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 metropolitan 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.


©2012 American Medical Association. All rights reserved.
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 measured in days were divided into 5 intervals: 0 days, 1-2 months (31-60 days), 3 months (61-120 days), 6 months (121-240 days), 1 year (241-480 days), and greater than 1 year (>480 days).

#### 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 entire 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.
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: echocardiography \( (r = 0.87) \), 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.

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, our investigation provides an addition 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-
Echocardiography (55)

Imaging Stress Test (44)

Pulmonary Function Test (49)

Chest Computed Tomography (46)

Cystoscopy (41)

Upper Endoscopy (35)

Eye Examination (79)

Screening Mammography (72)

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. This is despite the specific recommendation by the American College of Cardiology Foundation Appropriate Use Criteria Task Force.
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. Similarly, our data suggest another practice viewed unfavorably by the task force, namely, routine annual imaging stress tests.
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.

---

<table>
<thead>
<tr>
<th>Test</th>
<th>Metropolitan Area</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion of Population Tested Between 2004 and 2006</td>
<td>Proportion Repeated Among Those Tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within 3 y</td>
</tr>
<tr>
<td>Need For Repetition Uncertain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Echocardiography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Miami, Florida</td>
<td>48.0</td>
</tr>
<tr>
<td>Lowest</td>
<td>Portland, Oregon</td>
<td>17.5</td>
</tr>
<tr>
<td>Imaging stress test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Orlando, Florida</td>
<td>34.4</td>
</tr>
<tr>
<td>Lowest</td>
<td>Portland, Oregon</td>
<td>11.7</td>
</tr>
<tr>
<td>Pulmonary function test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Detroit, Michigan</td>
<td>19.0</td>
</tr>
<tr>
<td>Lowest</td>
<td>San Francisco, California</td>
<td>6.5</td>
</tr>
<tr>
<td>Chest computed tomography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Jacksonville, Florida</td>
<td>17.2</td>
</tr>
<tr>
<td>Lowest</td>
<td>San Francisco, California</td>
<td>7.5</td>
</tr>
<tr>
<td>Cystoscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Miami, Florida</td>
<td>11.3</td>
</tr>
<tr>
<td>Lowest</td>
<td>Portland, Oregon</td>
<td>4.7</td>
</tr>
<tr>
<td>Upper endoscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Miami, Florida</td>
<td>16.1</td>
</tr>
<tr>
<td>Lowest</td>
<td>Sacramento, California</td>
<td>7.4</td>
</tr>
<tr>
<td>Repetition Routinely Expected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye examination</td>
<td>Sarasota, Florida</td>
<td>74.8</td>
</tr>
<tr>
<td>Lowest</td>
<td>San Bernardino, California</td>
<td>46.6</td>
</tr>
<tr>
<td>Screening mammography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>Sarasota, Florida</td>
<td>60.0</td>
</tr>
<tr>
<td>Lowest</td>
<td>Los Angeles, California</td>
<td>42.0</td>
</tr>
</tbody>
</table>
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 understudied 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


Our Failure to Curb Excessive Testing

A fter 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 al 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. 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, few in medicine or health