NIH Public Access
Author Manuscript
JAMA Intern Med. Author manuscript; available in PMC 2015 January 01.

Published in final edited form as:
JAMA Intern Med. 2014 July ; 174(7): 1067–1076. doi:10.1001/jamainternmed.2014.1541.

MEASURING LOW-VALUE CARE IN MEDICARE

Aaron L. Schwartz, BA, **Bruce E. Landon, MD, MBA**, **Adam G. Elshaug, PhD, MPH**, **Michael E. Chernew, PhD**, and **J. Michael McWilliams, MD, PhD**
Department of Health Care Policy, Harvard Medical School (A.L.S., B.E.L., A.G.E., M.E.C., J.M.M.), Boston, MA; Division of General Internal Medicine and Primary Care, Department of Medicine, Beth Israel Deaconess Medical Center (B.E.L), Boston, MA; The Commonwealth Fund (A.G.E.), New York, NY; Menzies Centre for Health Policy, The University of Sydney (A.G.E.), Sydney, NSW, Australia; and Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School (J.M.M.), Boston, MA

Abstract

Importance—Despite the importance of identifying and reducing wasteful health care utilization, few direct measures of overuse have been developed. Direct measures are appealing because they identify specific services to limit and can characterize low-value care even among the most efficient providers.

Objective—To develop claims-based measures of low-value services, examine service use (and associated spending) detected by these measures in Medicare, and determine if patterns of use are related across different types of low-value services.

Design, Setting and Participants—Drawing from evidence-based lists of services that provide minimal clinical benefit, we developed and trialed 26 claims-based measures of low-value services. Using 2009 claims for 1,360,908 Medicare beneficiaries, we assessed the proportion of beneficiaries receiving these services, mean per-beneficiary service use, and the proportion of total spending devoted to these services. We compared the amount of use and spending detected by versions of these measures with different sensitivity and specificity. We also estimated correlations between use of different services within geographic areas, adjusting for beneficiaries' sociodemographic and clinical characteristics.

Main Outcome Measures—Use and spending detected by 26 measures of low-value services in 6 categories: low-value cancer screening; low-value diagnostic and preventive testing; low-value preoperative testing; low-value imaging; low-value cardiovascular testing and procedures; and other low-value surgical procedures.

Results—Services detected by more sensitive versions of measures affected 41% of beneficiaries and constituted 2.7% of overall annual spending. Services detected by more specific versions of measures affected 24% of beneficiaries and constituted 0.6% of overall spending. In adjusted

Address correspondence to: J. Michael McWilliams, M.D., Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115; phone: 617-432-3290; fax: 617-432-0173; mcwilliams@hcp.med.harvard.edu.

Author contributions: see forthcoming authorship forms for details. Mr. Schwartz and Dr. McWilliams had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analyses, low-value spending detected in geographic regions at the 5th percentile of the regional distribution of low-value spending ($221/beneficiary) exceeded the difference in detected low-value spending between regions at the 5th and 95th percentiles ($186/beneficiary). Adjusted regional use was positively correlated among 5 of 6 categories of low-value services (*r* for pair-wise, between-category correlations ranged 0.14–0.56, mean 0.35; P≤0.01).

Conclusions and Relevance—Services detected by a limited number of measures of low-value care constituted modest proportions of overall spending but affected substantial proportions of beneficiaries and may be reflective of overuse more broadly. Performance of claims-based measures in supporting targeted payment or coverage policies to reduce overuse may depend heavily on measure definition.

Keywords

Health Expenditures; Medicare; Physician's Practice Patterns; Quality Indicators; Value-Based Purchasing

---

Several recent initiatives, including the "Choosing Wisely" campaign by the American Board of Internal Medicine Foundation,[1] have focused on directly defining wasteful health care services that provide little or no health benefit to patients. It is challenging, however, to translate evidence-based lists of low-value services generated by such initiatives into meaningful metrics that can be applied to available data sources such as insurance claims.[2] The value of most services depends on the clinical situation in which they are provided, and administrative data often lack the clinical detail necessary to distinguish appropriate from inappropriate use. Consequently, the number of low-value services that can be reliably identified in claims data may be limited, and the amount of low-value care detected by claims-based measures may be highly sensitive to how the measures are defined.

Direct approaches to measuring overuse may nevertheless be useful for characterizing the potential extent of wasteful care and informing policies to address low-value practices. Indirect approaches to measuring care efficiency, such as comparing total risk-adjusted spending per patient across geographic areas or provider organizations,[3] may be challenging for policymakers and providers to act upon because specific services contributing to wasteful spending are not identified.[4] Furthermore, such relative measures may fail to characterize the full extent of low-value practices if they are widespread. In contrast, direct measures could be used to identify specific instances of overuse and assess their frequency among even the most efficient providers. In addition, even a limited set of direct measures could be useful for monitoring low-value care if it reflects underlying drivers of overuse more broadly. For analogous reasons, many quality measures relating to underuse have been developed and applied widely in quality-improvement initiatives despite similar measurement challenges.[5,6]

Drawing from evidence-based lists and the medical literature, we created algorithms to measure selected low-value services that could be applied to insurance claims data with reasonable accuracy despite the limited clinical information in claims. Using 2009 Medicare claims, we examined the use of these services and their associated spending, varying the sensitivity and specificity with which the measures likely identified overuse. We also

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

examined whether use of different types of low-value care was correlated within regions; positive correlations might suggest that the measures reflect common drivers of overuse.

Methods

Data Sources and Sample Population

We analyzed 2008–2009 claims data for a random 5% sample of Medicare beneficiaries, as well as demographic information from enrollment files and chronic conditions from the Chronic Condition Data Warehouse (CCW).[7] We applied measures of low-value services to 2009 claims, using 2008 claims and the CCW for relevant clinical history. Our study population consisted of 1,360,908 beneficiaries who were continuously enrolled in Part A and B of traditional fee-for-service Medicare in 2008 and while alive in 2009. We further restricted the study population to individuals who, in 2009, were living in US states or Washington DC and were age 65 or older. Our study was approved by the Harvard Medical School Committee on Human Studies and the Privacy Board of the Centers for Medicare and Medicaid Services.

Measures of Low-Value Services

We considered services that have been characterized as low-value by the American Board of Internal Medicine Foundation's Choosing Wisely initiative,[8] the US Preventive Services Task Force "D" recommendations,[9] the National Institute for Health and Care Excellence "do not do" recommendations,[10] the Canadian Agency for Drugs and Technologies in Health health technology assessments,[11] or peer-reviewed medical literature.[12] These services have been found to provide little to no clinical benefit on average, either in general or in specific clinical scenarios. From these services, we selected a subset that is relevant to the Medicare population and could be detected using Medicare claims with reasonable specificity, meaning that major clinical factors distinguishing likely overuse from appropriate use could be identified or approximated with claims and enrollment data (eAppendix). We also required the evidence base characterizing each service as low-value to have been established prior to 2009. Many low-value services were not selected (e.g., imaging for pulmonary embolism without moderate or high pre-test probability[8]) because of difficulty distinguishing inappropriate from appropriate use with claims data.

For each selected service, we developed an operational definition of low-value occurrences using Current Procedural Terminology (CPT) procedure codes, Berenson-Eggers Type of Service (BETOS) codes, International Classification of Diseases (ICD-9) diagnostic codes, CCW condition indicators, timing of care, site of care, and demographic information (eTable1). When supported by clinical evidence or guidelines, we broadened the scope of some recommendations featured in lists of low-value services. For example, we expanded the Choosing Wisely definition of low-value preoperative pulmonary testing before cardiac surgery to include pre-operative pulmonary testing before low to intermediate-risk surgeries more broadly.[13] We also combined similar low-value services (e.g. various laboratory tests for hypercoaguable states) into single measures. Table 1 presents the operational definitions for the 26 measures of low-value care we developed and applied to claims.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Inherent in most of our claims-based measures of low-value care was a trade-off between sensitivity (greater capture of inappropriate use) and specificity (less misclassification of appropriate use as inappropriate). To assess the variability of our findings across a spectrum of these important measurement properties, we specified two versions of each measure, one with higher sensitivity (and lower specificity) and the other with higher specificity (and lower sensitivity) for detecting low-value care (Table 1). Even without a gold standard for assessing service appropriateness, the relative sensitivity and specificity of our measures can be inferred from the clinical criteria we applied. For example, limiting the colorectal cancer screening measure to beneficiaries over age 85 instead of 75 decreases its sensitivity (fewer low-value instances detected) but increases its specificity (smaller proportion of appropriate services misclassified as inappropriate).

We calculated spending on low-value services using standardized prices to adjust for regional differences in Medicare payments. We used the median spending per service nationally as the standardized price for each service, including payments from Medicare, beneficiary coinsurance amounts, and any payments from other primary payers. We included related services typically bundled with the low-value service in these price estimates (e.g. contrast administration for an imaging study or anesthesia for a procedure). These bundles were defined based on examination of the most frequent CPT codes appearing during the day a low-value service was provided and thus would not include subsequent care prompted by the service (e.g., further imaging for incidental findings on pre-operative chest x-rays). Additional information on service detection and pricing, including the specific codes (CPT, BETOS, etc.) employed, is available in the eAppendix.

Statistical Analyses

We counted the number of times each beneficiary experienced each low-value service and calculated the per-beneficiary spending for each service. From these values, we calculated the percentage of beneficiaries receiving at least 1 low-value service and the aggregate spending for all beneficiaries for each service and in each of 6 service categories: low-value cancer screening; low-value diagnostic and preventive testing; low-value preoperative testing; low-value imaging; low-value cardiovascular testing and procedures; and other low-value surgical procedures. Aggregate spending estimates were multiplied by 20 to approximate spending for the entire Medicare population from 5% samples. We also calculated the proportion of total spending for services covered by Part A and B of Medicare (including coinsurance amounts and payments from other primary payers) devoted to services detected by low-value care measures.

We used hospital referral regions (HRRs) to examine how utilization of different types of low-value services was related among the same groupings of providers. Although we were not interested in geographic areas per se and although practices patterns vary both within and between areas,[4] HRRs nevertheless served as a useful unit of comparison to determine if groups of providers that were more likely to provide one type of low-value service were more likely to provide another. First, we estimated mean per-beneficiary utilization counts in each service category at the HRR level using linear regression models with HRR fixed effects. To control for beneficiaries' sociodemographic and clinical characteristics, we

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

included as covariates age, age squared, sex, race, indicators of 21 CCW diagnoses present before 2009 (derived from claims dating back to 1999), indicators of having multiple comorbid conditions (2 to 7+), the Rural-Urban Continuum Code for beneficiaries' county of residence, and several socioeconomic measures of the elderly population at the zip code tabulation area level (median income, percent below the federal poverty level, and percent with a high school degree). To account for additional dimensions of case mix not captured by the CCW, we included indicators of conditions that qualified patients for potential receipt of several low-value services (e.g., a diagnosis of headache in 2009 qualifying beneficiaries for potentially inappropriate head imaging; see eAppendix for details). For each pair of low-value service categories, we then estimated correlations between regional means in adjusted utilization, weighted by the number of traditional fee-for-service Medicare beneficiaries in each HRR. Correlations were not substantially altered by use of random effects to estimate regional means or by the addition of indicators of qualifying conditions.

Results

Among 1,360,908 beneficiaries in the study sample, 1,050,775 instances of care provision (77 services per 100 beneficiaries) were detected by the more sensitive measures of low-value services, corresponding to 21.0 million instances for the entire traditional Medicare population in 2009. Forty-one percent of beneficiaries received at least 1 service detected by the more sensitive measures. Our more specific but less sensitive measures of low-value care detected 424,207 services (31 per 100 beneficiaries), corresponding to 8.5 million services for the entire Medicare population. Twenty-four percent of beneficiaries received at least 1 of these services.

Spending for services detected by our more sensitive measures of low-value care totaled $8.2 billion for the entire Medicare population, or $303 per beneficiary, while spending for services detected by our more specific measures totaled $1.8 billion, or $66 per beneficiary. These amounts comprised 2.7% and 0.6%, respectively, of total annual spending in 2009 on services covered by Part A and B of Medicare.

Figure 1 presents utilization rates and their associated spending, decomposed by category of low-value care measures. Imaging, cancer screening, and diagnostic and preventive testing measures detected most of the utilization, whereas measures of imaging and cardiovascular testing and procedures detected most of the spending (see eTable 2 for these results in tabular form). Table 2 presents utilization rates and associated spending captured by each of the 26 measures of low-value care. Individual measures with major contributions to spending included both high-price, low-utilization items such as percutaneous coronary intervention for stable coronary disease and low-price, high-utilization items such as screening for asymptomatic carotid artery disease.

Table 3 presents correlations between adjusted levels of regional service use in different categories of low-value care, as detected by our more sensitive measures. Per-beneficiary utilization counts were positively correlated with one another for 5 of the 6 categories. Correlation coefficients ranged from 0.14 to 0.56 across all pair-wise combinations of these 5 categories ($P \leq 0.01$), with a mean of 0.35. Non-cardiovascular surgical procedures were

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

not positively correlated with utilization in other categories of measures. The measures exhibited good internal consistency across all categories (Chronbach's alpha, 0.69).

Adjusted regional spending on services detected by more sensitive measures of low-value care ranged from $221 per-beneficiary in the 5th percentile to $407 per-beneficiary in the 95th percentile of HRRs (median, $297; inter-quartile range, $264 to $336). Thus, low-value spending detected in regions at the 5th percentile of the regional distribution exceeded the difference in detected low-value spending between regions at the 5th and 95th percentiles ($186/beneficiary).

Comment

In this national study of selected low-value services, Medicare beneficiaries commonly received care that was likely to provide minimal or no benefit on average. Even when applying narrower versions of our limited number of measures of overuse, we identified low-value care affecting roughly one quarter of Medicare beneficiaries. These findings are consistent with the notion that wasteful practices are pervasive in the US health care system.

Within regions, different types of low-value utilization generally exhibited significantly positive correlations with one another, ranging from weak to moderate in strength, although one category of low-value utilization (non-cardiovascular surgical procedures) was not positively correlated with the others. These findings suggest that many, but not all, low-value services may be driven by common factors. Therefore, claims-based measures, although limited in number and the amount of wasteful spending they detect, could be useful for monitoring low-value care more broadly, including some care that may be difficult to measure with claims.

While these findings suggest that direct approaches to measuring wasteful care may be tractable and informative, other findings underscore potential challenges in developing and applying direct measures of overuse. In particular, the amount of low-value care we detected varied substantially with the clinical specificity of our measures. Estimates of the proportion of Medicare beneficiaries receiving one or more measured low-value service decreased from 41% to 24%, and the contribution of low-value spending to total spending decreased from 2.7% to 0.6%, when we employed more restrictive definitions that traded off sensitivity for specificity. For example, our more sensitive measure of low-value imaging for low back pain captured more inappropriate use of imaging studies at the expense of including some appropriate use. Our more specific measure was less likely to include appropriate use but probably excluded many low-value studies, as suggested by the 3-fold reduction in the number of studies captured.

Thus, the performance of administrative rules to reduce overuse through coverage policy, cost-sharing, or value-based payment (e.g., pay for performance) may depend heavily on measure definition. Such strategies may be appropriate for select services whose value is invariably low or whose low-value applications can be identified with high reliability. For other services, however, more sensitive measures could result in unintended restriction of appropriate tests and procedures by coverage and payment policies, while more specific

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

measures could substantially limit the impact of these strategies. Provider groups seeking to minimize wasteful spending, for example in response to global budgets, may be able to distinguish appropriate from inappropriate practices at the point of care without having to employ rigid rules derived from incomplete clinical data.

We also found that, although spending on low-value services varied considerably across regions, spending on low-value services was substantial even in regions where it was lowest. For example, low-value spending at the 5th percentile of the regional distribution of low-value spending was greater than the difference in low-value spending between the 5th and 95th percentiles. This finding suggest potential advantages of direct measurement over relative spending comparisons as a basis for detecting overuse because overuse may be substantial even among more efficient providers.

Our study has several limitations. Most notably, we analyzed only 26 measures of low-value services. In selecting these measures, we emphasized the specificity with which overuse could be detected with claims data and created more restrictive versions that limited contributions of potentially valuable service use to low-value spending totals and utilization counts. Despite the limited number of services we examined, their frequency and correlations with one another suggest substantial and widespread wasteful care. Use of a broader set of less specific and more sensitive measures would capture more low-value care. Similarly, broader definitions of wasteful spending that include downstream costs of low-value service use (e.g., repeat imaging for incidental findings) would capture more spending than our measures did. For example, one study estimated that testing costs may account for just 2% of the lifetime costs of PSA screening.[14]

Clinical data from linked medical records might support a more extensive assessment of the properties of claims-based measures. However, we would not expect the incorporation of more detailed data to substantially alter the amount of low-value care captured by many of our measures (e.g. cancer screening above certain ages, inappropriately frequent bone mineral density testing, homocysteine testing for cardiovascular disease, renal artery stenting, and vertebroplasty). Furthermore, by varying the definitions of our measures, we were able to demonstrate potential limitations of claims-based measures without having to use medical record data; any inconsistencies between claims and medical records in the amount of low-value care detected would have similar implications for strategies to address wasteful practices. Moreover, we focused on the potential utility of claims-based measures because medical record review as a means to measure and monitor wasteful care is costly and thus not feasible on a large scale. Nevertheless, validation of claims-based measures against a gold standard of clinical appropriateness will be needed to more precisely define their strengths and weaknesses and assess their utility for different purposes, such as monitoring, profiling, payment policy, or coverage design.

Although our analysis suggests that common drivers of low-value care exist, our study did not identify specific determinants of wasteful care. Factors associated with low-value care also may be associated with high-value care.[15,16] Coupling measures of overuse with measures of underuse may therefore be important when evaluating programs intended to achieve more cost-effective care.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Finally, unmeasured variation in diagnostic coding practices or case mix may have contributed to positive correlations between regional use of different low-value services in our study. These were not likely sources of significant bias, however, because we found a significant positive correlation between categories of low-value services that did not rely on diagnosis codes to define (i.e. age-inappropriate cancer screening and preoperative testing) and because our results were not sensitive to adjustment for additional conditions qualifying beneficiaries for potential receipt of several low-value services.

Many quality measures have been developed to assess underuse but few to assess overuse. Our study illustrates the potential utility and limitations of a direct approach to detecting wasteful care. Despite their imperfections, claims-based measures of low-value care could be useful for tracking overuse and evaluating programs to reduce it. However, many direct claims-based measures of overuse may be insufficiently accurate to support targeted coverage or payment policies that have a meaningful impact on use without resulting in unintended consequences. Broader payment reforms such as global or bundled payment models could allow greater provider discretion in defining and identifying low-value services while incentivizing their elimination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Joseph P. Newhouse and Frank Levy for comments on an earlier draft of this manuscript. Dr. Newhouse and Dr. Levy were not compensated for their contributions.

Funding sources and role of sponsors: Supported by grants from the Beeson Career Development Award Program (National Institute on Aging K08 AG038354 and the American Federation for Aging Research), Doris Duke Charitable Foundation (Clinical Scientist Development Award #2010053), National Institute on Aging (P01 AG032952), Agency for Healthcare Research and Quality Institutional Training Grant (2T32HS000055-20), Harvard University (Christopher G.P. Walker Fellowship), and the Australian National Health and Medical Research Council Sidney Sax Public Health Fellowship (627061). 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.

References

1. Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. JAMA. 2012; 307(17):1801–1802. [PubMed: 22492759]
2. Elshaug AG, McWilliams JM, Landon BE. The value of low-value lists. JAMA. 2013; 309(8):775–776. [PubMed: 23443439]
3. Skinner, JS. Causes and consequences of regional variations in health care. In: Pauly, MV.; McGuire, T.; Barros, PP., editors. Handbook of Health Economics. Amsterdam, North-Holland: Elsevier; 2012. p. 45-94.
4. Newhouse, JP.; Garber, AM.; Graham, RP.; McCoy, MA.; Mancher, M.; Kibria, A. Variation in health care spending: Target decision making, not geography. Washington DC: Institute of Medicine; 2013.
5. Wachter RM. Expected and unanticipated consequences of the quality and information technology revolutions. JAMA. 2006; 295(23):2780–2783. [PubMed: 16788133]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

6. National Committee for Quality Assurance. The state of health care quality 2012: focus on obesity and on Medicare plan improvement. 2012 http://www.ncqa.org/Portals/0/State of Health Care/2012/SOHC Report Web.pdf.
7. Center for Medicare and Medicaid Services. Chronic Conditions Data Warehouse. 2013 http://www.ccwdata.org/.
8. Choosing Wisely. Lists of five things physicians and patients should question. Feb.2013 http://www.choosingwisely.org/doctor-patient-lists/.
9. U.S. Preventive Services Task Force. Recommendations for adults. 2013 http://www.uspreventiveservicestaskforce.org/adultrec.htm.
10. National Institute for Health and Care Excellence. NICE “do not do” recommendations. 2011 http://www.nice.org.uk/usingguidance/donotdorecommendations/index.jsp.
11. Canadian Agency for Drugs and Technologies in Health. Health technology assessments. (http://cadth.ca/en/products/health-technology-assessment).
12. Elshaug AG, Watt AM, Mundy L, Willis CD. Over 150 potentially low-value health care practices: an Australian study. Med J Aust. 2012; 197(10):556–560. [PubMed: 23163685]
13. Qaseem A, Snow V, Fitterman N, et al. Risk assessment for and strategies to reduce perioperative pulmonary complications for patients undergoing noncardiothoracic surgery: a guideline from the American College of Physicians. Ann Intern Med. 2006; 144(8):575–580. [PubMed: 16618955]
14. Shteynshlyuger A, Andriole GL. Cost-effectiveness of prostate specific antigen screening in the United States: extrapolating from the European study of screening for prostate cancer. J Urol. 2011; 185(3):828–832. [PubMed: 21239021]
15. Landrum MB, Meara ER, Chandra A, Guadagnoli E, Keating NL. Is spending more always wasteful? The appropriateness of care and outcomes among colorectal cancer patients. Health Aff. 2008; 27(1):159–168.
16. Leape LL, Park RE, Solomon DH, Chassin MR, Kosecoff J, Brook RH. Does inappropriate use explain small-area variations in the use of health care services? JAMA. 1990; 263(5):669–672. [PubMed: 2404147]
17. Holley JL. Screening, diagnosis, and treatment of cancer in long-term dialysis patients. Clin J Am Soc Nephrol. 2007; 2(3):604–610. [PubMed: 17699470]
18. Vesco, K.; Whitlock, E.; Eder, M., et al. Screening for cervical cancer: a systematic evidence review for the U.S. Preventive Services Task Force. Rockville, MD: Agency for Healthcare Research and Quality; 2011 May.
19. Whitlock EP, Lin J, Liles E, Beil T, Fu R. Screening for colorectal cancer: A targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2008; 149(9): 638–658. [PubMed: 18838718]
20. Lin K, Lipsitz R, Miller T, Janakiraman S. Benefits and harms of prostate-specific antigen screening for prostate cancer: an evidence update for the U.S. Preventive Services Task Force. Ann Intern Med. 2008; 149(3):192. [PubMed: 18678846]
21. Bell KJL, Hayen A, Macaskill P, et al. Value of routine monitoring of bone mineral density after starting bisphosphonate treatment: secondary analysis of trial data. BMJ. 2009; 338:b2266. [PubMed: 19549996]
22. Hillier TA, Stone KL, Bauer DC, et al. Evaluating the value of repeat bone mineral density measurement and prediction of fractures in older women: the study of osteoporotic fractures. Arch Intern Med. 2007; 167(2):155–160. [PubMed: 17242316]
23. Martí-Carvajal AJ, Solà I, Lathyris D, Karakitsiou D-E, Simancas-Racines D. Homocysteine-lowering interventions for preventing cardiovascular events. Cochrane Database of Syst Rev. 2013; 1
24. Baglin T, Gray E, Greaves M, et al. Clinical guidelines for testing for heritable thrombophilia. Br J Haematol. 2010; 149(2):209–220. [PubMed: 20128794]
25. Levin A, Bakris GL, Molitch M, et al. 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 Int. 2007; 71(1):31–38. [PubMed: 17091124]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

26. Palmer SC, McGregor DO, Craig JC, Elder G, Macaskill P, Strippoli GF. Vitamin D compounds for people with chronic kidney disease not requiring dialysis. Cochrane Database of Syst Rev. 2009; 4
27. Mohammed, T.; Kirsch, J.; Amorosa, J., et al. ACR Appropriateness Criteria routine admission and preoperative chest radiography. Reston, VA: 2011.
28. Joo HS, Wong J, Naik VN, Savoldelli GL. The value of screening preoperative chest x-rays: a systematic review. Can J Anaesth. 2005; 52(6):568–574. [PubMed: 15983140]
29. Douglas PS, Garcia MJ, Haines DE, et al. 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, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the American College of Chest Physicians. J Am Coll Cardiol. 2011; 57(9):1126–1166. [PubMed: 21349406]
30. Fleisher LA, Beckman JA, Brown KA, et al. 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. Circulation. 2007; 116(17):e418–e500. [PubMed: 17901357]
31. Cornelius RS, Martin J, Wippold FJ, et al. ACR appropriateness criteria sinonasal disease. J Am Coll Radiol. 2013; 10(4):241–246. [PubMed: 23420025]
32. Moya A, Sutton R, Ammirati F, et al. Guidelines for the diagnosis and management of syncope (version 2009). Eur Heart J. 2009; 30(21):2631–2671. [PubMed: 19713422]
33. Jordan, J.; Wippold, FI.; Cornelius, R., et al. ACR appropriateness criteria headache. Reston, VA: 2009.
34. Gronseth GS, Greenberg MK. The utility of the electroencephalogram in the evaluation of patients presenting with headache: a review of the literature. Neurology. 1995; 45(7):1263–1267. [PubMed: 7617180]
35. Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. Lancet. 2009; 373(9662):463–472. [PubMed: 19200918]
36. Wolff T, Guirguis-Blake J, Miller T, Gillespie M, Harris R. Screening for carotid artery stenosis: an update of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2007; 147(12):860–870. [PubMed: 18087057]
37. Hendel RC, Berman DS, Di Carli MF, et al. 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 College of Radiology, the American Heart Association, the American Society of Echocardiography, the Society of Cardiovascular Computed Tomography, the Society for Cardiovascular Magnetic Resonance, and the Society of Nuclear Medicine. Circulation. 2009; 119(22):e561–e587. [PubMed: 19451357]
38. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med. 2007; 356(15):1503–1516. [PubMed: 17387127]
39. Lin GA, Dudley RA, Redberg RF. Cardiologists' use of percutaneous coronary interventions for stable coronary artery disease. Arch Intern Med. 2007; 167(15):1604–1609. [PubMed: 17698682]
40. Wheatley K, Ives N, Gray R, et al. Revascularization versus medical therapy for renal-artery stenosis. N Engl J Med. 2009; 361(20):1953–1962. [PubMed: 19907042]
41. Cooper CJ, Murphy TP, Cutlip DE, et al. Stenting and medical therapy for atherosclerotic renal-artery stenosis. N Engl J Med. 2014; 370(1):13–22. [PubMed: 24245566]
42. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2011; 42(2):517–584. [PubMed: 21127304]
43. PREPIC Study Group. 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. 2005; 112(3):416–422. [PubMed: 16009794]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

44. Sarosiek S, Crowther M, Sloan JM. Indications, complications, and management of inferior vena cava filters: the experience in 952 patients at an academic hospital with a level I trauma center. JAMA Intern Med. 2013; 173(7):513–517. [PubMed: 23552968]
45. Kallmes DF, Comstock BA, Heagerty PJ, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. N Engl J Med. 2009; 361(6):569–579. [PubMed: 19657122]
46. Buchbinder R, Osborne RH, Ebeling PR, et al. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. N Engl J Med. 2009; 361(6):557–568. [PubMed: 19657121]
47. Boonen S, Van Meirhaeghe J, Bastian L, et al. Balloon kyphoplasty for the treatment of acute vertebral compression fractures: 2-year results from a randomized trial. J Bone Miner Res. 2011; 26(7):1627–1637. [PubMed: 21337428]
48. McCullough BJ, Comstock BA, Deyo RA, Kreuter W, Jarvik JG. Major medical outcomes with spinal augmentation vs conservative therapy. JAMA Intern Med. 2013; 173(16):1514–1521. [PubMed: 23836009]
49. Laupattarakasem W, Laopaiboon M, Laupattarakasem P, Sumananont C. Arthroscopic debridement for knee osteoarthritis. Cochrane Database of Syst Rev. 2008; 1
50. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. N Engl J Med. 2013; 368(18):1675–1684. [PubMed: 23506518]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

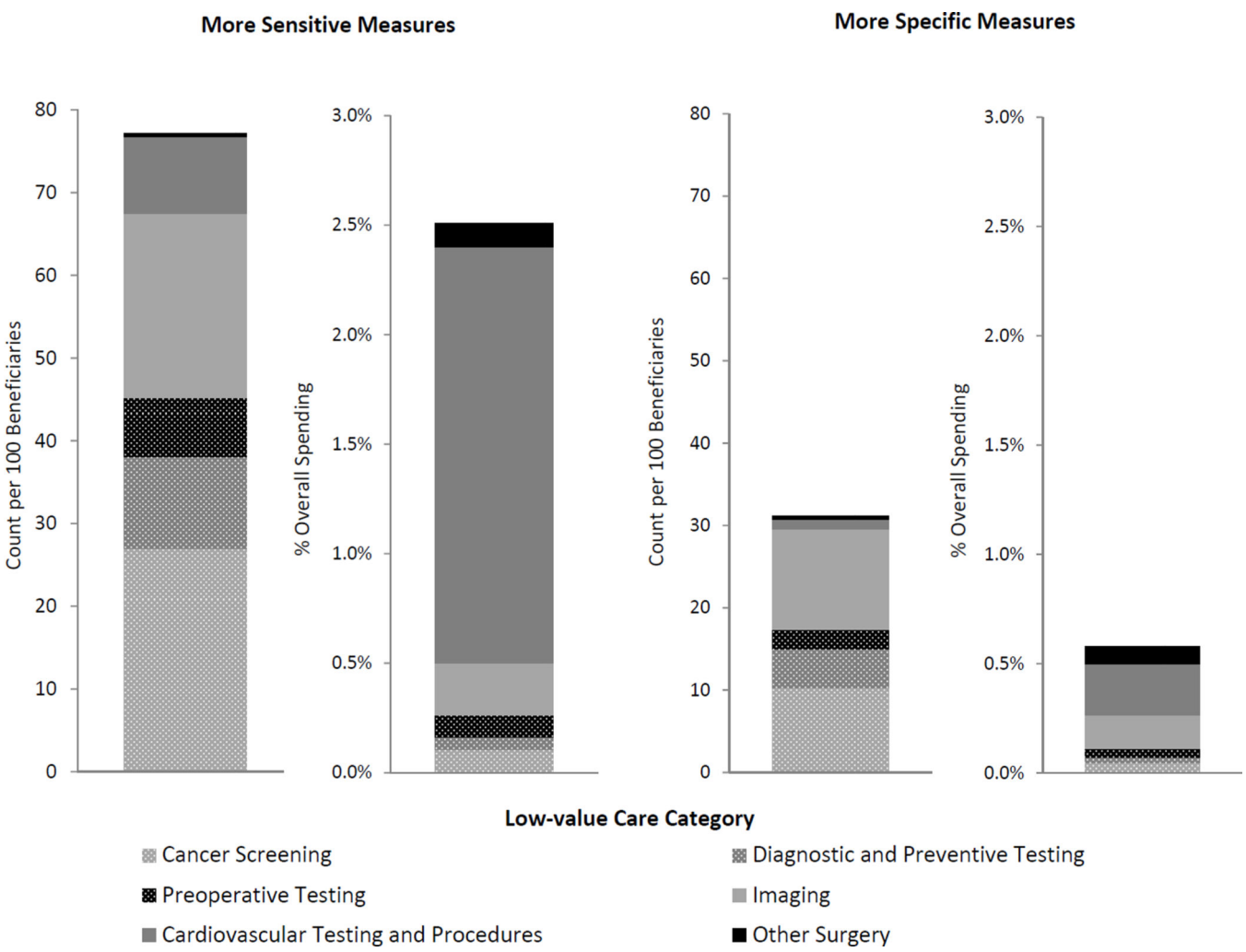

Figure 1. Utilization rates and associated spending for services detected by low-value care measures among Medicare beneficiaries in 2009

Count refers to the number of unique incidences of service provision. Overall spending refers to total spending on all services covered by Part A and B of Medicare. See Table 1 for services included in each category and for operational definitions of all measures.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Measures of low-value services

			Operational Definition	
	Measure	**Source and supporting literature**	**More sensitive, less specific Base definition**	**Less sensitive, more specific Additional restrictions**
Cancer Screening	Cancer screening for patients with chronic kidney disease (CKD) receiving dialysis	CW[17]	Screening for cancer of the breast, cervix, colon, or prostate for patients with chronic kidney disease receiving dialysis services	Only patients over age 75[a]
	Cervical cancer screening for women over age 65	CW, USPSTF[18]	Screening Papanicolaou test for women over age 65	No personal history of cervical cancer or dysplasia noted in claim or in prior claims[b] No diagnoses of other female genital cancers, abnormal Papanicolaou findings, or human papillomavirus positivity in prior claims[b]
	Colorectal cancer screening for older elderly patients	USPSTF[19]	Colorectal cancer screening (colonoscopy, sigmoidoscopy, barium enema, or fecal occult blood testing) for patients over age 75	No history of colon cancer Only screening (i.e. not diagnostic) procedure codes Only patients over age 85
	Prostate-specific antigen (PSA) testing for men over age 75	USPSTF[20]	PSA test for patients over age 75	No history of prostate cancer Only screening (i.e. not diagnostic) procedure codes
Diagnostic and Preventive Testing	Bone mineral density testing at frequent intervals	Literature[21,22]	Bone mineral density test less than two years after a prior bone mineral density test	Only patients with a diagnosis of osteoporosis prior to the initial bone mineral density test[c]
	Homocysteine testing for cardiovascular disease	Literature[23]	Homocysteine testing	No diagnoses of folate or B12 deficiencies in claim and no folate or B12 testing in prior claims[b]
	Hypercoagulability testing for patients with deep vein thrombosis	CW[24]	Lab tests for hypercoagulable states within 30 days following diagnosis of lower extremity deep vein thrombosis or pulmonary embolism	No evidence of recurrent thrombosis, defined by diagnosis of deep vein thrombosis or pulmonary embolism more than 90 days prior to claim
	Parathyroid hormone (PTH) measurement for patients with stage 1–3 CKD	NICE[25,26]	PTH measurement in patients with chronic kidney disease	No dialysis services before PTH testing or within 30 days following testing No hypercalcemia diagnosis in any 2009 claim
Preoperative Testing	Preoperative chest radiography	CADTH CW[27,28]	Chest x-ray specified as a preoperative assessment or occurring within 30 days prior to a low or intermediate risk non-cardiothoracic surgical procedure[e]	No x-rays related to inpatient or emergency care[d] Only x-rays that preceded a low or intermediate risk non-cardiothoracic surgical procedure (i.e. excluding x-rays specified as preoperative before other procedures)[e]
	Preoperative echocardiography	CW[29]	Echocardiogram specified as a preoperative assessment or occurring within 30 days prior to a	No echocardiograms related to inpatient or emergency care[d]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

			Operational Definition	
	Measure	**Source and supporting literature**	**More sensitive, less specific Base definition**	**Less sensitive, more specific Additional restrictions**
			low or intermediate risk non-cardiothoracic surgical procedure[e]	Only echocardiograms that preceded a low or intermediate risk non-cardiothoracic surgical procedure[e]
	Preoperative pulmonary function testing (PFT)	CW[13]	PFT specified as a preoperative assessment or occurring within 30 days prior to a low or intermediate risk surgical procedure[f]	No PFTs related to inpatient or emergency care[d] Only PFTs that preceded a low or intermediate risk surgical procedure[f]
	Preoperative stress testing	CW[30]	Stress electrocardiogram, echocardiogram, or nuclear medicine imaging specified as a preoperative assessment or occurring within 30 days prior to a low or intermediate risk non-cardiothoracic surgical procedure[e]	No stress testing related to inpatient or emergency care[d] Only stress testing that preceded a low or intermediate risk non-cardiothoracic surgical procedure[e]
Imaging	Computed tomography (CT) of the sinuses for uncomplicated acute rhinosinusitis	CW[31]	Maxillofacial CT study with a diagnosis of sinusitis in the imaging claim	No complications of sinusitis,[g] immune deficiencies, nasal polyps, or head/face trauma noted in claim No chronic sinusitis patients, defined by sinusitis diagnosis between 1 year and 30 days prior to imaging
	Head imaging in the evaluation of syncope	CW NICE[32]	CT or Magnetic Resonance Imaging (MRI) of the head with a diagnosis of syncope in the imaging claim	No diagnoses in claim warranting imaging[h]
	Head imaging for uncomplicated headache	CW[33]	Head CT/MRI with diagnosis of (non-thunderclap, non-post-traumatic) headache	No diagnoses in claim warranting imaging[i]
	Electroencephalogram for headaches	CW[34]	EEG with headache diagnosis in the claim	No epilepsy or convulsions noted in current or prior claims[b]
	Back imaging for patients with non-specific low back pain	CW, NICE[35]	Back imaging with a diagnosis of lower back pain	No diagnoses in claim warranting imaging[j] Imaging occurred within six weeks of the first diagnosis of back pain
	Screening for carotid artery disease in asymptomatic adults	CW, USPSTF[36]	Carotid imaging for patients without a history of stroke or transient ischemic attack (TIA) and without a diagnosis of stroke, TIA, or focal neurological symptoms in claim	Test not associated with inpatient or emergency care[k]
	Screening for carotid artery disease for syncope	CW[32]	Carotid imaging with syncope diagnosis	No history of stroke or TIA No stroke, TIA, or focal neurological symptoms noted in claim
Cardiovascular testing and procedures	Stress testing for stable coronary disease	CW[37] Literature[38]	Stress testing for patients with an established diagnosis of ischemic heart disease or angina (at	Test not associated with inpatient or emergency

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

			Operational Definition	
	Measure	**Source and supporting literature**	**More sensitive, less specific Base definition**	**Less sensitive, more specific Additional restrictions**
			least 6 months prior to the stress test) and thus not done for screening purposes	care, which might be indicative of unstable angina[k] Only patients with a past diagnosis of myocardial infarction in order to exclude patients with a history of non-cardiac chest pain in accurately coded as angina (i.e., those with no underlying ischemic heart disease who might benefit from screening and optimization of medical management)
	Percutaneous coronary intervention with balloon angioplasty or stent placement for stable coronary disease	Literature[38,39]	Coronary stent placement or balloon angioplasty for patients with an established diagnosis of ischemic heart disease or angina (at least 6 months prior to the procedure) Procedure not associated with an ER visit,[k] which might be indicative of acute coronary syndrome	Only patients with a past diagnosis of myocardial infarction in order to exclude patients with a history of non-cardiac chest pain inaccurately coded as angina
	Renal artery angioplasty or stenting	Literature[40,41]	Renal/visceral angioplasty or stent placement	Diagnosis of renal atherosclerosis or renovascular hypertension noted in procedure claim
	Carotid endarterectomy in asymptomatic patients	CW[36,42]	Carotid endarterectomy for patients without a history of stroke or TIA and without stroke, TIA, or focal neurological symptoms noted in claim	Operation not associated with an ER visit[k] Only female patients[l]
	Inferior vena cava filters for the prevention of pulmonary embolism	Literature[43,44]	Any IVC filter placement	
Other surgery	Vertebroplasty or kyphoplasty for osteoporotic vertebral fractures	Literature[45–48]	Vertebroplasty/kyphoplasty for vertebral fracture	No bone cancers, myeloma, or hemangioma noted in procedure claim
	Arthroscopic surgery for knee osteoarthritis	NICE[49,50]	Arthroscopic debridement/ chondroplasty of the knee with diagnosis of osteoarthritis or chondromalacia in the procedure claim	No meniscal tear noted in procedure claim

CW = Choosing Wisely; USPSTF = U.S. Preventive Services Task Force C or D recommendations; NICE = National Institute for Health and Care Excellence “do not do” list; CADTH = Canadian Agency for Drugs and Technologies in Health health technology assessments.

[a] This age cutoff is included because the distribution of kidney transplant ages within the sample suggests transplantation is uncommon over age 75.

[b] Prior claims refer to all claims from 01/01/2008 until one day prior to the service of interest.

[c] This restriction limits the measure to testing of patients with osteoporosis.

[d] Inpatient-associated is defined as occurring during within 30 days following an inpatient stay. ER-associated is defined as occurring during or one day after an ER visit.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[e] Procedures include surgeries of the breast, colectomy, cholecystectomy, transurethral resection of the prostate, hysterectomy, orthopedic surgeries besides hip and knee replacement, corneal transplant, cataract removal, retinal detachment, hernia repair, lithotripsy, arthroscopy, and cholecystectomy. 30-day window between preoperative testing and surgery was derived empirically based on distribution of intervals between test and procedure.

[f] Procedures include surgeries listed immediately above as well as coronary artery bypass graft, aneurysm repair, thromboendarterectomy, percutaneous transluminal coronary angioplasty, and pacemaker insertion.

[g] Complications of sinusitis include eyelid inflammation, acute inflammation of orbit, orbital cellulitis, or visual problems.

[h] Exclusion diagnoses include epilepsy, giant cell arteritis, head trauma, convulsions, altered mental status, nervous system symptoms (e.g. hemiplegia), disturbances of skin sensation, speech problems, stroke, transient ischemic attack, history of stroke.

[i] Exclusion diagnoses include those listed in vii as well as cancer and history of cancer.

[j] Exclusion diagnoses include cancer, trauma, intravenous drug abuse, neurological impairment, endocarditis, septicemia, tuberculosis, osteomyelitis, fever, weight loss, loss of appetite, night sweats, and anemia.

[k] Inpatient-associated is defined as occurring during an inpatient stay. ER-associated is defined as occurring during or within 14 days after an ER visit.

[l] Restriction is based on sex-specific subgroup analyses of procedure efficacy in the referenced literature.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

JAMA Intern Med. Author manuscript; available in PMC 2015 January 01.

Table 2

Service counts and associated spending detected by measures of low-value care

	More sensitive version of measures						More specific version of measures					
Measure (abbreviated)	**count (per 100 benefi-ciaries)[a]**	**% of low-value count**	**% of benefi-ciaries affected**	**spending (millions)**	**% of low-value spending**	**% of overall spending[b]**	**count (per 100 benefi-ciaries)[a]**	**% of low-value count**	**% of benefi-ciaries affected**	**spending (millions)**	**% of low-value spending**	**% of overall spending[b]**
Imaging for non-specific low back pain	12.4	16%	9.4%	226	3%	0.07%	4.5	14%	4.1%	82	5%	0.03%
PSA screening over age 75	12.0	16%	8.3%	98	1%	0.03%	2.8	9%	2.7%	23	1%	0.01%
PTH testing in early CKD	7.9	10%	2.5%	137	2%	0.04%	3.1	10%	1.7%	53	3%	0.02%
Stress testing for stable coronary disease	7.8	10%	7.3%	2,065	25%	0.67%	0.8	3%	0.8%	212	12%	0.07%
Colon cancer screening for older elderly patients	7.7	10%	6.9%	573	7%	0.18%	0.9	3%	0.8%	7	0%	0.00%
Cervical cancer screening over age 65	7.0	9%	6.9%	120	1%	0.04%	6.5	21%	6.4%	111	6%	0.04%
Carotid artery disease screening for asymptomatic patients	6.6	9%	6.0%	323	4%	0.10%	5.6	18%	5.1%	274	15%	0.09%
Preoperative x-ray	5.5	7%	5.1%	75	1%	0.02%	1.6	5%	1.6%	22	1%	0.01%
Homocysteine testing for cardiovascular disease	2.0	3%	1.5%	15	0%	0.00%	0.8	3%	0.6%	6	0%	0.00%
Head imaging for syncope	1.5	2%	1.4%	87	1%	0.03%	1.1	3%	1.0%	61	3%	0.02%
Bone mineral density testing at frequent intervals	1.0	1%	1.0%	20	0%	0.01%	0.8	3%	0.8%	17	1%	0.01%
Carotid artery disease screening for syncope	1.0	1%	1.0%	49	1%	0.02%	0.7	2%	0.7%	33	2%	0.01%
PCI/stenting for stable coronary disease	0.8	1%	0.7%	2,810	34%	0.91%	0.1	0%	0.1%	212	12%	0.07%
Preoperative echocardiography	0.8	1%	0.8%	58	1%	0.02%	0.3	1%	0.3%	21	1%	0.01%
Preoperative stress testing	0.7	1%	0.7%	180	2%	0.06%	0.3	1%	0.3%	81	4%	0.03%

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

	More sensitive version of measures						More specific version of measures					
Measure (abbreviated)	**count (per 100 benefi-ciaries)[a]**	**% of low-value count**	**% of benefi-ciaries affected**	**spending (millions)**	**% of low-value spending**	**% of overall spending[b]**	**count (per 100 benefi-ciaries)[a]**	**% of low-value count**	**% of benefi-ciaries affected**	**spending (millions)**	**% of low-value spending**	**% of overall spending[b]**
CT scan for rhinosinusitis	0.6	1%	0.6%	42	1%	0.01%	0.3	1%	0.3%	23	1%	0.01%
Renal artery stenting	0.4	0%	0.3%	705	9%	0.23%	0.1	0%	0.1%	139	8%	0.04%
Vertebroplasty	0.3	0%	0.3%	199	2%	0.06%	0.3	1%	0.3%	196	11%	0.06%
Arthroscopic surgery for knee osteoarthritis	0.2	0%	0.2%	143	2%	0.05%	0.1	0%	0.1%	63	4%	0.02%
Cancer screening for CKD patients on dialysis	0.2	0%	0.2%	4	0%	0.00%	0.1	0%	0.1%	1	0%	0.00%
IVC filter placement	0.2	0%	0.2%	43	1%	0.01%	0.2	1%	0.2%	43	2%	0.01%
Preoperative PFT	0.2	0%	0.2%	2	0%	0.00%	0.1	0%	0.1%	1	0%	0.00%
Head imaging for headache	0.1	0%	0.1%	8	0%	0.00%	0.1	0%	0.1%	5	0%	0.00%
Carotid endarterectomy for asymptomatic patients	0.1	0%	0.1%	263	3%	0.08%	0.1	0%	0.0%	110	6%	0.04%
Hypercoagulability testing after DVT	0.1	0%	0.1%	3	0%	0.00%	0.0	0%	0.0%	1	0%	0.00%
EEG for headache	0.0	0%	0.0%	1	0%	0.00%	0.0	0%	0.0%	0	0%	0.00%
Total	71	100%	41%[c]	8,248	100%	2.7%	31.2	1	24%[c]	1,798	100%	0.6%

CKD = chronic kidney disease;CT = computed tomography; DVT = deep vein thrombosis; EEG = Electroencephalogram; IVC = inferior vena cava; PCI = Percutaneous coronary intervention; PFT = pulmonary function testing; PSA = prostate-specific antigen;PTH = parathyroid hormone.

[a] Count refers to the number of unique incidences of service provision.

[b] Overall spending refers to annualspending for services covered by Part A and B of Medicare. See Table 1 for service category assignments and for operational definitions of all measures.

[c] Total does not equal column sum because some patients received multiple different services.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Pearson's correlation coefficients relating regional use between different categories of measures of low-value care

Category	*Cancer screening*	*Diagnostic and preventive testing*	*Preoperative testing*	*Imaging*	*Cardiovascular testing and procedures*	*Other surgery*
Cancer screening	1	-	-	-	-	-
Diagnostic and preventive testing	0.34**	1	-	-	-	-
Preoperative testing	0.31**	0.14*	1	-	-	-
Imaging	0.56**	0.38**	0.33**	1	-	-
Cardiovascular testing and procedures	0.29**	0.29**	0.27**	0.55**	1	-
Other surgery	−0.13*	−0.07	−0.16**	−0.03	0.05	1

* p-value <0.05

** p-value <0.01