nuna, Inc. Dr. Chernew reports that he is a partner in VBID Health, LLC, which has a contract with Milliman to develop and market a tool to help insurers and employers quantify spending on low-value services.

Disclaimers: None.

REFERENCES

- Abaluck, J., L. Agha, C. Kabrhel, A. Raja, and A. Venkatesh. 2014. *Negative Tests and the Efficiency of Medical Care: What Determines Heterogeneity in Imaging Behavior?* Working paper No. 19956. Cambridge, MA: National Bureau of Economic Research.
- Agency for Healthcare Research and Quality. 2015. "Instructions for Analyzing Data from CAHPS Surveys" [accessed on December 15, 2015]. Available at <https://cahps.ahrq.gov/surveys-guidance/survey4.0-docs/2015-Instructions-for-Analyzing-Data-from-CAHPS-Surveys.pdf>
- Asch, D. A., S. Nicholson, S. Srinivas, J. Herrin, and A. J. Epstein. 2009. "Evaluating Obstetrical Residency Programs Using Patient Outcomes." *Journal of the American Medical Association* 302 (12): 1277-83. doi:10.1001/jama.2009.1356.
- Asch, D. A., A. Epstein, and S. Nicholson. 2007. "Evaluating Medical Training Programs by the Quality of Care Delivered by Their Alumni." *Journal of the American Medical Association* 298 (9): 1049-51. doi:10.1001/jama.298.9.1049.
- Baglin, T., E. Gray, M. Greaves, B. J. Hunt, D. Keeling, S. Machin, I. Mackie, M. Makris, T. Nokes, D. Perry, R. C. Tait, I. Walker, and H. Watson. 2010. "Clinical Guidelines for Testing for Heritable Thrombophilia." *British Journal of Haematology* 149 (2): 209-20. doi:10.1111/j.1365-2141.2009.08022.x.
- Bell, K. J. L., A. Hayen, P. Macaskill, L. Irwig, J. C. Craig, K. Ensrud, and D. C. Bauer. 2009. "Value of Routine Monitoring of Bone Mineral Density after Starting Bisphosphonate Treatment: Secondary Analysis of Trial Data." *British Medical Journal* 338. doi:10.1136/bmj.b226.
- Boden, W. E., R. A. O'Rourke, K. K. Teo, P. M. Hartigan, D. J. Maron, W. J. Kostuk, M. Knudtson, M. Dada, P. Casperson, C. L. Harris, B. R. Chaitman, L. Shaw, G. Gosselin, S. Nawaz, L. M. Title, G. Gau, A. S. Blaustein, D. C. Booth, E. R. Bates, J. A. Spertus, D. S. Berman, G. B. J. Mancini, and W. S. Weintraub. 2007. "Optimal Medical Therapy with or without PCI for Stable Coronary Disease." *New England Journal of Medicine* 356 (15): 1503-16. doi:10.1056/NEJMoa070829.
- Boonen, S., J. Van Meirhaeghe, L. Bastian, S. R. Cummings, J. Ranstam, J. B. Tillman, R. Eastell, K. Talmadge, and D. Wardlaw. 2011. "Balloon Kyphoplasty for the Treatment of Acute Vertebral Compression Fractures: 2-Year Results from a Randomized Trial." *Journal of Bone and Mineral Research* 26 (7): 1627-37. doi:10.1002/jbmr.364.

- Bright, T. J., A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R. R. Coeytaux, G. Samsa, V. Hasselblad, J. W. Williams, M. D. Musty, L. Wing, A. S. Kendrick, G. D. Sanders, and D. Lobach. 2012. "Effect of Clinical Decision-Support Systems: A Systematic Review." *Annals of Internal Medicine* 157 (1): 29–43. doi:10.7326/0003-4819-157-1-201207030-00450.
- Buchbinder, R. 2004. "Plantar Fasciitis." *New England Journal of Medicine* 350 (21): 2159–66. doi:10.1056/NEJMcp032745.
- Buchbinder, R., R. H. Osborne, P. R. Ebeling, J. D. Wark, P. Mitchell, C. Wriedt, S. Graves, M. P. Staples, and B. Murphy. 2009. "A Randomized Trial of Vertebroplasty for Painful Osteoporotic Vertebral Fractures." *New England Journal of Medicine* 361 (6): 557–68. doi:10.1056/NEJMoa0900429.
- Burwell, S. M. 2015. "Setting Value-Based Payment Goals—HHS Efforts to Improve U.S. Health Care." *New England Journal of Medicine* 372: 897–9. doi:10.1056/NEJMp1500445.
- Canadian Agency for Drugs and Technologies in Health. "Health Technology Assessments" [accessed on February 16, 2015]. Available at <http://cadth.ca/en/products/health-technology-assessment>
- Centers for Medicare and Medicaid Services. 2015. "Chronic Conditions Data Warehouse" [accessed on December 15, 2015]. Available at <http://www.ccwdata.org/>
- Choosing Wisely. 2014. "Lists of Five Things Physicians and Patients Should Question" [accessed on February 16, 2015]. Available at <http://www.choosing-wisely.org/doctor-patient-lists/>
- Chou, R., R. Fu, J. A. Carrino, and R. A. Deyo. 2009. "Imaging Strategies for Low-Back Pain: Systematic Review and Meta-Analysis." *Lancet* 373 (9662): 463–72. doi:10.1016/S0140-6736(09)60172-0.
- Code of Federal Register. Title 42. Part 425.110. "Number of ACO Professionals and Beneficiaries."
- Cooper, C. J., T. P. Murphy, D. E. Cutlip, K. Jamerson, W. Henrich, D. M. Reid, D. J. Cohen, A. H. Matsumoto, M. Steffes, M. R. Jaff, M. R. Prince, E. F. Lewis, K. R. Tuttle, J. I. Shapiro, J. H. Rundback, J. M. Massaro, R. B. D'Agostino, and L. D. Dworkin. 2014. "Stenting and Medical Therapy for Atherosclerotic Renal Artery Stenosis." *New England Journal of Medicine* 370 (1): 13–22. doi:10.1056/NEJMoa1310753.
- Cornelius, R. S., J. Martin, F. J. Wippold, A. H. Aiken, E. J. Angtuaco, K. L. Berger, D. C. Brown, P. C. Davis, C. T. McConnell, L. L. Mechtler, B. Nussenbaum, C. J. Roth, and D. J. Seidenwurm. 2013. "ACR Appropriateness Criteria Sinonasal Disease." *Journal of the American College of Radiology* 10 (4): 241–6. doi:10.1016/j.jacr.2013.01.001.
- Currie, J., W. B. MacLeod, and J. Van Parys. 2015. *Physician Practice Style and Patient Health Outcomes: The Case of Heart Attacks*, NBER Working Paper. Cambridge, MA.
- Dimick, J. B., G. Welch, and J. D. Birkmeyer. 2004. "Surgical Mortality as an Indicator of Hospital Quality: The Problem with Small Sample Size." *Journal of the American Medical Association* 292 (7): 847–51. doi:10.1001/jama.292.7.847.

- Douglas, P. S., M. J. Garcia, D. E. Haines, W. W. Lai, W. J. Manning, A. R. Patel, M. H. Picard, D. M. Polk, M. Ragosta, R. P. Ward, and R. B. Weiner. 2011. "ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for Echocardiography: A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association." *Journal of the American College of Cardiology* 57 (9): 1126–66. doi:10.1016/j.jacc.2010.11.002.
- Doyle, J. J., S. M. Ewer, and T. H. Wagner. 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams." *Journal of Health Economics* 29 (6): 866–82. doi:10.1016/j.jhealeco.2010.08.004.
- Doyle, J. J., J. A. Graves, J. Gruber, and S. A. Kleiner. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy* 123 (1): 170–214. doi:10.1086/677756.
- Elshaug, A. G., J. R. Moss, P. Littlejohns, J. Karnon, T. L. Merlin, and J. E. Hiller. 2009. "Identifying Existing Health Care Services That Do Not Provide Value for Money." *Medical Journal of Australia* 190 (5): 269–73.
- Fay, R., and R. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74 (366): 269–77.
- Fleisher, L. A., J. A. Beckman, K. A. Brown, H. Calkins, E. Chaikof, K. E. Fleischmann, W. K. Freeman, J. B. Froehlich, E. K. Kasper, J. R. Kersten, B. Riegel, J. F. Robb, S. C. Smith, A. K. Jacobs, C. D. Adams, J. L. Anderson, E. M. Antman, C. E. Buller, M. A. Creager, S. M. Ettinger, D. P. Faxon, V. Fuster, J. L. Halperin, L. F. Hiratzka, S. A. Hunt, B. W. Lytle, R. Nishimura, J. P. Ornato, R. L. Page, L. G. Tarkington, and C. W. Yancy. 2007. "ACC/AHA 2007 Guidelines on Perioperative Cardiovascular Evaluation and Care for Noncardiac Surgery: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guideline)." *Circulation* 116 (17): e418–99. doi:10.1161/CIRCULATIONAHA.107.185699.
- Garber, J. R., R. H. Cobin, H. Gharib, J. V. Hennessey, I. Klein, J. I. Mechanick, R. Pesah-Pollack, P. A. Singer, and K. A. Woeber. 2012. "Clinical Practice Guidelines for Hypothyroidism in Adults: Cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association." *Endocrine Practice* 18 (6): 988–1028. doi:10.4158/EP12280.GL.
- Goldstein, L. B., C. D. Bushnell, R. J. Adams, L. J. Appel, L. T. Braun, S. Chaturvedi, M. A. Creager, A. Culebras, R. H. Eckel, R. G. Hart, J. A. Hinchey, V. J. Howard, E. C. Jauch, S. R. Levine, J. F. Meschia, W. S. Moore, J. V. I. Nixon, T. A. Pearson, American Heart Association Stroke Council; Council on Cardiovascular Nursing; Council on Epidemiology and Prevention; Council for High Blood Pressure Research; Council on Peripheral Vascular Disease, and Interdisciplinary Council on Quality of Care and Outcomes Research. 2011. "Guidelines for the Primary Prevention of Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association." *Stroke* 42 (2): 517–84. doi:10.1161/STR.0b013e3181fcb238.

- Gronseth, G. S., and M. K. Greenberg. 1995. "The Utility of the Electroencephalogram in the Evaluation of Patients Presenting with Headache: A Review of the Literature." *Neurology* 45 (7): 1263–7.
- Hendel, R. C., D. S. Berman, M. F. Di Carli, P. A. Heidenreich, R. E. Henkin, P. A. Pellikka, G. M. Pohost, K. A. Williams, American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Nuclear Cardiology, American College of Radiology, American Heart Association, American Society Echocardiography, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, and Society of Nuclear Medicine. 2009. "ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate Use Criteria for Cardiac Radionuclide Imaging: A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the American Society of Nuclear Cardiology, the American Col." *Circulation* 119 (22): e561–87. doi:10.1161/CIRCULATIONAHA.109.192519.
- Hillier, T. A., K. L. Stone, D. C. Bauer, J. H. Rizzo, K. L. Pedula, J. A. Cauley, K. E. Ensrud, M. C. Hochberg, and S. R. Cummings. 2007. "Evaluating the Value of Repeat Bone Mineral Density Measurement and Prediction of Fractures in Older Women: The Study of Osteoporotic Fractures." *Archives of Internal Medicine* 167 (2): 155–60. doi:10.1001/archinte.167.2.155.
- Holick, M. F. 2007. "Vitamin D Deficiency." *New England Journal of Medicine* 357 (3): 266–81. doi:10.1056/NEJMr070553.
- Holley, J. L. 2007. "Screening, Diagnosis, and Treatment of Cancer in Long-Term Dialysis Patients." *Clinical Journal of the American Society of Nephrology* 2 (3): 604–10. doi:10.2215/CJN.03931106.
- Institute of Medicine. 2013. *Variation in Health Care Spending: Target Decision Making, Not Geography*, edited by J. P. Newhouse, A. M. Garber, R. P. Graham, M. A. McCoy, M. Mancher, and A. Kibria. Washington, DC: National Academies Press.
- Jackson, G. L., B. J. Powers, R. Chatterjee, J. Prvu Bettger, A. R. Kemper, V. Hasselblad, R. J. Dolor, R. J. Irvine, B. L. Heidenfelder, A. S. Kendrick, R. Gray, and J. W. Williams. 2013. "The Patient-Centered Medical Home: A Systematic Review." *Annals of Internal Medicine* 158 (3): 169. doi:10.7326/0003-4819-158-3-201302050-00579.
- Javitt, J. C., J. B. Rebitzer, and L. Reisman. 2008. "Information Technology and Medical Missteps: Evidence from a Randomized Trial." *Journal of Health Economics* 27 (3): 585–602. doi:10.1016/j.jhealeco.2007.10.008.
- Joo, H. S., J. Wong, V. N. Naik, and G. L. Savoldelli. n.d. "The Value of Screening Pre-operative Chest X-Rays: A Systematic Review." *Canadian Journal of Anaesthesia* 52 (6): 568–74. doi:10.1007/BF03015764.
- Jordan, J., F. I. Wippold, R. Cornelius, S. Amin-Hanjani, J. Brunberg, P. Davis, R. De La Paz, D. Dormont, I. Germano, L. Gray, S. Mukherji, S. DJ, M. Sloan, P. Turski, R. Zimmerman, G. Zipfel, and Expert Panel on Neurologic Imaging. 2009. *ACR Appropriateness Criteria Headache*. Reston, VA.
- Kallmes, D. F., B. A. Comstock, P. J. Heagerty, J. A. Turner, D. J. Wilson, T. H. Diamond, R. Edwards, L. A. Gray, L. Stout, S. Owen, W. Hollingworth, B. Ghdoke,

- D. J. Annesley-Williams, S. H. Ralston, and J. G. Jarvik. 2009. "A Randomized Trial of Vertebroplasty for Osteoporotic Spinal Fractures." *The New England Journal of Medicine* 361 (6): 569–79. doi:10.1056/NEJMoa0900563.
- Keating, N. L., E. M. Kouri, Y. He, R. A. Freedman, R. Volya, and A. M. Zaslavsky. 2016. "Location Isn't Everything: Proximity, Hospital Characteristics, Choice of Hospital, and Disparities for Breast Cancer Surgery Patients." *Health Services Research* 51 (4): 1561–83. doi:10.1111/1475-6773.12443.
- Landon, B. E. 1998. "A Conceptual Model of the Effects of Health Care Organizations on the Quality of Medical Care." *Journal of the American Medical Association* 279 (17): 1377. doi:10.1001/jama.279.17.1377.
- Landon, B. E., N. L. Keating, M. L. Barnett, J.-P. Onnela, S. Paul, A. J. O'Malley, T. Keegan, and N. A. Christakis. 2012. "Variation in Patient-Sharing Networks of Physicians across the United States." *Journal of the American Medical Association* 308 (3): 265–73. doi:10.1001/jama.2012.7615.
- Laupattarakasem, W., M. Laopaiboon, P. Laupattarakasem, and C. Sumananont. 2008. "Arthroscopic Debridement for Knee Osteoarthritis." *Cochrane Database of Systematic Reviews* 1: CD005118. doi:10.1002/14651858.CD005118.pub2.
- Levin, A., G. L. Bakris, M. Molitch, M. Smulders, J. Tian, L. A. Williams, and D. L. Andress. 2007. "Prevalence of Abnormal Serum Vitamin D, PTH, Calcium, and Phosphorus in Patients with Chronic Kidney Disease: Results of the Study to Evaluate Early Kidney Disease." *Kidney International* 71 (1): 31–8. doi:10.1038/sj.ki.5002009.
- Lin, G. A., R. A. Dudley, and R. F. Redberg. 2007. "Cardiologists' Use of Percutaneous Coronary Interventions for Stable Coronary Artery Disease." *Archives of Internal Medicine* 167 (15): 1604–9. doi:10.1001/archinte.167.15.1604.
- Lin, K., R. Lipsitz, T. Miller, and S. Janakiraman. 2008. "Benefits and Harms of Prostate-Specific Antigen Screening for Prostate Cancer: An Evidence Update for the U.S. Preventive Services Task Force." *Annals of Internal Medicine* 149 (3): 192. doi:10.7326/0003-4819-149-3-200808050-00009.
- Martí-Carvajal, A. J., I. Solà, D. Lathyris, D.-E. Karakitsiou, and D. Simancas-Racines. 2013. "Homocysteine-Lowering Interventions for Preventing Cardiovascular Events." *The Cochrane Database of Systematic Reviews* 1: CD006612. doi:10.1002/14651858.CD006612.pub3.
- McWilliams, J. M., M. E. Chernew, J. B. Dalton, and B. E. Landon. 2014a. "Outpatient Care Patterns and Organizational Accountability in Medicare." *Journal of the American Medical Association Internal Medicine* 174 (6): 938–45. doi:10.1001/jamainternmed.2014.1073.
- McWilliams, J. M., B. E. Landon, M. E. Chernew, and A. M. Zaslavsky. 2014b. "Changes in Patients' Experiences in Medicare Accountable Care Organizations." *New England Journal of Medicine* 371 (18): 1715–24. doi:10.1056/NEJMsa1406552.
- McWilliams, J. M., M. E. Chernew, B. E. Landon, and A. L. Schwartz. 2015. "Performance Differences in Year 1 of Pioneer Accountable Care Organizations." *The*

- New England Journal of Medicine* 372 (20): 1927–36. doi:10.1056/NEJMsal414929.
- McWilliams, J. M., L. A. Hatfield, M. E. Chernew, B. E. Landon, and A. L. Schwartz. 2016. “Early Performance of Accountable Care Organizations in Medicare.” *New England Journal of Medicine* 374: 2357–66. doi:10.1056/NEJMsal600142.
- Mohammed, T., J. Kirsch, J. Amorosa, K. Brown, J. Chung, D. Dyer, M. Ginsburg, D. Heitkamp, J. Kanne, E. Kazerooni, L. Ketai, J. Ravenel, A. Saleh, R. Shah, and E. P. on T. Imaging. 2011. *ACR Appropriateness Criteria Routine Admission and Preoperative Chest Radiography*. Reston, VA.
- Moya, A., R. Sutton, F. Ammirati, J.-J. Blanc, M. Brignole, J. B. Dahm, J.-C. Deharo, J. Gajek, K. Gjesdal, A. Krahn, M. Massin, M. Pepi, T. Pezawas, R. Ruiz Granell, F. Sarasin, A. Ungar, J. G. van Dijk, E. P. Walma, and W. Wieling. 2009. “Guidelines for the Diagnosis and Management of Syncope (Version 2009).” *European Heart Journal* 30 (21): 2631–71. doi:10.1093/eurheartj/ehp298.
- O’Malley, A. J., and A. M. Zaslavsky. 2008. “Domain-Level Covariance Analysis for Multilevel Survey Data With Structured Nonresponse.” *Journal of the American Statistical Association* 103 (484): 1405–18. doi:10.1198/016214508000000724.
- Palmer, S. C., D. O. McGregor, J. C. Craig, G. Elder, P. Macaskill, and G. F. Strippoli. 2009. “Vitamin D Compounds for People with Chronic Kidney Disease Not Requiring Dialysis.” *The Cochrane Database of Systematic Reviews* 4: CD008175. doi:10.1002/14651858.CD008175.
- Pham, H. H., D. Schrag, A. S. O’Malley, B. Wu, and P. B. Bach. 2007. “Care Patterns in Medicare and Their Implications for Pay for Performance.” *The New England Journal of Medicine* 356 (11): 1130–9. doi:10.1056/NEJMsal063979.
- Pinto, R. Z., C. G. Maher, M. L. Ferreira, M. Hancock, V. C. Oliveira, A. J. McLachlan, B. Koes, and P. H. Ferreira. 2012. “Epidural Corticosteroid Injections in the Management of Sciatica: A Systematic Review and Meta-Analysis.” *Annals of Internal Medicine* 157 (12): 865–77.
- PREPIC Study Group. 2005. “Eight-Year Follow-up of Patients with Permanent Vena Cava Filters in the Prevention of Pulmonary Embolism: The PREPIC (Prevention Du Risque d’Embolie Pulmonaire Par Interruption Cave) Randomized Study.” *Circulation* 112 (3): 416–22. doi:10.1161/CIRCULATIONAHA.104.512834.
- Qaseem, A., V. Snow, N. Fitterman, E. R. Hornbake, V. A. Lawrence, G. W. Smetana, K. Weiss, D. K. Owens, M. Aronson, P. Barry, D. E. Casey, J. T. Cross, K. D. Sherif, and K. B. Weiss. 2006. “Risk Assessment for and Strategies to Reduce Perioperative Pulmonary Complications for Patients Undergoing Noncardiothoracic Surgery: A Guideline from the American College of Physicians.” *Annals of Internal Medicine* 144 (8): 575–80. doi:10.7326/0003-4819-144-8-200604180-00008.
- Rajaram, S. S., N. K. Desai, A. Kalra, M. Gajera, S. K. Cavanaugh, W. Brampton, D. Young, S. Harvey, and K. Rowan. 2013. “Pulmonary Artery Catheters for Adult Patients in Intensive Care.” *The Cochrane Database of Systematic Reviews* 2: 1–59. doi:10.1002/14651858.CD003408.pub3.

- Rebitzer, J. B., and M. Votruba. 2014. "Organizational Economics and Physician Practices." In *Encyclopedia of Health Economics*, edited by A. J. Culyer, pp. 414–24. Oxford, UK: Elsevier.
- Robinson, J. C. 2001. "Theory and Practice in the Design of Physician Payment Incentives." *The Milbank Quarterly* 79 (2): 149–77. doi:10.1111/1468-0009.00202.
- Sarosiek, S., M. Crowther, and J. M. Sloan. 2013. "Indications, Complications, and Management of Inferior Vena Cava Filters: The Experience in 952 Patients at an Academic Hospital with a Level I Trauma Center." *Journal of the American Medical Association Internal Medicine* 173 (7): 513–7. doi:10.1001/jamainternmed.2013.343.
- Schwartz, A. L., B. E. Landon, A. G. Elshaug, M. E. Chernew, and J. M. McWilliams. 2014. "Measuring Low-Value Care in Medicare." *Journal of the American Medical Association Internal Medicine* 174 (7): 1067–76. doi:10.1001/jamainternmed.2014.1541.
- Schwartz, A. L., M. E. Chernew, B. E. Landon, and J. M. McWilliams. 2015. "Changes in Low-Value Services in Year 1 of the Medicare Pioneer Accountable Care Organization Program." *Journal of the American Medical Association Internal Medicine* 175 (11): 1815–25. doi:10.1001/jamainternmed.2015.4525.
- Song, Z., S. Rose, D. G. Safran, B. E. Landon, M. P. Day, and M. E. Chernew. 2014. "Changes in Health Care Spending and Quality 4 Years into Global Payment." *New England Journal of Medicine* 371 (18): 1704–14. doi:10.1056/NEJMsa1404026.
- Staal, J. B., R. de Bie, H. C. de Vet, J. Hildebrandt, and P. Nelemans. 2008. "Injection Therapy for Subacute and Chronic Low-Back Pain." *The Cochrane Database of Systematic Reviews* 3: CD001824. doi:10.1002/14651858.CD001824.pub3.
- U.S. Census Bureau. 2012. "2009-2011 American Community Survey Summary File" [accessed on July 8, 2014]. Available at <https://www.census.gov/programs-surveys/acs/data/summary-file.2011.html>
- U.S. Preventive Services Task Force. 2014. "Published Recommendations" [accessed on February 16, 2015]. Available at <http://www.uspreventiveservicestaskforce.org/BrowseRec/Index/browse-recommendations>
- Vesco, K., E. Whitlock, M. Eder, J. Lin, B. Burda, C. Senger, R. Holmes, R. Fu, and S. Zuber. 2011. *Screening for Cervical Cancer: A Systematic Evidence Review for the U.S. Preventive Services Task Force*. Rockville, MD: Agency for Healthcare Research and Quality.
- Welch, W. P., A. E. Cuellar, S. C. Stearns, and A. B. Bindman. 2013. "Proportion of Physicians in Large Group Practices Continued to Grow in 2009-11." *Health Affairs* 32 (9): 1659–66. doi:10.1377/hlthaff.2012.1256.
- Wheatley, K., N. Ives, R. Gray, P. A. Kalra, J. G. Moss, C. Baigent, S. Carr, N. Chalmers, D. Eadington, G. Hamilton, G. Lipkin, A. Nicholson, and J. Scoble. 2009. "Revascularization versus Medical Therapy for Renal-Artery Stenosis." *New England Journal of Medicine* 361 (20): 1953–62. doi:10.1056/NEJMoa0905368.
- Whitlock, E. P., J. Lin, E. Liles, T. Beil, and R. Fu. 2008. "Screening for Colorectal Cancer: A Targeted, Updated Systematic Review for the U.S. Preventive Services

- Task Force.” *Annals of Internal Medicine* 149 (9): 638. doi:10.7326/0003-4819-149-9-200811040-00245.
- Wolff, T., J. Guirguis-Blake, T. Miller, M. Gillespie, and R. Harris. 2007. “Screening for Carotid Artery Stenosis: An Update of the Evidence for the U.S. Preventive Services Task Force.” *Annals of Internal Medicine* 147 (12): 860–70. doi:10.7326/0003-4819-147-12-200712180-00006.
- Zaslavsky, A. M. 2001. “Statistical Issues in Reporting Quality Data: Small Samples and Casemix Variation.” *International Journal for Quality in Health Care* 13 (6): 481–8.
- . 2007. “Using Hierarchical Models to Attribute Sources of Variation in Consumer Assessments of Health Care.” *Statistics in Medicine* 26 (8): 1885–900. doi:10.1002/sim.2808.
- Zhang, Y., S. H. Baik, A. M. Fendrick, and K. Baicker. 2012. “Comparing Local and Regional Variation in Health Care Spending.” *New England Journal of Medicine* 367 (18): 1724–31. doi:10.1056/NEJMsa1203980.

SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.
Data S1: Methods Appendix.