



Assessment of Overuse of Medical Tests and Treatments at US Hospitals Using Medicare Claims

Kelsey Chalmers, PhD; Paula Smith, MPH; Judith Garber, MPP; Valerie Gopinath, MS; Shannon Brownlee, MS; Aaron L. Schwartz, MD, PhD; Adam G. Elshaug, MPH, PhD; Vikas Saini, MD

Abstract

IMPORTANCE Overuse of health care services exposes patients to unnecessary risk of harm and costs. Distinguishing patterns of overuse among hospitals requires hospital-level measures across multiple services.

OBJECTIVE To describe characteristics of hospitals associated with overuse of health care services in the US.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cross-sectional analysis used Medicare fee-for-service claims data for beneficiaries older than 65 years from January 1, 2015, to December 31, 2017, with a lookback of 1 year. Inpatient and outpatient services were included, and services offered at specialty and federal hospitals were excluded. Patients were from hospitals with the capacity (based on a claims filter developed for this study) to perform at least 7 of 12 investigated services. Statistical analyses were performed from July 1, 2020, to December 20, 2020.

MAIN OUTCOMES AND MEASURES Outcomes of interest were a composite overuse score ranging from 0 (no overuse of services) to 1 (relatively high overuse of services) and characteristics of hospitals clustered by overuse rates. Twelve published low-value service algorithms were applied to the data to find overuse rates for each hospital, normalized and aggregated to a composite score and then compared across 6 hospital characteristics using multivariable regression. A k-means cluster analysis was used on normalized overuse rates to identify hospital clusters.

RESULTS The primary analysis was performed on 2415 cohort A hospitals (ie, hospitals with capacity for 7 or more services), which included 1 263 592 patients (mean [SD] age, 72.4 [14] years; 678 549 women [53.7%]; 1 017 191 White patients [80.5%]). Head imaging for syncope was the highest-volume low-value service (377 745 patients [29.9%]), followed by coronary artery stenting for stable coronary disease (199 579 [15.8%]). The mean (SD) composite overuse score was 0.40 (0.10) points. Southern hospitals had a higher mean score than midwestern (difference in means: 0.06 [95% CI, 0.05-0.07] points; $P < .001$), northeast (0.08 [95% CI, 0.06-0.09] points; $P < .001$), and western hospitals (0.08 [95% CI, 0.07-0.10] points; $P < .001$). Nonprofit hospitals had a lower adjusted mean score than for-profit hospitals (-0.03 [95% CI, -0.04 to -0.02] points; $P < .001$). Major teaching hospitals had significantly lower adjusted mean overuse scores vs minor teaching hospitals (difference in means, -0.07 [95% CI, -0.08 to -0.06] points; $P < .001$) and nonteaching hospitals (-0.10 [95% CI, -0.12 to -0.09] points; $P < .001$). Of the 4 clusters identified, 1 was characterized by its low counts of overuse in all services except for spinal fusion; the majority of major teaching hospitals were in this cluster (164 of 223 major teaching hospitals [73.5%]).

CONCLUSIONS AND RELEVANCE This cross-sectional study used a novel measurement of hospital-associated overuse; results showed that the highest scores in this Medicare population were

(continued)

Key Points

Question What hospital characteristics are associated with overuse of health care services in the US?

Findings In this cross-sectional study of 1 325 256 services performed at 3351 hospitals, we found that hospitals in the South, for-profit hospitals, and nonteaching hospitals were associated with the highest rates of overuse.

Meaning Variation within specific hospital types and regions may uncover opportunities for targeted interventions to address overuse.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

associated with nonteaching and for-profit hospitals, particularly in the South.

JAMA Network Open. 2021;4(4):e218075.

Corrected on June 4, 2021. doi:10.1001/jamanetworkopen.2021.8075

Introduction

Overuse is defined as the delivery of tests and procedures that provide little or no clinical benefit, are unlikely to have an impact on clinician decisions, increase health care spending without improving health outcomes, or risk patient harm in excess of potential benefits.¹ Estimates suggest that overuse contributes \$75.7 billion to \$101.2 billion to wasted US health care spending annually.²⁻⁴ Studies at the level of physicians, organizations, and hospital referral regions have measured overuse patterns in claims data, including Medicare, Medicaid, and commercially insured populations.⁵⁻¹⁰ These results show considerable variation across physician organizations, including within hospital referral regions and across physicians within the same organization, although the included physician demographic characteristics did not explain a substantial amount of such variation.¹⁰

Although clinicians are responsible for ordering tests and treatments, their practice patterns may be influenced by hospital policies and culture. Hospital-level interventions to reduce overuse exist,¹¹ but to measure and compare their success, a hospital-level measure is required. This study offers such a measure, based on the overuse rates of 12 low-value services, and compares rates across hospital regions, ownership type, safety net status, and teaching status. We also use cluster analysis to investigate patterns of overuse and whether these patterns are associated with particular hospital characteristics.

Methods

Data Sources

This cross-sectional study used a 100% sample from the Centers for Medicare & Medicaid Services' (CMS) Chronic Conditions Data Warehouse of Medicare Fee-For-Service data from the Medicare Provider Analysis and Review table, inpatient, outpatient, and carrier claims filed at short-term general or critical access hospitals from January 1, 2015, to December 31, 2017. We excluded Medicare Advantage claims and Kaiser Permanente hospitals dominated by patients with Medicare Advantage, specialty hospitals (hospitals with more than 20% of their inpatient admissions as either orthopedic or cardiac diagnosis-related groups), hospitals not on the 2019 CMS Hospital Compare website,¹² and federal hospitals. This study was approved and granted a patient waiver of consent by the New England institutional review board because there were minimal risks for participants and the authors had no contact with any individuals in the study. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.²⁸

Overuse Indicators

We selected 13 low-value services from Schwartz et al⁵ and Segal et al¹³ that we agreed were likely to be provided by hospitals. The included services were knee arthroscopy, vertebroplasty, inferior vena cava filter, renal artery stenting, pulmonary artery catheterization in the intensive care unit, hysterectomy, carotid endarterectomy, coronary artery stenting, spinal fusion, electroencephalogram for 2 low-value indications (syncope and headaches), carotid artery imaging, and head imaging.

Our unit of observation was a unique service date per beneficiary. We modified 6 of these overuse indicators after quality checks on the results indicated some potential misclassification of appropriate services as low value. To enhance the specificity, we added additional exclusion criteria

not in the original published reports. The details of these updated algorithms are listed in eTable 2 in the [Supplement](#).

Within the Medicare claims data, we converted the *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)* procedure and diagnosis codes (present in the data after October 2015) to *International Classification of Diseases, Ninth Revision (ICD-9)* using CMS' general equivalence mapping tables¹⁴ in order to apply these algorithms, which used ICD-9 codes.

We decided to exclude pulmonary artery catheterization because of its low volume (290 total services in 2015-2017). Our composite overuse score therefore included 12 services.

To avoid labeling hospitals as having no overuse because they could not offer a service (eg, if they lacked the necessary equipment), we created a capacity filter for each service. This filter included hospitals with at least 1 claim per year for services similar to, or using similar facilities as, the low-value service in question (eTable 1 in the [Supplement](#)).

There were 3359 hospitals that had capacity to provide at least 1 service. Our primary study population included hospitals with the capacity for 7 or more services (n = 2415, cohort A). We assessed the stability of these findings with a subanalysis on a second cohort of hospitals with the capacity for all services (n = 1350 hospitals, cohort B).

The Composite Overuse Score

Overuse Rates

Calculating a composite score was done in 4 steps: (1) calculating overuse rates for each service, (2) reliability adjustment of these rates for denominator volume, (3) normalizing the range of rates across services, and (4) calculating the weighted sum of these values for each hospital.

Developing an overuse metric from multiple indicators that use different denominators and patient populations presented a challenge. Chalmers et al¹⁵ described 3 types of denominators for quantifying low-value care: the specified service volume, the volume of patients with a specific condition, or the volume of all patients. We used the total patient volume as the denominator for those services that are low value in most cases (vertebroplasty, knee arthroscopy, renal stenting, and inferior vena cava filter). For the remaining services, where there was some benefit in certain circumstances, we used a service-specific (for the procedures) or diagnosis-specific (for tests and imaging) denominator.

We used an empirical Bayes reliability adjustment on these overuse rates to adjust small-denominator hospitals toward the overall mean.¹⁶ This adjustment assumes there is a prior distribution of hospital overuse counts and that hospital estimates with small denominators are less reliable than those with larger volumes. For each service, we fit a β distribution to all hospital overuse rates not equal to 0 or 1 in order to obtain a prior distribution of the overuse rates; this was done in R using the `fitdistrplus` package (R Foundation).¹⁷ The histograms of all rates and these fitted distributions are shown in eFigure 1 in the [Supplement](#). Using the estimated parameters for each service, α and β , the adjusted rate for hospital i was as follows:

$$R_{adj\ i} = (s_i + \alpha) / (d_i + \alpha + \beta),$$

where s_i and d_i are the numerator and denominator count for the hospital's service overuse rate.

We then standardized the adjusted overuse rates from 0 to 1 using minimum-maximum normalization, as the overuse rates varied widely across the services owing to differences in denominator volumes. In order to remove the effect of a small number of hospitals with outlier rates on this rescaling, we first limited the rates to 3 times the SD away from the mean hospital rate for each service by replacing any rates greater or lower than this with the upper and lower bound.

Overuse Score Calculation

The overuse score was a sum of the normalized adjusted overuse rates weighted by the total counts of low-value services across all hospitals. This calculation prioritized services with the highest effect

(by volume) on patients nationally. For cohort A, we redistributed the weights of any missing (that is, no capacity) services in our composite score calculation.

Cluster Analysis

To investigate patterns of overuse across the 12 services, we used k-means cluster analysis to group hospitals based on their normalized adjusted overuse rates using scikit-learn software for the Python programming language.¹⁸ We selected the number of clusters visually using a scree plot and then assigned labels to each cluster based on the apparent patterns across services.

Hospital Characteristics

We defined the following hospital characteristics for our comparative analysis: safety net, teaching and financial status, size, geographic region, and core-based statistical area. We ranked hospitals by their proportion of patient stays billed as dual eligible and designated the highest 20% as safety net hospitals. We derived the geographic region from the 2010 Census Regions and Divisions of the United States report. The remaining characteristics were defined using the American Hospital Association 2017 data set.¹⁹ Hospital size was based on bed counts. Designation as a major teaching hospital required membership in the Council of Teaching Hospitals or the Association of American Medical Colleges. Minor teaching hospitals needed only a medical school affiliation as reported to the American Medical Association. For the core-based statistical area, metropolitan areas have 50 000 or more people, micropolitan regions have 10 000 to 50 000 people, and all other areas are considered rural.²⁰ Hospitals designated government or nonfederal and nongovernment or not-for-profit were labeled as nonprofit; the remaining category of investor-owned (for-profit) was considered for-profit hospitals. We excluded 8 hospitals with missing American Hospital Association data.

Statistical Analysis

We used multiple linear regression to report the adjusted composite overuse means for each hospital characteristic level, adjusted for the other hospital characteristics.²¹ We made post-hoc pairwise comparisons of hospital characteristics with Tukey *P* value and CI adjustment. A *P* value of 0.05 was used to indicate significance, and all tests were 2-sided. For the cluster comparison, we compared the proportions of each hospital characteristic within each cluster against its proportion in the entire cohort of hospitals. Because this difference in proportions is largely affected by sample size, we also calculated the Cohen *h* value and reported results where *h* was greater than 0.2.²²

Claims analysis was performed using SAS Enterprise, version 7.15 HF8 (SAS Institute) on the CMS Virtual Research Data Center, and statistical analyses were performed from July 1, 2020, to December 20, 2020, using Python programming, version 3.7 and R, version 4.0.0 (using the tidyverse, ggplot2, ggridges, and matplotlib packages; R Foundation).²³⁻²⁷ The hospital normalized rates, characteristics, and clusters output are available for reference.³⁵

Results

Table 1 reports the patient and hospital characteristics in our sample, and **Table 2** reports the observed low-value service counts and the denominator counts for cohorts A and B. There were 1 325 256 low-value services from January 1, 2015, to December 31, 2017, in the entire population (3351 hospitals) with the capacity to perform at least 1 of the 12 services. The primary analysis was performed on 2415 cohort A hospitals (ie, hospitals with capacity for 7 or more services), which included 1 263 592 patients (mean [SD] age, 72.4 [14] years; 678 549 women [53.7%]; 1 017 191 White patients [80.5%]). Head imaging for syncope was the highest-volume low-value service (377 745 [29.9%]), followed by coronary artery stenting for stable coronary disease (199 579 [15.8%]) and carotid artery imaging for syncope (131 236 [10.8%]).

Within visits where syncope was the primary diagnosis and facial/head trauma diagnoses were excluded, 377 745 patients (27.0%) received head imaging (interquartile range [IQR], 22.1%-37.8% across hospitals), the highest proportion among the 4 investigated diagnostic services. The overuse rates and their density across all hospitals are shown in eFigure 1 in the Supplement.

For any visit with a percutaneous coronary stent, 24.8% of visits were for a patient with likely stable coronary disease and no unstable angina or acute myocardial infarction (IQR, 13.8%-27.1% across hospitals). Overall 11.0% of patients with syncope had carotid artery imaging (IQR, 7.1%-15.9%).

Overuse Scores

Overuse scores ranged across hospitals from 0.13 to 0.73 points, with a mean (SD) composite overuse score of 0.40 (0.10) points. The distribution of the overuse scores across hospitals is shown in eFigure 2 in the Supplement. Major teaching hospitals had significantly lower adjusted mean overuse scores vs minor teaching hospitals (difference in means, -0.07 [95% CI, -0.08 to -0.06] points; *P* < .001) and nonteaching hospitals (-0.10 [95% CI, -0.12 to -0.09] points; *P* < .001) (Table 3 shows unadjusted and adjusted results). Nonprofit hospitals had a lower adjusted mean score than for-profit hospitals (-0.03 [95% CI, -0.04 to -0.02] points; *P* < .001). There were significant regional differences; southern hospitals had a higher mean score than midwestern (difference in means: 0.06 [95% CI, 0.05-0.07] points; *P* < .001), northeast (0.08 [95% CI, 0.06-0.09] points; *P* < .001), and western hospitals (0.08 [95% CI, 0.07-0.10] points; *P* < .001). Smaller hospitals (<200 beds) had a

Table 1. Patient and Hospital Characteristics in Our Sample

Characteristic	No. (%) ^a		
	All hospitals (N = 3351)	Cohort A hospitals: capacity for 7 or more services (n = 2415)	Cohort B hospitals: capacity for 12 services (n = 1350)
Total low-value services, No.	1 325 256	1 263 592	1 012 489
Patient age, mean (SD), y	73.4 (14)	72.4 (14)	72.3 (14)
Patient sex			
Women, No. (%)	727 566 (54.9)	678 549 (53.7)	539 657 (53.3)
Men, No. (%)	597 690 (45.1)	585 043 (46.3)	472 832 (46.7)
Hospital safety net status			
Non-safety net	2661 (79)	1998 (83)	1204 (89)
Safety net	690 (21)	417 (17)	146 (11)
Hospital ownership			
For profit	542 (16)	444 (18)	250 (19)
Nonprofit	2809 (84)	1971 (82)	1100 (81)
Teaching class			
Major teaching	224 (6.7)	223 (9.2)	216 (16)
Minor teaching	808 (24)	745 (31)	516 (38)
Nonteaching	2319 (69)	1447 (60)	618 (46)
Hospital size			
≥200 beds	1164 (35)	1154 (48)	1013 (75)
6-199 beds	2187 (65)	1261 (52)	337 (25)
Hospital CBSA type			
Urban	2117 (63)	1853 (77)	1257 (93)
Suburban	601 (18)	421 (17)	89 (6.6)
Rural	633 (19)	141 (5.8)	4 (0.3)
Region			
Midwest	995 (30)	638 (26)	338 (25)
Northeast	489 (15)	432 (18)	226 (17)
South	1281 (38)	917 (38)	539 (40)
West	586 (17)	428 (18)	247 (18)

Abbreviation: CBSA, core-based statistical area.

^a Values are listed as No. (%) unless otherwise specified.

larger adjusted mean than larger hospitals (0.02 [95% CI, 0.01-0.03] points; $P < .001$). **Figure 1** shows the density of these scores by hospital characteristics so readers can visualize these differences across all hospitals.

Hospital Clusters

Overuse rates for each service fell into 4 distinct clusters in cohort A (eFigures 3 and 4 in the Supplement show the selection and visualization of these clusters). **Figure 2** shows the quintile counts of the rates across these clusters. For each cluster, we report the hospital characteristics with a significantly and largely different (that is, if Cohen $h > 0.2$) proportion within the cluster compared with all hospitals in the cohort (eTable 3 in the Supplement).

Cluster 1 had hospitals with generally low overuse except for spinal fusion. Major teaching hospitals tended to be found in this cluster (41.2% in cluster 1 vs 16.0% overall; t statistic, 17.5; $P < .001$; Cohen h value, 0.57), as did nonprofit hospitals (92.9% in cluster 1 vs 81.5% overall; t statistic, 6.9; $P < .001$; Cohen h value, 0.35), and large hospitals (>200 beds) (90.8% in cluster 1 vs 75.0% overall; t statistic, 6.9; $P < .001$; Cohen h value, 0.50). Cluster 2 showed higher overuse rates across most invasive procedures than the other 3 clusters, and had more for-profit hospitals (35.7% in cluster 2 vs 18.5% overall; t statistic, 6.4; $P < .001$; Cohen h value, 0.39) and southern hospitals (61.1% in cluster 2 vs 40.0% overall; t statistic, 7.6; $P < .001$; Cohen h value, 0.43). Cluster 3 hospitals had higher overuse of the 4 diagnostic services compared with other clusters and had a larger share of nonteaching hospitals (59.8% in cluster 3 vs 45.8% overall; t statistic, 4.1; $P < .001$; Cohen h value, 0.28).

Table 2. The 12 Low-Value Services and Denominator Descriptions, as Well as the Total Low-Value Service Counts and Spread Across Hospitals

Measure	Numerator	Denominator	Cohort A hospitals (n = 2415) ^a		Cohort B hospitals (n = 1350)		Comp. weight (%)
			Total No.	Rate (N/D) per 100	Total No.	Rate per 100	
Procedures							
Knee arthroscopy	Arthroscopic debridement/chondroplasty of the knee with diagnosis of osteoarthritis or chondromalacia in the procedure claim	Patient volume	105 459	0.03	71 296	0.03	8.4
Vertebroplasty	Vertebroplasty for osteoporosis fractures	Patient volume	94 200	0.03	80 429	0.03	7.2
IVC filter	IVC filter	Patient volume	40 916	0.01	35 974	0.01	3.1
Renal artery stenting	Renal artery stenting for hypertension	Patient volume	12 239	0.003	11 207	0.004	0.9
Hysterectomy	Hysterectomy for benign disease	All hysterectomies	97 831	65.4	81 601	62.5	7.6
CEA	Carotid endarterectomy for patients without stroke/TIA	All carotid endarterectomies	47 612	52.4	44 556	52.3	3.6
Coronary stents	Coronary artery stenting for stable heart disease	All coronary artery stents	199 579	24.8	186 550	24.8	15.1
Spinal fusion	Spinal fusion without radicular pain, herniated disc	All spinal fusions	72 258	20.5	65 866	20.4	5.5
Diagnostic tests and imaging							
EEG for syncope	EEG for syncope	All claims with primary diagnosis of syncope (with exclusions) ^b	77 084	3.6	60 988	3.6	5.9
EEG for headache	EEG for headache	All claims with primary diagnosis of headache (with exclusions)	7433	0.38	5886	0.38	0.6
Carotid artery imaging for syncope	Carotid artery imaging for syncope	All claims with primary diagnosis of syncope (with exclusions) ^c	131 236	11.0	96 231	10.6	10.8
Head imaging for syncope	Head imaging for syncope	All claims with primary diagnosis of syncope (with exclusions) ^c	377 745	27.0	271 905	25.3	31.4

Abbreviations: CEA, carotid endarterectomy; comp, composite; EEG, electroencephalogram; IVC, inferior vena cava; N/D, numerator/denominator; TIA, transient ischemic attack.

^b Syncope defined using *International Classification of Diseases, Ninth Revision (ICD-9)* codes from Segal: 78.02, 99.21, 33.701.

^c Syncope defined using *ICD-9* codes from Schwartz: 78.02, 99.21.

^a Cohort A hospitals had capacity for 7 or more services, whereas cohort B hospitals had capacity for 12 services (excluding pulmonary artery catheterization).

Hospitals in cluster 4 had higher rates of overuse of hysterectomy than other clusters, but lower overuse scores for vertebroplasty, inferior vena cava filters, renal stenting, and the diagnostic services of electroencephalogram and carotid imaging. This group had a higher share of smaller hospitals (40.4% in cluster 4 vs 25.0% overall; *t* statistic, 5.9; *P* < .001; Cohen *h* value, 0.33).

Results for Cohort B: Hospitals With Capacity for All 12 Services

Cohort B had fewer smaller, safety net and rural hospitals than cohort A. Differences in the mean overuse scores across hospital characteristics were similar to cohort A results (Table 3), except that the difference between the small and large hospitals in the smaller cohort was no longer significant.

We also set the number of clusters as 4 in the k-means analysis for cohort B. Results were similar to the first analysis, including 1 cluster in which hospitals tended to have low overuse scores across all services except for spinal fusion—the majority of major teaching hospitals were in this cluster (164 of 223 major teaching hospitals [73.5%])—and another cluster where hospitals had high overuse scores for imaging services (eFigure 5 in the Supplement). The proportions of hospital characteristics within each cluster are shown in eTable 4 in the Supplement, with similar findings as cohort A.

Discussion

To our knowledge, the method of scoring of low-value services reported here represents the first metric that can be applied at a hospital level, allowing for comparisons across hospitals and examination of hospital characteristics associated with low-value care. Our findings that larger hospitals, major teaching hospitals, and nonprofit hospitals are more likely to avoid overuse may provide guidance for targeted improvement efforts. For example, payers such as CMS might consider structuring financial incentives for reducing overuse around specific hospital factors in our data. Our

Table 3. Unadjusted and Adjusted Means of the Composite Overuse Score Across Hospitals

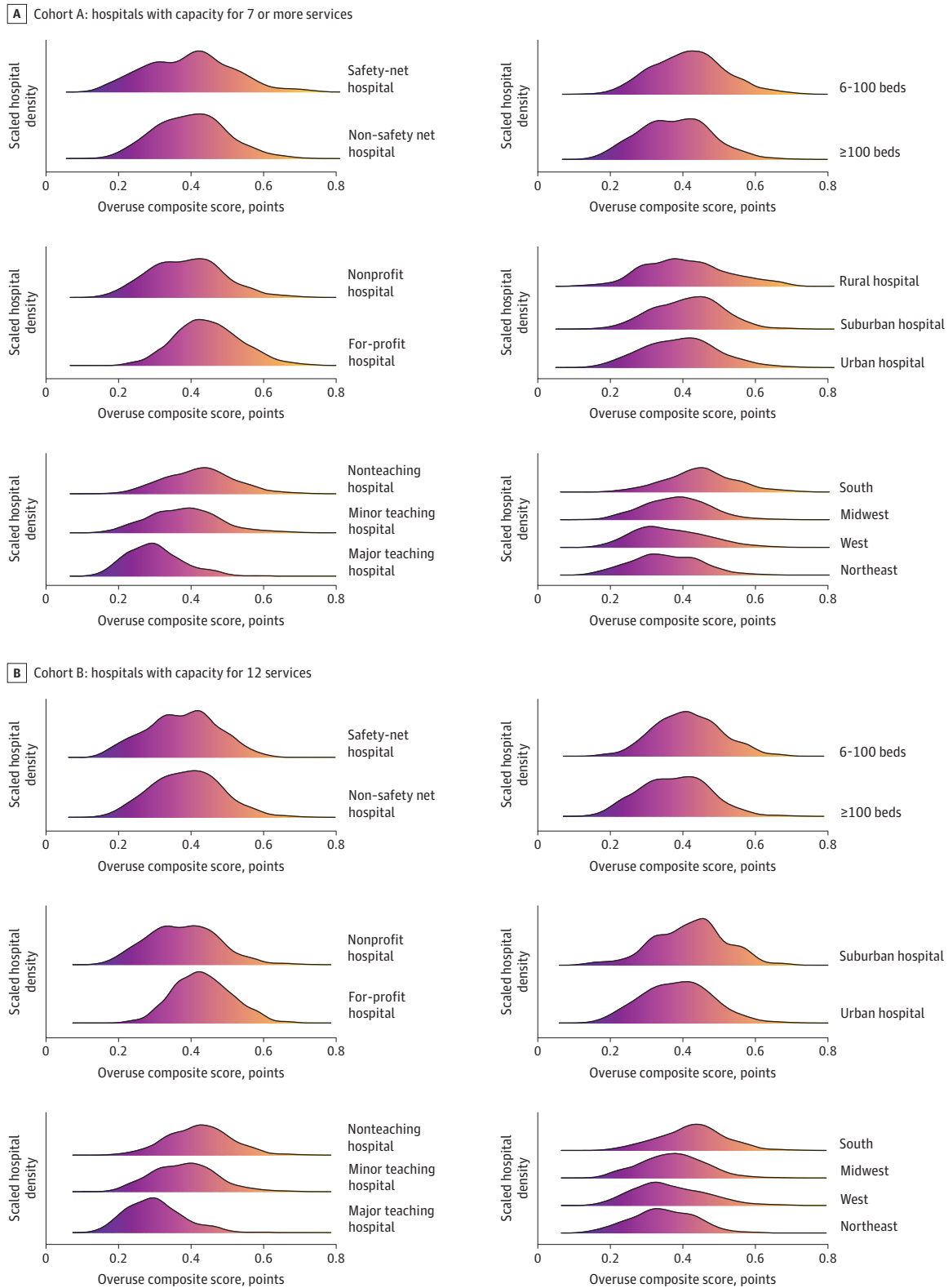
Hospital characteristic	Composite scores for cohort A			Composite scores for cohort B		
	Mean (SD)	Adjusted mean (95% CI) ^a	<i>P</i> value ^b	Mean (SD)	Adjusted mean (95% CI)	<i>P</i> value
Safety net status						
Non-safety net	0.4 (0.1)	0.4 (0.4-0.4)	.32	0.4 (0.1)	0.4 (0.4-0.41)	.06
Safety net	0.4 (0.12)	0.4 (0.39-0.41)		0.39 (0.1)	0.4 (0.39-0.42)	
Ownership type						
For profit	0.45 (0.09)	0.42 (0.41-0.43)	<.001	0.45 (0.08)	0.42 (0.41-0.43)	<.001
Nonprofit	0.39 (0.1)	0.39 (0.39-0.4)		0.39 (0.1)	0.4 (0.39-0.4)	
Teaching class						
Major teaching	0.3 (0.08)	0.32 (0.3-0.33)	<.001	0.31 (0.08)	0.32 (0.31-0.33)	<.001
Minor teaching	0.38 (0.1)	0.39 (0.38-0.39)		0.4 (0.09)	0.4 (0.39-0.41)	
Nonteaching	0.42 (0.1)	0.42 (0.41-0.42)		0.44 (0.09)	0.43 (0.43-0.44)	
Hospital size						
≥200 beds	0.38 (0.1)	0.39 (0.38-0.4)	.02	0.39 (0.1)	0.4 (0.39-0.4)	.23
6-199 beds	0.42 (0.1)	0.41 (0.4-0.41)		0.43 (0.09)	0.41 (0.4-0.42)	
CBSA type						
Urban	0.39 (0.1)	0.4 (0.4-0.41)	.60	0.4 (0.1)	0.4 (0.4-0.41)	.30
Suburban	0.41 (0.09)	0.39 (0.38-0.4)		0.44 (0.1)	0.4 (0.39-0.42)	
Rural	0.41 (0.11)	0.38 (0.37-0.4)		0.41 (0.04)	0.38 (0.29-0.46)	
Region						
Midwest	0.38 (0.09)	0.39 (0.38-0.39)	<.001	0.38 (0.09)	0.39 (0.38-0.4)	<.001
Northeast	0.35 (0.09)	0.37 (0.36-0.37)		0.36 (0.09)	0.38 (0.37-0.39)	
South	0.45 (0.09)	0.44 (0.44-0.45)		0.44 (0.09)	0.43 (0.43-0.44)	
West	0.37 (0.1)	0.36 (0.35-0.37)		0.38 (0.1)	0.37 (0.36-0.38)	

Abbreviation: CBSA, core-based statistical area.

^a Adjusted means are based on linear regression using all hospital characteristics.

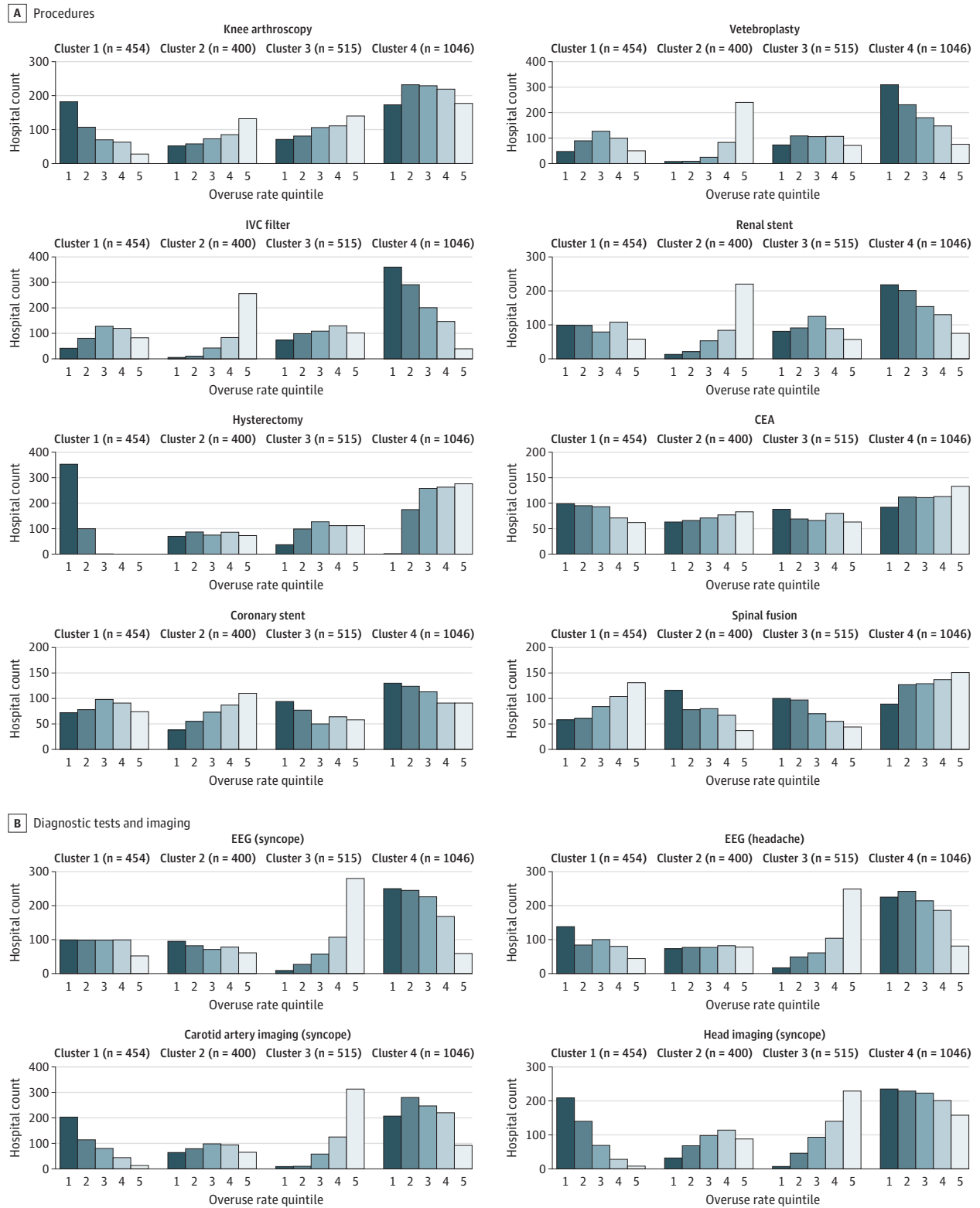
^b Comparison *P* value is from the analysis of variance comparison of the linear regression model and tests whether the hospital characteristic is significant in this model.

Figure 1. Overuse Composite Scores by Hospital Characteristic



A, Density plots of the overuse composite score for hospitals with capacity for 7 or more services (cohort A) in safety and non-safety net hospitals, nonprofit and for-profit hospitals, teaching and nonteaching hospitals, number of beds per hospital, rural, suburban, and urban hospitals, and hospitals based on geographic location. B, Density plots of the overuse composite score for hospitals with capacity for 12 services (cohort B) in safety and non-safety net hospitals, nonprofit and for-profit hospitals, teaching and nonteaching hospitals, number of beds per hospital, rural, suburban, and urban hospitals, and hospitals based on geographic location.

Figure 2. Counts Within Quintiles for 12 Low-Value Services in 4 Identified Hospital Clusters



A, Cluster profiles for hospitals with capacity for 7 or more services (cohort A, N = 2415 hospitals) in reference to the following procedures: knee arthroscopy, vertebroplasty, IVC filter, renal stent, hysterectomy, CEA, coronary stent, and spinal fusion. B, Cluster profiles for hospitals with capacity for 7 or more services in reference to the following diagnostic tests and imaging: electroencephalogram (EEG) (syncope), EEG (headache), carotid artery imaging (syncope), and head imaging (syncope). Bars show the counts of quintiles of the normalized overuse hospital rates for each service across the 4 clusters. CEA indicates carotid endarterectomy; IVC, inferior vena cava.

cluster analyses might also point to ways for payers to target incentives for reducing particular types of overuse; diagnostic testing, for example, is already low in major teaching hospitals but higher in others.

We also found regional differences in hospitals' avoidance of overuse, and CMS could prioritize its efforts by regions. Colla et al⁸ also found their overuse composite measure (at the hospital referral region level) of tests and treatments was highest in the southern US.

We used total numerator volumes to weight the composite overuse score in order to underemphasize services with low volumes, and our conclusions based on the composite score are dependent on this choice. We could have used weights based on the total costs of each of the services, the likely patient harm from each of these services, or how certain the evidence is to avoid a service. Each of these weightings would create an overall score for hospitals based on different judgments about the consequences of delivering low-value services (eg, the value of a low-volume, expensive procedure vs a high-volume, low-cost service).

The cluster analysis revealed underlying patterns of hospital characteristics associated with overuse that were stable within and across the 2 study populations. For example, both cohorts included a cluster where hospitals had high rates of imaging overuse; this could mean many or even all of the hospitals in this cluster share common business practices, culture, or payer mix. This consistency reveals a structure within the data but is hypothesis generating. Further studies will be required to elucidate the factors responsible for these observations.

Within both cohorts A and B, 1 cluster exhibited notably lower overuse scores on all services with the exception of spinal fusion. This cluster had a disproportionate share of larger, metropolitan-area nonprofit teaching hospitals in the northeast. Why this service might be an outlier among these hospitals is unclear. It may be driven by patient demand for spinal fusion, but more likely factors for its entrenchment include the sparsity of high-quality evidence²⁹ and such hospital-level factors³⁰ as investment in devices, local market competition,³¹ and the procedure's relatively high profit margin.³²

Limitations

This study has some limitations. Clinical details are not always captured in claims data, and indicators of low-value care may underestimate or overestimate true rates.³³ We used a set of published indicators, some of which are from another overuse index that has external validation against regional costs and outcomes.¹³ In addition, our improvements to ICD-9-based claims data algorithms for classifying low-value services enhance the specificity of our results.

Our analysis was based on Medicare fee-for-service claims. There may be different trends of overuse among commercially insured persons, perhaps owing to different policies and coverage or provider reimbursements. At a clinician level, however, both Charlesworth et al⁷ and Colla et al³⁴ showed that clinicians did not change their practices regarding provision of low-value services depending on a patient's insurance (Medicare vs commercially insured). They found, instead, that geography was a bigger driver in variation of low-value service utilization.

Our results do not apply to specialty hospitals, which we defined conservatively as those with more than 20% of orthopedic or cardiac cases. These hospitals may have substantially different rates of overuse than general hospitals.

Although the patterns across hospital characteristics in the smaller group of hospitals in cohort B were similar to those in cohort A, they may not persist in the full population of 3359 hospitals with capacity for at least 1 service. Our findings are also limited by the set of specific low-value services we investigated. Other patterns may emerge when more services are included.

Conclusions

Results of this cross-sectional study show that measurements of low-value services using Medicare claims data can be applied to individual hospitals to compare their overall rates of overuse. This

analysis revealed differences in overuse by hospital characteristics such as teaching status, region, and nonprofit status. This novel measurement of hospital-associated overuse is a useful method for combining results across multiple indicators of overuse and comparing overall overuse within US hospitals.

ARTICLE INFORMATION

Accepted for Publication: March 9, 2021.

Published: April 27, 2021. doi:[10.1001/jamanetworkopen.2021.8075](https://doi.org/10.1001/jamanetworkopen.2021.8075)

Correction: This article was corrected on June 4, 2021, to correct the number of White patients, which was written incorrectly in the Abstract and Results.

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2021 Chalmers K et al. *JAMA Network Open*.

Corresponding Author: Vikas Saini, MD, Lown Institute, 21 Longwood Ave, Brookline, MA 02446 (vsaini@lowninstitute.org).

Author Affiliations: Lown Institute, Brookline, Massachusetts (Chalmers, Smith, Garber, Gopinath, Brownlee, Saini); Menzies Centre for Health Policy, Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia (Chalmers); Department of Medical Ethics and Health Policy, Perelman School of Medicine, The University of Pennsylvania, Philadelphia (Schwartz); Division of General Internal Medicine, Perelman School of Medicine, The University of Pennsylvania, Philadelphia (Schwartz); Center for Health Equity Research and Promotion, Corporal Michael J. Crescenz Veterans Administration Medical Center, Philadelphia, Pennsylvania (Schwartz); Centre for Health Policy, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia (Elshaug); University of Southern California, Brookings Schaeffer Initiative for Health Policy, The Brookings Institution, Washington, DC (Elshaug).

Author Contributions: Dr Chalmers and Ms Gopinath had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Chalmers, Smith, Garber, Brownlee, Saini.

Acquisition, analysis, or interpretation of data: Chalmers, Smith, Gopinath, Schwartz, Elshaug, Saini.

Drafting of the manuscript: Chalmers, Smith, Garber, Elshaug, Saini.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Chalmers, Smith, Garber, Gopinath, Schwartz, Elshaug.

Obtained funding: Saini.

Administrative, technical, or material support: Schwartz.

Supervision: Elshaug, Saini.

Conflict of Interest Disclosures: Dr Chalmers report receiving personal fees from Queensland Health Department, Victoria Health Department, and Private Healthcare Australia for previous data analysis consulting, and grants from Australian Department of Veterans' Affairs outside the submitted work; Dr Chalmers reported that the Lown Institute received grant funding from Arnold Ventures on low-value care research, unrelated to the current work, between 2020-2021. Dr Schwartz reported receiving personal fees from the Lown Institute, CVS Health, and Medicare Payment Advisory Commission, and grants from Phyllis & Jerome Lyle Rappaport Foundation outside the submitted work. Dr Elshaug reported receiving personal fees from the Australian state government health departments-Victoria, Queensland, South Australia, as well as the Australian Department of Veterans Affairs, Medibank Ltd, Private Healthcare Australia, and the Australian Defense Force Joint Health Command, for low-value care analytics and advice, grants from Arnold Ventures LLC, and grants from the National Health and Medical Research Council (Australia) outside the submitted work. No other disclosures were reported.

Funding/Support: This research had no external sponsors and was funded by the Lown Institute.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. MacLeod S, Musich S, Hawkins K, Schwebke K. Highlighting a common quality of care delivery problem: overuse of low-value healthcare services. *J Healthc Qual*. 2018;40(4):201-208. doi:[10.1097/JHQ.0000000000000095](https://doi.org/10.1097/JHQ.0000000000000095)

2. Shrank WH, Rogstad TL, Parekh N. Waste in the US health care system: estimated costs and potential for savings. *JAMA*. 2019;322(15):1501-1509. doi:10.1001/jama.2019.13978
3. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA*. 2012;307(14):1513-1516. doi:10.1001/jama.2012.362
4. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156-168. doi:10.1016/S0140-6736(16)32585-5
5. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring low-value care in Medicare. *JAMA Intern Med*. 2014;174(7):1067-1076. doi:10.1001/jamainternmed.2014.1541
6. Reid RO, Rabideau B, Sood N. Low-value health care services in a commercially insured population. *JAMA Intern Med*. 2016;176(10):1567-1571. doi:10.1001/jamainternmed.2016.5031
7. Charlesworth CJ, Meath THA, Schwartz AL, McConnell KJ. Comparison of low-value care in Medicaid vs commercially insured populations. *JAMA Intern Med*. 2016;176(7):998-1004. doi:10.1001/jamainternmed.2016.2086
8. Colla CH, Morden NE, Sequist TD, Schpero WL, Rosenthal MB. Choosing wisely: prevalence and correlates of low-value health care services in the United States. *J Gen Intern Med*. 2015;30(2):221-228. doi:10.1007/s11606-014-3070-z
9. Oakes AH, Sen AP, Segal JB. Understanding geographic variation in systemic overuse among the privately insured. *Med Care*. 2020;58(3):257-264. doi:10.1097/MLR.0000000000001271
10. Schwartz AL, Zaslavsky AM, Landon BE, Chernew ME, McWilliams JM. Low-value service use in provider organizations. *Health Serv Res*. 2018;53(1):87-119. doi:10.1111/1475-6773.12597
11. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions aimed at reducing use of low-value health services: a systematic review. *Med Care Res Rev*. 2017;74(5):507-550. doi:10.1177/1077558716656970
12. Centers for Medicare & Medicaid Services. Hospital value-based purchasing (HVBP) – safety. Accessed December 22, 2020. <https://data.cms.gov/provider-data/dataset/dgmgq-aat3>
13. Segal JB, Nassery N, Chang H-Y, Chang E, Chan K, Bridges JFP. An index for measuring overuse of health care resources with Medicare claims. *Med Care*. 2015;53(3):230-236. doi:10.1097/MLR.0000000000000304
14. Centers for Medicare & Medicaid Services. 2018 ICD-10 CM and GEMs. Accessed January 15, 2021. <https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs>
15. Chalmers K, Pearson S-A, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. *BMJ Qual Saf*. 2017;26(10):855-858. doi:10.1136/bmjqs-2017-006678
16. MacKenzie TA, Grunkemeier GL, Grunwald GK, et al. A primer on using shrinkage to compare in-hospital mortality between centers. *Ann Thorac Surg*. 2015;99(3):757-761. doi:10.1016/j.athoracsur.2014.11.039
17. Delignette-Muller ML, Dutang C. Fitdistrplus: an R package for fitting distributions. *J Stat Softw*. 2015;64(4):1-34. doi:10.18637/jss.v064.i04
18. Pedregosa F, Varoquaux G, Gramfort A, et al Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12(85):2825-2830.
19. American Hospital Association. AHA hospital statistics, 2017 edition. Accessed December 15, 2020. <https://www.aha.org/2016-12-27-aha-hospital-statistics-2017-edition>
20. United States Census Bureau. Core-based statistical areas. Accessed December 22, 2020. <https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html>
21. Lenth RV. Emmeans: estimated marginal means, aka least-squares means. Accessed December 1, 2020. <https://CRAN.R-project.org/package=emmeans>
22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates; 1988.
23. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer; 2016.
24. Wilke CO. Ggrridges: ridgeline plots in "ggplot2". Accessed December 1, 2020. <https://CRAN.R-project.org/package=ggrridges>
25. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. doi:10.1109/MCSE.2007.55
26. R Core Team. The R project for statistical computing. Accessed December 1, 2020. <https://www.R-project.org/>
27. Wickham H, Averick M, Bryan J, et al Welcome to the tidyverse. *J Open Source Softw*. 2019;4(43):1686. doi:10.21105/joss.01686

28. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-577. doi:10.7326/0003-4819-147-8-200710160-00010
29. Machado GC, Ferreira PH, Yoo RI, et al. Surgical options for lumbar spinal stenosis. *Cochrane Database Syst Rev.* 2016;11(11):CD012421. doi:10.1002/14651858.CD012421
30. Krimphove MJ, Cole AP, Friedlander DF, et al. The current landscape of low-value care in men diagnosed with prostate cancer: what is the role of individual hospitals? *Urol Oncol.* 2019;37(9):575.e9-575.e18. doi:10.1016/j.urolonc.2019.04.001
31. Wright JD, Tergas AI, Hou JY, et al. Effect of regional hospital competition and hospital financial status on the use of robotic-assisted surgery. *JAMA Surg.* 2016;151(7):612-620. doi:10.1001/jamasurg.2015.5508
32. Lyons KW, Klare CM, Kunkel ST, et al. A 5-year review of hospital costs and reimbursement in the surgical management of degenerative spondylolisthesis. *Int J Spine Surg.* 2019;13(4):378-385. doi:10.14444/6052
33. de Vries EF, Struijs JN, Heijink R, Hendriks RJP, Baan CA. Are low-value care measures up to the task? a systematic review of the literature. *BMC Health Serv Res.* 2016;16(1):405. doi:10.1186/s12913-016-1656-3
34. Colla CH, Morden NE, Sequist TD, Mainor AJ, Li Z, Rosenthal MB. Payer type and low-value care: comparing choosing wisely services across commercial and Medicare populations. *Health Serv Res.* 2018;53(2):730-746. doi:10.1111/1475-6773.12665
35. Harvard Dataverse. Replication data for: "assessment of overuse of medical tests and treatments at US hospitals using Medicare claims." Accessed March 1, 2021. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/T22QNO>

SUPPLEMENT.

- eTable 1. Details of the Service Capacity Filter Rules for Each Service
- eTable 2. Details of the Low-Value Service Algorithms and Any Modifications
- eFigure 1. Distribution of the Overuse Rates for Hospitals (Cohort A, N = 2,415)
- eFigure 2. Density Plots of the Normalized Overuse Rates and Overuse Score
- eFigure 3. Scree Plot for K-Means Cluster Analysis of ($n \geq 7$) Service Overuse Rates
- eFigure 4. Cluster Visualization of Results From K-Means Analysis
- eTable 3. Hospital Characteristics Significantly Associated With K-Means Clusters (Cohort A)
- eFigure 5. Cluster Comparison for Hospitals With Capacity for All Services (N = 1,350 Hospitals)
- eTable 4. Hospital Characteristics Significantly Associated With K-Means Clusters (Cohort B)