

HEALTH CARE REFORM

Repeat Testing Among Medicare Beneficiaries

H. Gilbert Welch, MD, MPH; Kevin J. Hayes, PhD; Carol Frost, BA

Background: Although the tendency to repeat examinations is a major determinant of the capacity to serve new patients and of the ability to contain health care costs, little research has described the patterns observed in actual practice.

Methods: We investigated patterns of repeat testing in a longitudinal study of a 5% random sample of Medicare beneficiaries, restricted to 743 478 fee-for-service patients who were alive for a 3-year period after their index test between January 1, 2004, and December 31, 2006. Using the 50 largest metropolitan statistical areas as the unit of analysis, we examined the relationship between the proportion of the population tested and the proportion of tests repeated among those tested.

Results: Among beneficiaries undergoing echocardiography, 55% had a second test within 3 years. Repeat testing following other examinations was also common: 44% of imaging stress tests were repeated within 3 years, as were 49% of pulmonary function tests, 46% of chest computed tomography, 41% of cystoscopies, and 35% of upper endoscopies. The proportion of the population tested and the proportion of tests repeated varied across met-

ropolitan statistical areas. The proportion who underwent echocardiography was highest in Miami, Florida (48%, among whom 66% of examinations were repeated in 3 years), and was lowest in Portland, Oregon (18%, among whom 47% of examinations were repeated in 3 years). Across 50 metropolitan statistical areas, the proportion of the population tested was consistently positively correlated with the proportion of tests repeated for echocardiography (Spearman $r=0.87$, $P<.001$), imaging stress test ($r=0.65$, $P<.001$), pulmonary function test ($r=0.62$, $P<.001$), chest computed tomography ($r=0.66$, $P<.001$), cystoscopy ($r=0.21$, $P=.13$), and upper endoscopy ($r=0.59$, $P<.001$).

Conclusions: Repeat testing is common among Medicare beneficiaries. Patients residing in metropolitan statistical areas with high rates of population testing are more likely to be tested and are more likely to have their test repeated.

Arch Intern Med. 2012;172(22):1745-1751.

Published online November 19, 2012.

doi:10.1001/2013.jamainternmed.727

THE TENDENCY TO REPEAT EXAMINATIONS is a major determinant of the capacity to serve new patients and of the ability to contain health care costs. For example, physicians who typically see their patients every year can care for twice as many patients as those who typically see their patients every 6 months. At the same time, these physicians have cut the provider costs in half. Varying thresholds to repeat diagnostic tests and varying intervals to repeat testing (ie, 6 months or 1 year) have similar implications. Tests that are routinely repeated following a brief period require that more capacity (more diagnostic equipment, such as imaging systems, and more personnel) must be in place to be able to provide access for new patients. Low thresholds to repeat diagnostic tests and short intervals to repeat testing also raise costs.

However, little is known about appropriate thresholds and intervals for repeat

testing. The American College of Cardiology Foundation Appropriate Use Criteria Task Force has considered the suitability of various cardiac imaging modalities and has made a few judgments about surveillance intervals (ie, 1 year or 3 years) in specific clinical settings.¹ The absence of relevant research forced the college to rely on the clinical judgment of expert physician panels.² The

See Invited Commentary at end of article

appropriateness criteria of the American College of Radiology³ (also relying on expert panels) focus on imaging for specific clinical conditions, with to date minimal consideration of testing intervals. The evidence base is similarly lacking in the case of diagnostic procedures. For example, the American Urological Association⁴ notes that the most common recommendation on surveillance for bladder cancer is cys-

Author Affiliations: Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Hanover, New Hampshire (Dr Welch); and Medicare Payment Advisory Commission, Washington, DC (Dr Hayes and Ms Frost).

Table 1. Summary of Study Measures

Measure	Denominator	Example
Proportion tested	Population (beneficiaries)	28.5% Of Medicare beneficiaries who were alive from 2004 to 2009 underwent echocardiography between 2004 and 2006
Proportion repeated	Beneficiaries tested	Among those beneficiaries undergoing echocardiography, 55.2% had repeat echocardiography within 3 y
Repeat testing interval	Encounter (the test)	Among those echocardiography procedures that were repeated, the most common interval to the next echocardiography was 1 y

toscopy every 3 months in the first 2 years after initial treatment, followed by every 6 months for the subsequent 2 to 3 years, and then annually thereafter. Seemingly precise, this recommendation dates back to at least 1936 and has an uncertain origin.⁵

The combination of the productivity and cost implications and the scarcity of relevant research led us to investigate the tendency to repeat selected examinations among Medicare beneficiaries. To our knowledge, this investigation represents the first description of repeat testing intervals used in actual practice in the United States.

METHODS

EXAMINATIONS

We selected 6 examinations (1) that are common and familiar to physicians and (2) for which uncertainty exists about whether to repeat them and how often. The 6 included cystoscopy, upper endoscopy, 2 pulmonary examinations (pulmonary function test and chest computed tomography), and 2 cardiology examinations (echocardiography and imaging stress test [nuclear stress and stress echocardiography]). In an effort to provide a frame of reference, we selected 2 examinations that are routinely expected to be repeated at a recommended interval, namely, screening mammography (every 1 year⁶ or 2 years⁷) and eye examination (every year⁸). Examinations were defined using *Current Procedural Terminology* codes (details are available in a technical appendix from the author), and we avoided double counting by removing add-on services (eg, Doppler color flow studies are an add-on service for echocardiography and result in 2 codes in the Medicare Part B data but would count as only one echocardiography session in our data).

PATIENTS

We analyzed data from a 5% random sample of Medicare beneficiaries between January 1, 2004, and December 31, 2009. Beneficiaries enrolled in Medicare Part A only or in a risk contract health maintenance organization were excluded because their claims data are incomplete. Given our interest in repeat testing, we restricted the analysis to those beneficiaries who were "at risk" for repeat testing (ie, those who were alive for the en-

tire 6-year period). Of 1 579 553 Medicare fee-for-service beneficiaries enrolled in 2004, a total of 936 370 were alive through 2009.

For this population, which was continuously enrolled in Medicare fee-for-service from 2004 through 2009, we then reviewed all the claims for the aforementioned examinations performed during those same 6 years. Because we wanted a follow-up window of 3 years, we then discarded all the claims for a service when the beneficiary's earliest examination occurred after December 31, 2006. This produced the cohort in which we studied repeat testing, namely, 743 478 beneficiaries who received one of the aforementioned examinations between January 1, 2004, and December 31, 2006. For each individual, we only counted those claims made within the 3-year period following his or her index test.

DESCRIPTIVE ANALYSES OF REPEAT EXAMINATIONS

For each examination, we first obtained simple descriptive statistics on the proportion of the population tested (between 2004 and 2006) and the proportion of tests repeated among those tested in a specified period (ie, the proportion undergoing repeat testing within 1 or 3 years). We only count the first repeat test of each individual tested in the numerator, such that the proportion repeated can never exceed 100%. We then examined the distribution of repeat testing intervals. Because beneficiaries may have more than 1 repeat test, the denominator is the encounter (the test) and not the individual (the beneficiary). These 3 measures are summarized in **Table 1**.

To depict the distribution of repeat testing intervals, we collapsed the observed intervals (the duration between the 2 examinations, measured in days) into a few discrete categories commonly used in practice (ie, 3 months, 6 months, and annually). For example, understanding the clinical reality that a recommendation for a second examination at 1 year almost never occurs exactly at 1 year, we included a range of intervals around the stated category of plus or minus 33% (ie, 3 months included between 2 and <4 months, 6 months included between 4 and <8 months, and 1 year included between 8 and 16 months). Incorporating values at both extremes yielded 5 categories with the following criteria: less than 3 months (≤ 60 days), 3 months (61-120 days), 6 months (121-240 days), 1 year (241-480 days), and greater than 1 year (>480 days).

RELATIONSHIP TO PROPORTION TESTED

Finally, we explored the relationship between the tendency to test a population (proportion of the population tested) and the tendency to repeat the test (proportion of tests repeated among those tested). The unit of analysis is the metropolitan statistical area (MSA). To enhance the precision of the point estimates for each MSA, this analysis was restricted to the 50 largest MSAs (which comprise >40% of the Medicare population in the present analysis). For each of 8 examinations, our primary analysis was a correlation of (1) the proportion of the MSA population tested during 3 years (2004-2006) with (2) the proportion of those tested in the MSA population who subsequently underwent repeat testing within 3 years of the index test.

Claims-level analyses were performed using commercially available software (SAS, version 9.2; SAS Institute, Inc). Spearman rank correlation coefficients were calculated using another software program (STATA, version 11.2; StataCorp LP). This research was determined to be exempt from review

Table 2. Descriptive Statistics on the Frequency of Examination and Repeat Testing for Various Tests Among Medicare Beneficiaries Between 2004 and 2009

Test	%				
	Proportion of Population Tested Between 2004 and 2006	Proportion Repeated Among Those Tested		Testing Interval Among Those Repeated, mo	
		Within 3 y	Within 1 y	Mean	Median
Need for Repetition Uncertain					
Echocardiography	28.5	55.2	31.1	12.8	12.1
Imaging stress test	23.3	43.6	20.8	15.9	14.1
Pulmonary function test	10.1	49.1	35.0	8.2	5.6
Chest computed tomography	12.0	45.9	34.6	8.4	5.9
Cystoscopy	7.1	40.5	30.8	6.8	3.9
Upper endoscopy	12.0	34.5	19.4	12.3	10.3
Repetition Routinely Expected					
Eye examination	57.7	78.9	59.8	8.0	6.1
Screening mammography	52.1	71.5	49.7	15.2	13.1

by the Dartmouth Committee for the Protection of Human Subjects.

RESULTS

Descriptive statistics on the frequency of various examinations among Medicare fee-for-service beneficiaries across the United States are given in **Table 2**. As expected, patients receiving eye examinations and screening mammography were most likely to undergo repeat testing (79% and 72% of tests were repeated within 3 years, respectively). The median intervals were 6.1 months for repeat eye examinations and 13.1 months for repeat screening mammography. Among 6 examinations selected because they are not routinely expected to be repeated, patients undergoing echocardiography had repeat testing most frequently (55% within 3 years). The median interval was 12.1 months for repeat echocardiography.

Figure 1 shows the distribution of repeat testing intervals after being collapsed into 5 discrete categories commonly used in clinical practice. Of the 2 cardiology examinations, echocardiography was most commonly repeated annually, and imaging stress test was most commonly repeated at intervals longer than 1 year. The 2 pulmonary examinations were typically repeated at shorter intervals: pulmonary function test was most commonly repeated in less than 3 months, and chest computed tomography was most commonly repeated in 6 months. Screening mammography had the most dominant successful testing interval (annually), serving as a test of face validity for the data.

We then considered the tendency to test a population (proportion of the population tested) and the tendency to repeat the test (proportion of tests repeated among those tested). This relationship, shown in **Figure 2**, uses the 50 largest MSAs in the United States as the unit of analysis. In 8 of 8 examinations, the sign on the correlation was positive (sign test $P = .008$). The Spearman rank correlation coefficient was statistically significant for the following 7 examinations: echocardi-

ography ($r = 0.87$), imaging stress test ($r = 0.65$), pulmonary function test ($r = 0.62$), chest computed tomography ($r = 0.66$), upper endoscopy ($r = 0.59$), eye examination ($r = 0.83$), and screening mammography ($r = 0.75$) ($P < .001$ for all). The sole exception was cystoscopy ($r = 0.21$, $P = .13$). The frequency of repeat testing in MSAs at the extremes in the proportion of the population tested (highest and lowest) is summarized in **Table 3**.

Finally, we considered whether MSAs with high repeat testing rates for one examination (eg, echocardiography) tended to have high repeat testing rates for others examinations (eg, upper endoscopy). Among 15 possible pairwise comparisons of the 6 examinations that are not routinely expected to be repeated, a positive correlation was found in 14 (sign test $P < .001$).

COMMENT

We examined repetitive testing for 6 commonly performed diagnostic tests in which repeat testing is not routinely anticipated. Although we expected a certain fraction of examinations to be repeated, we were struck by the magnitude of that fraction: one-third to one-half of these tests are repeated within a 3-year period. This finding raises the question whether some physicians are routinely repeating diagnostic tests.

The concern may be best exemplified in the use of echocardiography. This is a commonly performed examination among the population covered by Medicare, with more than one-quarter of all fee-for-service beneficiaries undergoing the diagnostic test between 2004 and 2006. Although this frequency has been previously reported,⁹ our investigation provides an additional dimension: more than half of those examined underwent repeat testing within 3 years. Combined with the additional finding that the most common repeat testing interval is 1 year, this suggests that some Medicare beneficiaries are undergoing routine annual echocardiography. The practice is confirmed by anecdotal observations of our car-

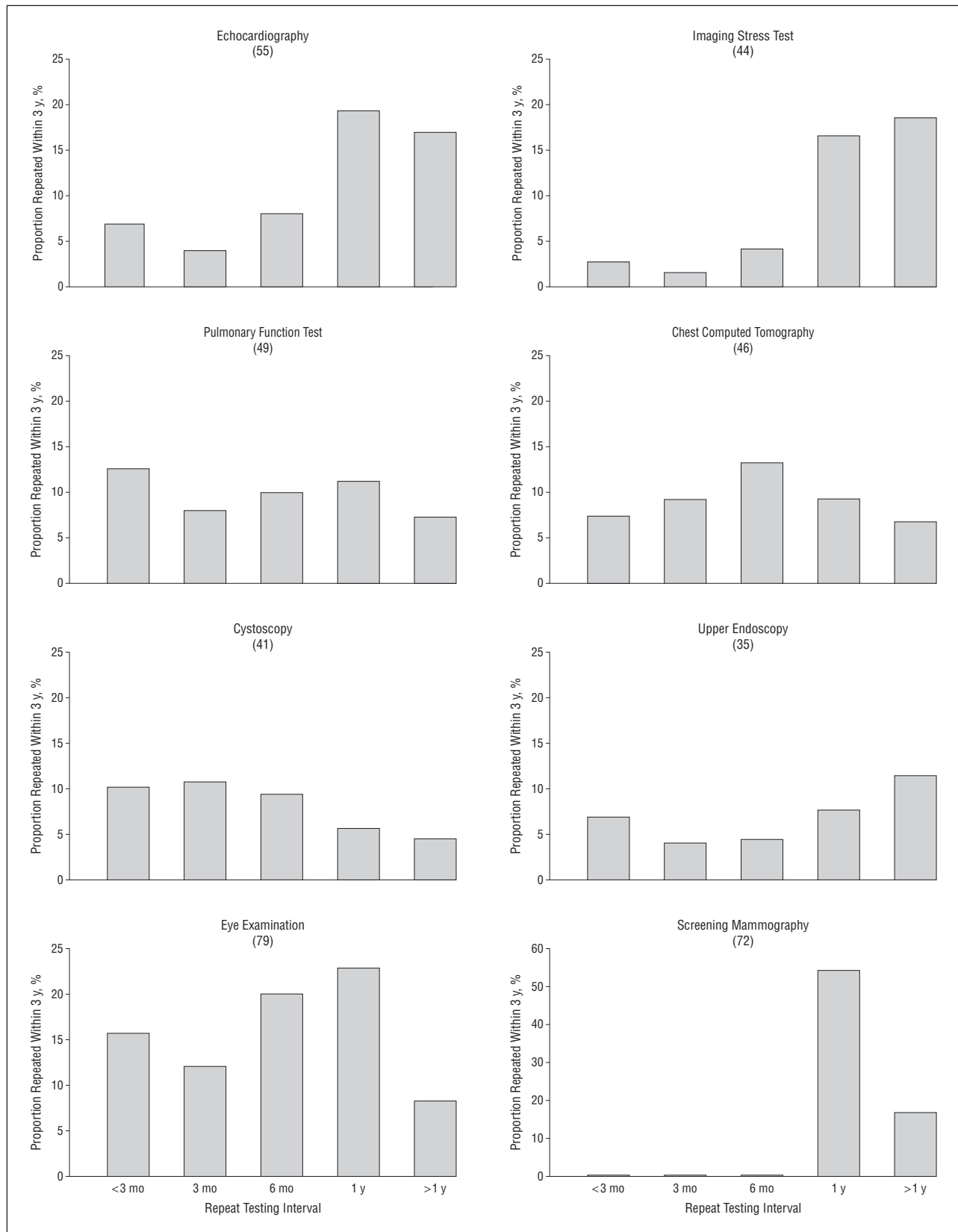


Figure 1. Distribution of repeat testing intervals among Medicare beneficiaries who underwent an examination between 2004 and 2006 and were “at risk” for repeat testing within 3 years. The percentages of patients who underwent repeat testing within 3 years are given in parentheses.

diology colleagues, and websites indicate that such frequent testing is purported to be useful.¹⁰ This is despite

the specific recommendation by the American College of Cardiology Foundation Appropriate Use Criteria Task

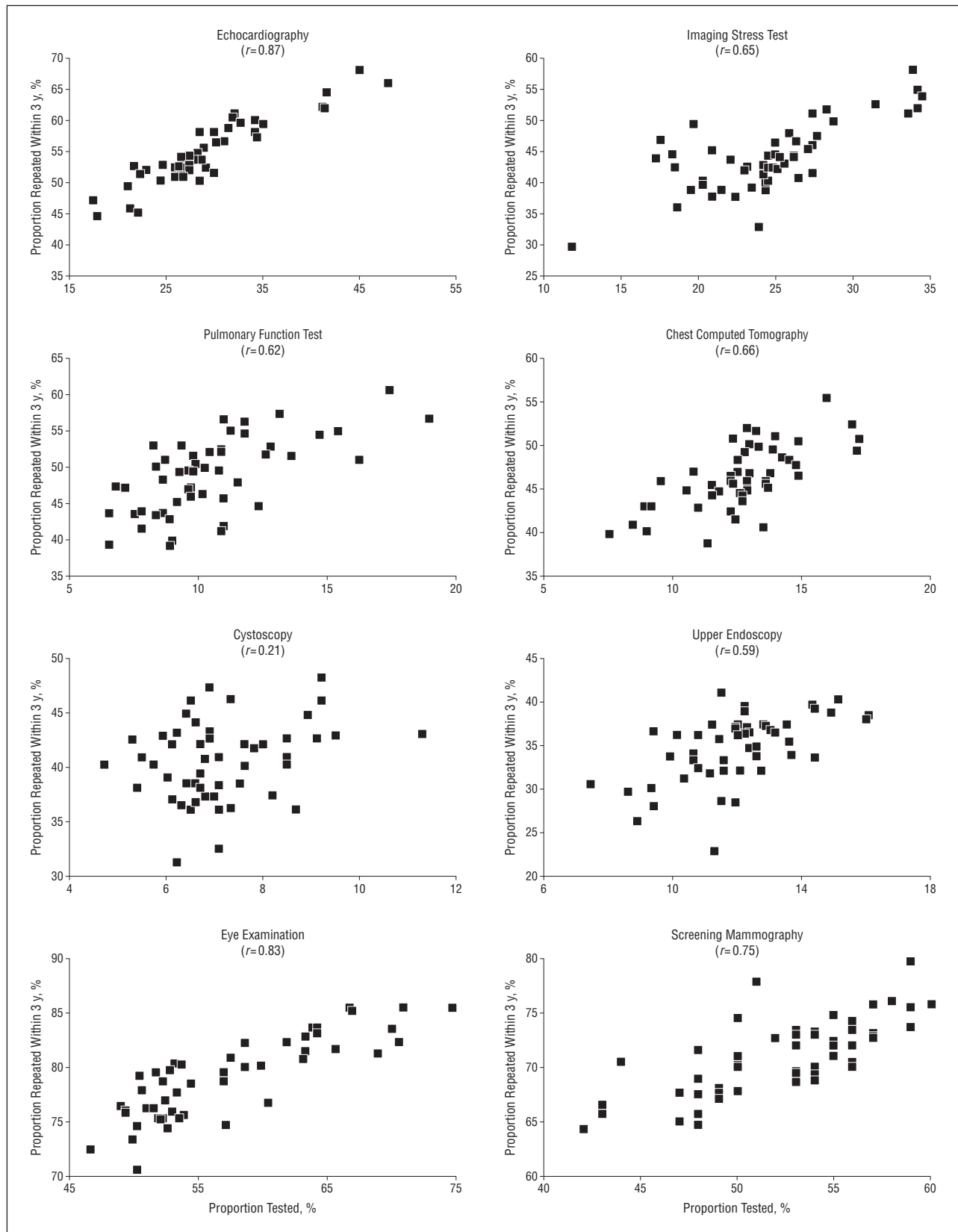


Figure 2. Relationship between the proportion of the population tested between 2004 and 2006 and the proportion of tests repeated within 3 years among those tested. Spearman rank coefficients are given in parentheses.

Force against routine surveillance echocardiography.¹¹ Similarly, our data suggest another practice viewed un-

favorably by the task force, namely, routine annual imaging stress tests.¹²

Table 3. Extremes in the Proportion of the Population Undergoing Various Tests Among Medicare Beneficiaries Residing in the 50 Largest Consolidated Metropolitan Statistical Areas in the United States

Test	Metropolitan Area	%		
		Proportion of Population Tested Between 2004 and 2006	Proportion Repeated Among Those Tested	
			Within 3 y	Within 1 y
Need For Repetition Uncertain				
Echocardiography				
Highest	Miami, Florida	48.0	66.0	39.0
Lowest	Portland, Oregon	17.5	47.1	25.5
Imaging stress test				
Highest	Orlando, Florida	34.4	53.8	24.2
Lowest	Portland, Oregon	11.7	29.8	11.0
Pulmonary function test				
Highest	Detroit, Michigan	19.0	56.6	39.2
Lowest	San Francisco, California	6.5	39.4	28.8
Chest computed tomography				
Highest	Jacksonville, Florida	17.2	50.7	39.5
Lowest	San Francisco, California	7.5	39.9	31.5
Cystoscopy				
Highest	Miami, Florida	11.3	43.1	32.0
Lowest	Portland, Oregon	4.7	40.3	30.9
Upper endoscopy				
Highest	Miami, Florida	16.1	38.5	19.9
Lowest	Sacramento, California	7.4	30.7	17.1
Repetition Routinely Expected				
Eye examination				
Highest	Sarasota, Florida	74.8	85.6	71.0
Lowest	San Bernardino, California	46.6	72.5	51.8
Screening mammography				
Highest	Sarasota, Florida	60.0	75.8	55.9
Lowest	Los Angeles, California	42.0	64.4	43.6

External standards for the other tests in our sample are less obvious. This ambiguity means that readers have to use their own clinical judgment about the appropriateness of repeat testing. Roughly one-third of both pulmonary examinations in our study were repeated within 1 year. Although we were surprised by this finding for pulmonary function tests, we were not surprised that this was the case for chest computed tomography given that more than one-third of the general population have lung parenchymal findings for which many would argue require surveillance.¹³ Our repeat cystoscopy findings confirm those expected by the American Urological Association, despite repeated suggestions that most patients would do well with less intensive surveillance.^{14,15} The finding that almost one-third of upper endoscopies were repeated within 3 years was considerably higher than we expected.

As a check of our analytic approach, we also investigated 2 examinations that are routinely expected to be repeated at a recommended interval. The finding that screening mammography is overwhelmingly repeated annually reflects both the recommendations existing during our study period and Medicare payment policies.¹⁶ Although we expected eye examinations to be commonly repeated, we were struck by the short repeat testing interval: almost half of examinations were repeated in 6 months or less.

Finally, because few external standards exist to judge the appropriateness of repeat testing, we explored the role

that local physician practice style might have in explaining its variation across the nation's 50 largest MSAs. If physicians had similar thresholds for diagnostic testing, we posited that the proportion of the population tested would largely reflect the population signals of disease burden in the MSA (symptom frequency, disease prevalence, etc). In other words, similar diagnostic thresholds would result in a homogeneous group of tested individuals. Therefore, little variation would exist in the need for repeat testing because the disease burden among the fraction tested would be roughly similar. However, the finding of a high positive correlation of the proportion of the population tested with the proportion of tests repeated (in the tested fraction) suggests substantial variation in physician testing thresholds across the United States.

Our work had several limitations. First, the correlation of the proportion of the population tested with the proportion of tests repeated could be evidence that market factors (such as physician supply and the presence of managed care plans) influence testing thresholds. Second, and most obvious, we did not attempt to incorporate any diagnostic information. Diagnostic codes are available in many Medicare claims and may include the diagnosis that resulted from the examination findings or the sign or symptom that was the indication for the examination.¹⁷ In the future, we intend to focus on individual tests to explore how market factors and the diagnosis at the first examination may affect repeat testing.

In conclusion, diagnostic tests are frequently repeated among Medicare beneficiaries. This has important implications not only for the capacity to serve new patients and the ability to contain costs but also for the health of the population. Although the tests themselves pose little risk, repeat testing is a major risk factor for incidental detection and overdiagnosis. Our findings should foster further research in this unstudied area.

Accepted for Publication: July 18, 2012.

Published Online: November 19, 2012. doi:10.1001/2013.jamainternmed.727

Correspondence: H. Gilbert Welch, MD, MPH, Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, 35 Centerra Pkwy, Ste 202, Lebanon, NH 03766 (H.Gilbert.Welch@dartmouth.edu).

Author Contributions: *Study concept and design:* Welch, Hayes, and Frost. *Acquisition of data:* Hayes and Frost. *Analysis and interpretation of data:* Welch, Hayes, and Frost. *Drafting of the manuscript:* Welch, Hayes, and Frost. *Critical revision of the manuscript for important intellectual content:* Hayes. *Statistical analysis:* Hayes and Frost. *Administrative, technical, and material support:* Hayes. *Study supervision:* Welch and Hayes.

Conflict of Interest Disclosures: None reported.

Disclaimer: The views expressed are those of the authors and should not be attributed to the Medicare Payment Advisory Commission.

Additional Contributions: Mark E. Miller, PhD, provided guidance and support throughout the project.

REFERENCES

1. American College of Cardiology Foundation. Appropriate use criteria. <http://content.onlinejacc.org/collection.aspx?ArticleTypeId=11789>. Accessed September 12, 2012.
2. Patel MR, Spertus JA, Brindis RG, et al; American College of Cardiology Foundation. ACCF proposed method for evaluating the appropriateness of cardiovascular imaging. *J Am Coll Cardiol*. 2005;46(8):1606-1613.
3. American College of Radiology. ACR Appropriateness Criteria: overview. 2012. <http://www.acr.org/-/media/ACR/Documents/AppCriteria/Overview.pdf>. Accessed September 12, 2012.
4. American Urological Association. Guideline for the management of nonmuscle invasive bladder cancer: (stages Ta, T1, and Tis): chapter 1: the management of bladder cancer: diagnosis and treatment recommendations. 2007 Update. <http://www.auanet.org/content/clinical-practice-guidelines/clinical-guidelines/main-reports/bladcan07/chapter1.pdf>. Accessed November 5, 2011.
5. Lokeshwar VB, Merseburger AS, Hautmann SH, eds. *Bladder Tumors: Molecular Aspects and Clinical Management*. Totowa, NY: Humana Press; 2011.
6. American Cancer Society. Guidelines for the early detection of cancer. <http://www.cancer.org/Healthy/FindCancerEarly/CancerScreeningGuidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer>. Accessed November 1, 2011.
7. US Preventive Services Task Force Screening for breast cancer. <http://www.uspreventiveservicestaskforce.org/uspstf/uspstfbrca.htm>. Accessed November 1, 2011.
8. American Optometric Association. Recommended eye examination frequency for pediatric patients and adults. <http://www.aoa.org/x5502.xml>. Accessed November 1, 2011.
9. Pearlman AS, Ryan T, Picard MH, Douglas PS. Evolving trends in the use of echocardiography: a study of Medicare beneficiaries. *J Am Coll Cardiol*. 2007;49(23):2283-2291.
10. Echocardiogram. HeartSite.com. <http://www.heartsite.com/html/echocardiogram.html>. Accessed November 1, 2011.
11. Douglas PS, Garcia MJ, Haines DE, et al; 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; Society for Cardiovascular Magnetic Resonance. 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.
12. Hendel RC, Berman DS, Di Carli MF, et al; American College of Cardiology Foundation Appropriate Use Criteria Task Force; American Society of Nuclear Cardiology; American College of Radiology; American Heart Association; American Society of Echocardiography; Society of Cardiovascular Computed Tomography; Society for Cardiovascular Magnetic Resonance; Society of Nuclear Medicine. 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. *J Am Coll Cardiol*. 2009;53(23):2201-2229.
13. Furtado CD, Aguirre DA, Sirlin CB, et al. Whole-body CT screening: spectrum of findings and recommendations in 1192 patients. *Radiology*. 2005;237(2):385-394.
14. Hall RR, Parmar MK, Richards AB, Smith PH. Proposal for changes in cystoscopic follow up of patients with bladder cancer and adjuvant intravesical chemotherapy. *BMJ*. 1994;308(6923):257-260.
15. Guney S, Guney N, Canogullari Z, Ergenekon E. Ta T1 low and intermediate transitional cell carcinoma of the bladder: recurrence rates and the timing of check cystoscopies within the first year. *Urol Int*. 2008;80(2):124-128.
16. Cancer screenings. CMS.gov. http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/cancer_screening.pdf. Accessed September 12, 2012.
17. Medicare claims processing manual: chapter 23: fee schedule administration and coding requirements. <https://www.cms.gov/manuals/downloads/clm104c23.pdf>. Accessed November 9, 2011.

INVITED COMMENTARY

Our Failure to Curb Excessive Testing

After decades of attention to unsustainable growth in health spending and its degradation of worker wages, employer economic vitality, state educational funding, and federal fiscal integrity, it is discouraging to contemplate fresh evidence by Welch et al¹ of our failure to curb waste of health care resources. These wide geographic variations in the frequency of repeated

diagnostic testing persist more than 30 years after the first salient description of wide geographic variation in health service use.² Most subsequent studies that embedded increasingly refined risk adjustment have failed to find evidence of superior treatment outcomes in high-use regions. Even after finer calculations shrink prior estimates of some regional variations,³ few in medicine or health