DOI: 10.1111/1475-6773.13150

RESEARCH ARTICLE

HSR Health Services Research

Changes in screening colonoscopy following Medicare reimbursement and cost-sharing changes

Lina D. Song MA[1,2] | Joseph P. Newhouse PhD[3,4,5,6] | Xabier Garcia-De-Albeniz MD, PhD[2,7] | John Hsu MD, MBA, MSCE[2,3]

[1]PhD Program in Health Policy, The Graduate School of Arts and Sciences, Harvard University, Cambridge, Massachusetts

[2]Health Policy Research Center, Mongan Institute, Massachusetts General Hospital, Boston, Massachusetts

[3]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts

[4]Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

[5]The John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts

[6]Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts

[7]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

Correspondence
Lina D. Song, MA, PhD Program in Health Policy, The Graduate School of Arts and Sciences, Harvard University, Cambridge, MA.
Email: dsong@fas.harvard.edu

Funding information
National Institutes of Health (NIH), Grant/Award Number: R01CA164023, P01AG032952, P01CA134294; Agency for Healthcare Research and Quality (AHRQ), Grant/Award Number: R01HS023128

Objectives: To compare existing algorithms for classifying screening vs diagnostic colonoscopies and to quantify the increase in screening colonoscopy rates when Medicare began reimbursement in 2001 and when the Affordable Care Act (ACA) eliminated cost-sharing.

Data Sources: Twenty percent random sample of fee-for-service (FFS) Medicare claims, 2000-2012.

Study Design: Using recent administrative codes as tarnished gold standards, we examined the sensitivity and specificity of five published algorithms for classifying colonoscopies and calculated annual screening colonoscopy rates. We estimated the change in rates after Medicare began reimbursement and used difference-in-differences analysis to estimate the effects of eliminating cost-sharing by comparing states with and without a mandate to cover screening colonoscopy prior to the ACA.

Findings: Model-based algorithms have higher sensitivity (0.53-0.99) than expert-based algorithms (0.35-0.39), but lower specificity (0.43-0.65 vs 0.79-0.88). All algorithms detected increases in screening from both Medicare's reimbursement change (range: 24-93/10 000) and the 2011 cost-sharing change (range: 1.1-34/10 000). Difference-in-difference estimates of the ACA's effect varied from 51 to 155 tests per 10 000 depending on the algorithm.

Conclusions: Screening colonoscopy rates increased after eliminating cost-sharing in 2011, but the increase's size varied depending on the algorithm used to classify the indication. Improvements are needed in Medicare coding for screening.

KEYWORDS
Affordable Care Act, algorithm, claims data, cost-sharing, screening colonoscopy

1 | INTRODUCTION

Following the recommendations of the US Preventive Services Task Force (USPSTF), the Affordable Care Act (ACA) eliminated cost-sharing for preventive services starting in January 2011, including colorectal cancer (CRC) screening in Medicare beneficiaries. Although CRC has a high disease burden[1,2] and screening colonoscopy reduced CRC morbidity and mortality,[3-7] the screening rate prior to the ACA remained suboptimal at an estimated range between 29 and 60 percent.[2,8,9] Screening colonoscopy for average-risk Medicare beneficiaries had been covered since July 2001, but beneficiaries without supplementary insurance were still responsible for out-of-pocket payments arising from Part B coinsurance and deductibles, a potential deterrent to screening.[10] The degree to which screening colonoscopy rates increased after the ACA eliminated cost-sharing is, therefore, an important question.

Whether overall screening colonoscopy rates increased following this policy change is not known, however, because most available data come from single centers that do not provide a national estimate or surveys that can be subject to overreporting of preventive service use.[10-14] Use of Medicare claims can potentially improve on estimates from these sources, but inferring screening intent of colonoscopy from claims data poses a challenge because one must distinguish screening from diagnostic colonoscopies, and information on the indication is often absent—or if present is inaccurate.[15-17] Some providers bill screening colonoscopies as therapeutic if, during the procedure, the endoscopist removes colonic polyps or adenomas. CMS' reimbursement rules specify that the pretest intent should determine the indication, independent of any procedures during the test,[18] and required including a screening indication code for colonoscopy. The impact of these rules has been little studied, and confusion remains over defining screening status based on completed activities vs intent.[18-20] Moreover, with the ACA's elimination of all cost-sharing for preventive services classifying the colonoscopy as a screening rather than diagnostic can save beneficiaries without supplemental coverage several hundred dollars.

Accurately assessing which colonoscopies are likely to be screening colonoscopies can thus indicate the impact of the policy change, as well as the number of beneficiaries wrongly being billed for their colonoscopy. Various algorithms have been proposed and used to classify colonoscopy indications using electronic or paper medical records,[21-27] but to date, no studies have compared them or evaluated them. This paper uses FFS Medicare claims to calculate the sensitivity and specificity of five such algorithms using recent Medicare claims-based codes. Then, it assesses changes in screening colonoscopy use after two major policy changes: (a) the introduction of Medicare reimbursement for screening colonoscopy in 2001; and (b) the elimination of cost-sharing for screening colonoscopy in 2011. Finally, it uses a difference-in-differences method to estimate the impact of ACA's elimination of cost-sharing on screening colonoscopy rates by comparing states that had enacted mandates for private insurance plans to cover CRC screening prior to the ACA with those that had not.

2 | METHODS

2.1 | Study population

We used a 20 percent sample of Medicare FFS claims (Parts A and B) and the Medicare denominator file from years 2000 to 2012. We extracted the beneficiaries' demographic, vital status, and enrollment information from the denominator file. We used Healthcare Common Procedures Coding System (HCPCS) codes and International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) diagnosis codes to identify colonoscopies, other relevant procedures, and relevant symptoms and comorbidities from Carrier, Outpatient, and Inpatient files (see Technical Appendix).

Our study population was beneficiaries who received colonoscopies between January 1, 2001, and December 31, 2012, aged 65 or older on the date they received a colonoscopy, and whose original reason for entitlement was "aged without ESRD." Although the USPSTF recommends starting CRC screening at age 50, we restricted the analysis to the beneficiaries 65 or older because of concern that those under 65 and eligible for Medicare because of disability or end-stage renal disease could differ in their use of preventive services.[28,29] We also restricted the analysis to the beneficiaries with continuous enrollment in FFS Medicare (Parts A and B) for 12 months before the date of the colonoscopy, since applying the algorithms required precolonoscopy information on utilization, diagnosis, and symptoms, yet the claim history is incomplete during any period that the beneficiary was enrolled in Medicare Advantage. After removing 51.4 percent of the original population based on these criteria, we were left with a sample of 5 871 301 colonoscopies performed on 3 808 755 beneficiaries between 2001 and 2012.

2.2 | Algorithms to identify screening colonoscopies

We began by searching PubMed (2000-Present) for existing approaches for classifying colonoscopy indications using search terms "screening colonoscopy" and "colorectal cancer" in combination with "algorithm," "intent," "claims data," "model," or "classification." There were 14 relevant studies that met our criteria, excluding those requiring pathology and laboratory test data, review articles, or case studies.[21-27,30-35] We categorized these 14 studies into four groups based on their methodologic similarity, in particular, how the algorithms were developed, what types of data they used, and whether they were validated (Technical Appendix). We selected one or two from each category based on the rate of adoption by other researchers. The process resulted in five algorithms we used to identify colonoscopy indications: El-Serag, Goodwin, Fassil, Sewitch, and Ko. Table 1 provides a detailed comparison of these five algorithms.

Among the five algorithms, three are "expert-opinion-based," (El-Serag, Goodwin, and Fassil) meaning they are designed to reflect the clinical intuition of what experts think of as true screening, defining true screening as the absence of certain CRC-related symptoms or procedures before or at the time of the examination. The two other algorithms (Sewitch and Ko) are "statistical-model-based" because they use statistical models to identify a set of variables that predict the screening indication from administrative data. Sewitch used survey data from endoscopists and patients as the gold standard of screening indication. Ko used data from the Clinical Outcomes Research Initiative (CORI) database, which includes data from the medical records of gastrointestinal endoscopy procedures from participating sites, as a gold standard[36] and matched these data to Medicare claims data. Sewitch and Ko each used two statistical methods to make their predictions: Sewitch used a multivariate logistic regression model (log) and recursive partitioning (tree), and Ko used linear discriminant analysis (LDA, a parametric model) and Classification and Regression Trees (CART, a nonparametric method). As a result, we effectively compare seven different methods.

2.3 | Test characteristics of the classification algorithms

We calculated the sensitivity, specificity, and positive predictive value (PPV) of these various algorithms in our study population using two administrative codes (CPT modifier = 33 and ICD-9 code = V76.51) as tarnished gold standards. Because we are interested in measuring the tests that originally had screening intent, we relied on the two administrative codes that were designed to reflect the patient's original screening intent, the CPT modifier 33 and the ICD-9 code V76.51, instead of the two HCPCS codes G0105 and G0121 (G0105—colorectal cancer screening, colonoscopy on individual at high risk; G0121—colorectal cancer screening, colonoscopy on individual not meeting criteria for high risk) that define screening colonoscopy more broadly. The modifier 33 was created specifically to allow providers to indicate the delivery of preventive services.[37] V76.51 is a diagnosis code for CRC screening when the patient seeks preventive care that has historically been underused, but our analysis of data indicates its use increased as a proportion of colonoscopy volume since CMS' 2009 clarification that coding should reflect the initial screening intent. The tarnished gold standard is therefore either V76.51 or modifier 33 present. We did not require the presence of both codes as a gold standard, because both codes are rarely present on the same claim; in our sample, only 0.1 percent of the claims with V76.51 also had the modifier 33 present.

We defined sensitivity as the true-positive rate, the total number of colonoscopies categorized as screening according to each algorithm divided by the total number of truly positive screening colonoscopies according to our tarnished gold standards. Similarly, we defined specificity as the true-negative rate, the total number of colonoscopies not categorized as screening according to each algorithm divided by the total number of truly negative screening colonoscopies according to our tarnished gold standards. We defined PPV as the proportion of true screening colonoscopies out of all colonoscopies categorized as screening according to each algorithm. All of these values were calculated using claims from 2011 to 2012 after modifier 33 had been implemented.

The various algorithms required some adaptation to the Medicare claims data. The Fassil algorithm was designed for application to clinical data and did not include ICD-9 codes, so we translated each criterion in the algorithm into corresponding ICD-9 and HCPCS codes. Because the symptoms in the Fassil algorithm's classification criteria were also used in other claims-based algorithms that relied on consultation with clinical experts,[24,25] we used the corresponding ICD-9 codes from those other algorithms. Since some of the original criteria in the Fassil algorithm required more detailed information than what was available from the claims data, we were only able to classify claims into four indication types (screening, diagnostic, surveillance, and unknown) instead of the eight suggested by the original model. The algorithms differed in their look-back periods, so we standardized all of them to one year so that the same set of individuals met the eligibility criteria. Goodwin, for example, considered symptoms and procedures in the three months before the colonoscopy, whereas Sewitch looked back five years.

2.4 | Robustness checks for application of the classification algorithms

We evaluated the robustness of our classification method in three ways. We first used each algorithm's original definition of the look-back period. Second, we standardized the ICD-9 codes for similar symptoms across all algorithms. For example, when referring to CRC-related symptoms and conditions such as anemia, abdominal pain, and gastrointestinal bleeding as criteria to determine indication, some models differed in the specific ICD-9 codes used in their algorithms. Thus, we made all models refer to the same set of conditions and ICD-9 codes. Third, we were concerned about distinguishing information that was obtained before rather than during or after the test, such as polypectomy performed concurrently with screening as a result of abnormal findings.[19] To do so, we removed the criteria related to post-test procedures as well as the patient's symptoms that were present on the colonoscopy claims from the algorithms. Details of these robustness checks are described in the Technical Appendix. All analyses were performed in R, version 3.2.0.

2.5 | Evaluation of policy effect

We examined changes in the use of screening colonoscopies defined by each algorithm following two Medicare policy changes. Effective from July 1, 2001, Medicare beneficiaries at average risk for CRC became eligible for a screening colonoscopy every 10 years. Screening colonoscopies for high-risk individuals already were reimbursed (Medicare considers an individual at high risk if he or she has one or more of the following: a close relative who has had colorectal cancer or an adenomatous polyp; a family history of familial adenomatous polyposis; a family history of hereditary nonpolyposis colorectal cancer; and a personal history of adenomatous polyps, colorectal cancer, or inflammatory bowel disease, including Crohn's disease, and ulcerative colitis).

Effective from January 1, 2011, section 4104 of the ACA eliminated cost-sharing for almost all preventive services covered by Medicare, including colonoscopy.[38] We measured the age-, sex-, and race-adjusted changes in screening rates following these two policy changes. To address the yearly variation in patient composition, we created a hypothetical population for each age-sex-race bracket for each year and calculated the bracket-specific screening colonoscopy rates according to the seven different algorithms. We then calculated the adjusted annual screening rate by multiplying the proportion of the population across study years in each bracket type with the bracket-specific annual rates. Last, we estimated the adjusted annual changes in screening rate by subtracting the adjusted screening rates for 2001 from those in 2002, and those in 2011 from those in 2012.

TABLE 1 Comparison of models

Model	El-Serag	Goodwin	Fassil	Sewitch	Ko
Category	Expert opinion algorithm	Expert opinion algorithm	Expert opinion algorithm	Model-based algorithm	Model-based algorithm
Year	2006	2011	2014	2013	2014
Goal	To measure the frequency trend in colorectal cancer screening in the Department of Veterans Affairs (VA)	To assess the timing of repeated colonoscopies after a negative screening colonoscopy findings among Medicare patients	To evaluate the effects of data source and adjudication on indication classification and on estimates of the effectiveness of screening colonoscopy on late-stage colorectal cancer diagnosis risk	To develop and evaluate the accuracy of model-based algorithms to identify screening colonoscopies in health administrative databases	To develop methods to classify colonoscopy indications from administrative data
Data					
Source	Inpatient and outpatient VA databases	Medicare claims and enrollment data	An observational data (patients' progress notes, referral notes, and procedure reports) of members of four U.S. health plans	Physician billing records and hospitalization data from the provincial administrative health databases of Canada	Clinical Outcomes Research Initiative (CORI) data and matched Medicare claims data
Observation period (y)	1998-2003 (6 y)	2000-2008 (9 y)	2006-2008 (3 y)	January to March 2007 (3 mo)	2006 (1 y)
Look-back period	4 y	3 mo	10 y	5 y	1 y
Classification algorithm					
Statistical model	Decision tree based on the clinical intuition of what experts think of as true screening	Decision tree based on the clinical intuition of what experts think of as true screening	Decision tree based on the clinical intuition of what experts think of as true screening	Multivariate logistic regression model and classification tree	Linear discriminant analysis (LDA, a parametric method) and classification and regression trees (CART, a nonparametric method)
Definition of screening	The absence of ICD-9 codes for 28 symptoms or conditions and no colonoscopy in the past 4 y	The absence of relevant procedures (snare, fulguration, biopsy) and diagnosis on claims or in prior 3 mo consistent with indication other than screening	Probable screening if both nonspecific symptoms and screening were recorded. High-risk screening if the test was performed for screening and the patient had a first-degree relative diagnosed with CRC before age 50, two or more second-degree relatives diagnosed at any age, or other familial syndromes. Definite average-risk screening if screening was recorded and none of the CRC conditions or risk factors noted above were recorded	For endoscopists: performed in asymptomatic people at average risk for developing colorectal cancer, or in people with a family history of colorectal cancer. For patients: The reason for colonoscopy was not to follow-up for a previous test or problem or specific lower abdominal symptoms and personal history of gastrointestinal (GI) condition were not reported	Indicated by the CORI data checkbox as average-risk colorectal cancer screening or high-risk colorectal cancer screening (family history of colorectal cancer or polyps)

(Continues)

TABLE 1 (Continued)

Model	El-Serag	Goodwin	Fassil	Sewitch	Ko
Validation	Yes	No	No	Yes	Yes
Codes used	ICD-9	ICD-9	Chart review	RAMQ, CCP, CCI, ICD-9	ICD-9, CPT
Number of categories	3 (screening, diagnostic, and surveillance)	2 (screening and diagnostic)	8 (surveillance, definitely diagnostic, probably diagnostic, possibly diagnostic, probably screening, high-risk screening, definitely screening (average-risk), and unknown)	2 (screening, nonscreening)	Four ways of grouping. 4 (diagnostic, average-risk screening, high-risk screening, and surveillance) or 3 (diagnostic, average-risk screening, and high-risk screening/surveillance combined) or 2 (diagnostic and all other indications combined) or 2 (average-risk screening and all other indications combined)
Outcome					
Sensitivity	70.1	N/A	N/A	84.5	0.57 (LDA) and 0.56 (CART)
Specificity	71.6	N/A	N/A	63.3	0.96 (LDA) and 0.96 (CART)

CCI, Canadian Classification of Health Interventions; CCP, Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures; CRC, colorectal cancer; ICD-9, The International Classification of Diseases, Ninth Revision; RAMQ, Régie de l'assurance maladie du Québec.

1. The range index colonoscopy dates in the data.
2. The number of previous years from the date of index colonoscopy considered for determining the indication of colonoscopy.
3. Sensitivity and specificity are for three-level groupings for Ko algorithm.

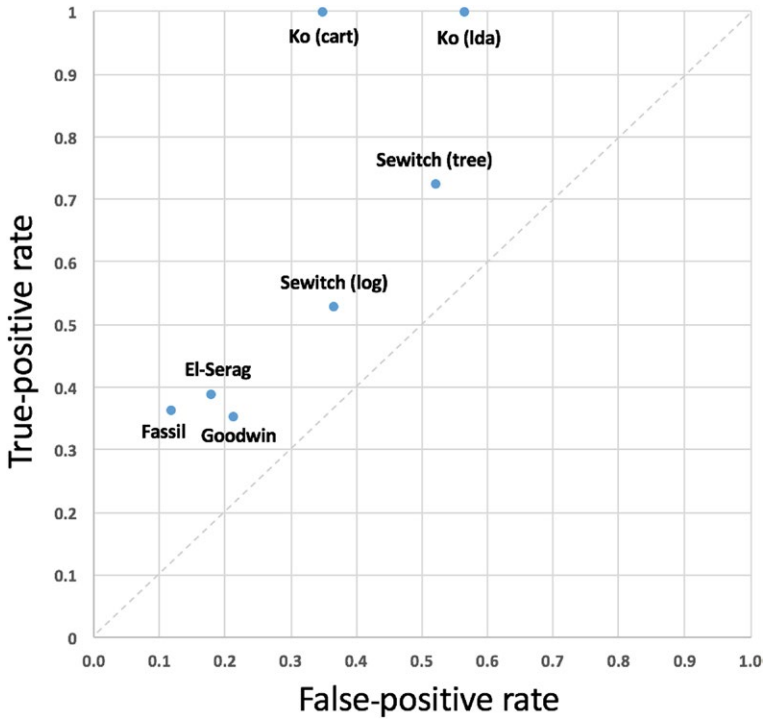

FIGURE 1 Receiver operation characteristic (ROC) curve with V76.51 or modifier 33 as gold standards, according to each model [Color figure can be viewed at wileyonlinelibrary.com]
Notes: ROC indicates receiver operating characteristic. All values were calculated using claims from 2011 to 2012 after modifier 33 has been implemented.

2.6 | Estimating the effect of ACA's elimination of cost-sharing

To estimate the effect of ACA's elimination of cost-sharing in the Medicare population, we used a difference-in-differences study design, leveraging state-level variation in the coverage of CRC screening in the private insurance market prior to the ACA, even though these mandates did not apply to Medicare beneficiaries. Prior to the ACA, 33 states and the District of Columbia had enacted mandates for private insurance plans to cover CRC screening. As a result, the screening-eligible population residing in these states ("no cost-sharing states") with private insurance had no cost-sharing for screening colonoscopy before joining Medicare, whereas those in the remaining states ("cost-sharing states") continued to face cost-sharing until it was eliminated by the ACA. The information on state-level variation in CRC screening laws prior to the ACA is publicly available.[39] Supported by the empirical evidence that the use of screening colonoscopy increased among the adults aged 50-64 in private insurance market following the state mandates,[40] we estimated the effect of the ACA's cost-sharing provision by comparing the changes in screening colonoscopy rates in the 17 no cost-sharing states and the 33 states and the District of Columbia that are cost-sharing states. We tested the parallel trend assumption by interacting the treatment variable with pre-ACA time dummies and then included state-specific time trends.

In sensitivity analyses, we excluded the three states that enacted mandates for private insurance plans in 2010 (Kentucky, Pennsylvania, and Wisconsin) for which the short time interval between the private insurance mandate and the ACA makes it difficult to separate the effects apart. We also examined excluding the ten states in which the private insurance mandates were enacted before 2001. This addressed the possibility that some individuals who received screening colonoscopies as a result of private insurance mandates before 2001 would be due for another screening in 2011 or later, which can contribute to an increase in screening colonoscopies after the ACA in these states.

Our dependent variable was a binary indicator of whether the individual received a screening colonoscopy in a given year. We included patient characteristics such as age, gender, race/ethnicity, Medicare and Medicaid dual eligibility, receipt of pneumococcal vaccine as a proxy for preventive service use, and patient comorbidities (calculated using the Charlson Comorbidity Index and Elixhauser Comorbidity Classification) to control for the patient heterogeneity.[41] We also included county-level sociodemographic and economic variables such as unemployment rate, poverty rate, and the number of gastroenterologists and hospitals per beneficiary, obtained from the Area Health Resource File. We estimated a multivariable linear regression model with a state fixed effect, interaction terms between state and a pre-/postindicator, and year. The difference-in-differences analysis addresses time-invariant unobservable state-level differences, such as variation in provider supply, market concentration, and patient clinical characteristics and preferences for preventive service use that can all be correlated with the rate of screening utilization. For an individual i in a state s in year t, our model specification is as follows:

$$Y_{ist} = \text{State}_s + \text{Year}_t + \beta_1 \text{Policy}_{s,t} + \beta_2 X_{it} + \beta_3 Z_{it} + \varepsilon_{ist}$$

where Y_{ist} is a binary variable of whether an individual i in a state s received a screening colonoscopy in year t; State_s is the state fixed effect; Year_t is the year fixed effect; $\text{Policy}_{s,t}$ is the

DID dummy which equals 1 for years 2011 and 2012 for the cost-sharing states, and equals 0 otherwise; X_{it} is a vector of individual-level control variables; and Z_{it} is a vector of county-level control variables. We clustered the standard errors only at the state level due to the potential serial correlation problem of a long time series data.[42]

3 | RESULTS

3.1 | Sensitivity and specificity

Although the model-based algorithms have higher sensitivity (0.53-0.99) than expert-based algorithms (0.35-0.39) according to our tarnished gold standard, presence of either V76.51 or modifier 33, they have lower specificity (0.43-0.65) than expert-based algorithms (0.79-0.88); in each case, the ranges do not overlap. The receiver operation characteristic (ROC) curve (Figure 1) shows the trade-off between sensitivity and specificity, where the model-based algorithms have both higher true-positive and higher false-positive rates than the expert-based algorithms. Results are comparable for the presence of V76.51 alone (Table S1).

3.2 | Robustness checks for application of classification algorithms

Our robustness checks yielded comparable findings. Although the proportion of screening vs other colonoscopy indications per algorithm changed as the various robustness tests were applied, the overall temporal trend of screening colonoscopy as well as the magnitude of variation in policy effect across classification algorithms was unchanged when we used the original look-back years, standardized the ICD-9 codes across models, and removed post-test procedures and symptoms (Figure S1).

3.3 | Number of screening colonoscopies

We compared the rate of screening, diagnostic, and surveillance colonoscopy as categorized by each algorithm adjusting for age, sex, and race. The algorithms differed markedly, a six- to sevenfold difference in the average number of screening colonoscopies identified (Figure 2). The most lenient is Ko's (LDA) algorithm and the strictest is Fassil's. We found similar magnitudes of difference in classifying diagnostic and surveillance colonoscopies.

3.4 | Estimated changes in screening following policy changes

All the algorithms indicated an increase in adjusted colonoscopy screening rates when Medicare began coverage in 2001 (Figure 2); but the magnitude of the change varied widely, from 24 (Goodwin) to 93 (Ko, LDA) per 10 000 eligible persons. Similarly, all algorithms showed an increase in screening with the ACA's elimination of beneficiary cost-sharing in 2011, but the amount varied by more than an order of magnitude across the algorithms, from 1.1 (Sewitch, log) to 34 (Ko, CART) per 10 000 eligible persons.

3.5 | Effect of ACA cost-sharing elimination in the Medicare population

The study population for estimating the ACA effect on screening colonoscopy included 9 706 726 FFS Medicare beneficiaries across 50 states and the District of Columbia between 2001 and 2012. This consists of the 6 352 340 beneficiaries residing in the 33 states and the District of Columbia in the no cost-sharing states and the 3 354 386 beneficiaries across 17 states in the cost-sharing states. Compared with the beneficiaries in the no cost-sharing states, those in the cost-sharing states were older, more likely to be white, less likely to be dually eligible, less likely

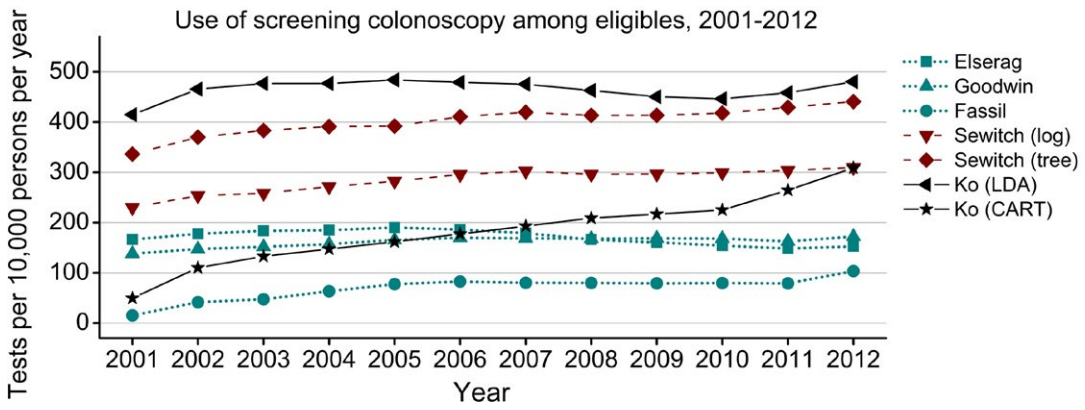

FIGURE 2 Number of age-, sex-, and race-adjusted screening colonoscopies according to seven classification algorithms [Color figure can be viewed at wileyonlinelibrary.com]

Notes: The changes in age-, sex-, race-adjusted annual screening colonoscopy rates according to each classification algorithm between 2001 and 2002 are as follows: 29.4 (El-Serag), 24.2 (Goodwin), 27.3 (Fassil), 47.6 (Sewitch, log), 69.0 (Sewitch, tree), 93.4 (Ko, LDA), and 64.5 (Ko, CART). The changes adjusted annual screening colonoscopy rates between 2011 and 2012 are as follows: 2.1 (El-Serag), 4.5 (Goodwin), 9.6 (Fassil), 1.1 (Sewitch, log), 3.8 (Sewitch, tree), 11.4 (Ko, LDA), and 34.4 (Ko, CART).

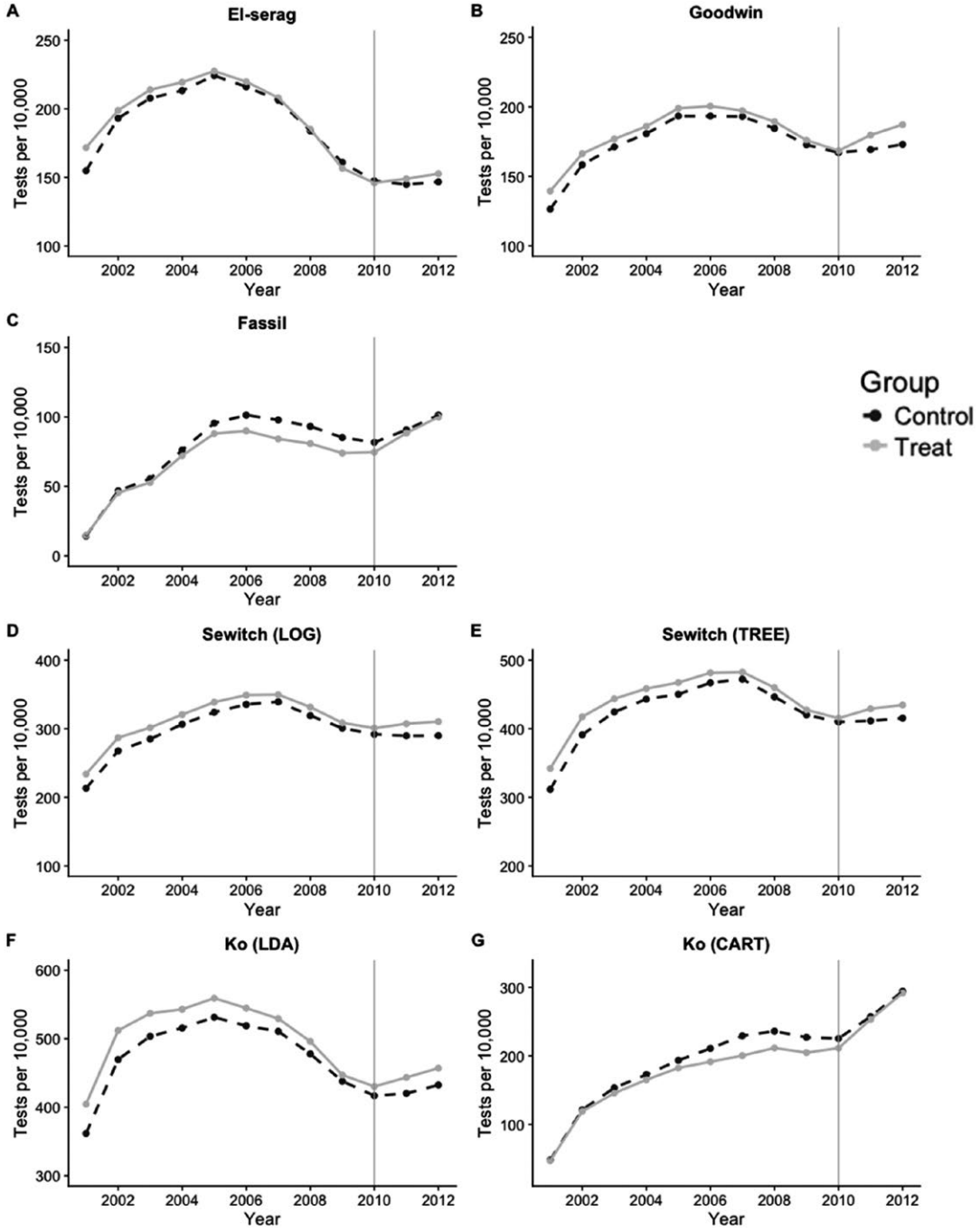

FIGURE 3 Trends in annual screening colonoscopy before and after the ACA elimination of cost-sharing according to seven classification algorithms: A, El-Serag; B, Goodwin; C, Fassil; D, Sewitch (log); E, Sewitch (tree); F, Ko (LDA); G, Ko (CART). The solid vertical line indicates the year ACA's elimination of cost-sharing for screening colonoscopy went into effect for the Medicare beneficiaries

to have comorbidities, less likely to receive a pneumococcal vaccine, and more likely to receive a colonoscopy (Table S2). All of the covariates except for the receipt of colonoscopy were included in the regression model to control for the individual-level heterogeneity. Prior to the ACA, the trends in annual screening colonoscopy rates in both groups were declining at a statistically different but reasonably comparable rate (Table S3). Following the ACA, the trend reversed and rose according to all algorithms (Figure 3). More noteworthy, states with cost-sharing prior to the ACA experienced a significantly greater increase in the annual screening colonoscopy after the ACA than those without according to four out of seven algorithms, ranging from 51.1 (95% CI: 7.2-95.0, El-Serag) to 155.2 (95% CI: 78.0-232.4, Ko CART) tests per 10 000 persons (Table 2). When we included the state-specific time trends, more (five out of seven) algorithms indicated a significant increase. When the three states with private

TABLE 2 Difference-in-differences models of the effect of eliminating cost-sharing for screening colonoscopy on annual screening colonoscopy rates according to seven classification algorithms

		Alternative models		
Outcomes (tests per 10 000)	Main model	State-specific time trends	Exclude 2010 mandate states	Exclude pre-2001 states
Screening colonoscopy, El-Serag	51.1* (22.4)	52.4 (32.7)	49.4* (22.9)	47.7* (20.9)
Screening colonoscopy, Goodwin	48.0 (26.7)	78.9*** (16.8)	66.7* (26.3)	23.2 (25.1)
Screening colonoscopy, Fassil	86.9*** (19.5)	44.4* (21.7)	85.0*** (19.5)	67.6*** (19.2)
Screening colonoscopy, Sewitch, log	41.8 (24.7)	66.5* (26.9)	32.9 (26.0)	32.3 (24.3)
Screening colonoscopy, Sewitch, tree	−22.9 (12.7)	54.4 (32.7)	−34.3** (13.3)	−22.3 (13.7)
Screening colonoscopy, Ko, LDA	52.7* (22.0)	82.9*** (22.5)	70.4** (23.9)	67.4 (21.9)
Screening colonoscopy, Ko, CART	155.2*** (39.4)	145.8*** (24.2)	166.9*** (39.3)	137.5** (41.8)

Notes: Values indicate DID effect (robust SE). Difference-in-difference estimates are from a multivariate linear regression model. Standard errors were clustered at the state level. *P < 0.05, **P < 0.01, ***P < 0.001.

insurance mandates enacted in 2010 or the ten states with private insurance mandates enacted before 2001 were removed from the analysis, similar number of algorithms (five and three out of seven, respectively) predicted an increase in screening colonoscopies.

4 | DISCUSSION

Accurate estimates of screening colonoscopy rates are important for setting and evaluating policy and improving care. We compared existing algorithms for identifying screening colonoscopies in claims data and found substantial differences in their construction and performance. Those algorithms we designated as statistical-model-based have higher sensitivity but poorer specificity than those we designated as expert-opinion-based. Sensitivity is likely a better test characteristic than specificity, however, since our tarnished gold standards are underused.

Nonetheless, any preference for the statistical-model-based algorithms requires qualification. First, the V76.51 or modifier 33 may be present on nonscreening colonoscopy claims from misclassification. Second, the model-based algorithms might be performing better than the expert-based algorithms because they are calibrated based on some of the administrative codes we used for validation. For example, Ko's algorithm has 99.9 percent sensitivity because it uses V76.51 as a criterion to determine screening colonoscopies. In general, there is a challenge with applying the machine learning approaches developed in one reimbursement context to another. The parameters that are estimated for these types of algorithms will be valid only if the variables in the data used to develop the algorithm hold the same relationship with the variables in the data that the algorithms would be applied to. All four model-based algorithms we examined were estimated using data from before the ACA changes, and their external validity may, therefore, be compromised in later years. Expert-based algorithms, however, rely on the clinical definition of screening and so are less dependent on the contextual setting of the original data.

Using these algorithms, we estimated the change in screening colonoscopy utilization among Medicare beneficiaries following two important policy changes, the decision to cover screening tests for all average-risk beneficiaries in 2001 and the removal of cost-sharing in 2011. Previous studies showed an increase in screening colonoscopy volume between 2001 and 2002,[43-46] and all the algorithms we studied also showed an increase in rates of use following the adoption of both policies. But the magnitude of the estimated increase differed by fourfold for the 2001 policy change and 30-fold for the 2011 policy change across the algorithms, indicating that evaluating the effects of changed coverage policies for colonoscopy screening from claims data is significantly model dependent.

We also exploited state-level variation in CRC screening mandates for commercial insurance before the ACA to estimate the effect of ACA cost-sharing removal on screening colonoscopies in Medicare. Elimination of cost-sharing modestly increased screening colonoscopy rates and was predicted consistently by four out of seven algorithms we examined. One previous study, for example, found no significant uptake in colonoscopy among high-risk FFS Medicare beneficiaries with the abolition of cost-sharing.[12] Another study that used the Behavioral Risk Factor Surveillance System (BRFSS) data found an increase in annual total colonoscopy rates among elderly men when cost-sharing was abolished, but that study combined screening with other indications.[47] Although the fact that majority of the algorithms indicate a significant increase strengthens the conclusion that the elimination of cost-sharing had a positive impact on screening colonoscopies, the magnitude of the ACA effect estimates varied more than threefold across algorithms, which emphasizes

the challenges in accurately estimating the policy effect using the classification algorithms. When we included state-specific time trend, more algorithms predicted a significant increase in screening colonoscopy, suggesting that the increase in screening colonoscopy is consistent under the assumption of variation in time trends by states. Our evaluation of the performance based on tarnished gold standards shows that some algorithms dominate others in terms of test characteristics. The efficiency frontier of the ROC curve is made up of the Fassil and Ko (CART) algorithms. These two algorithms predicted the highest increases (87-155 per 10 000) in screening colonoscopies after the ACA, a meaningful increase given the annual number of colonoscopies was in the range of 415-481 per 10 000 according to the most lenient algorithm (Ko, LDA).

A major strength of our study is to provide a common comparison point for multiple methods of identifying screening colonoscopies from claims data and to show that estimated effect sizes are model dependent. Comparing existing approaches to distinguish screening and diagnostic colonoscopies using FFS Medicare claims is attractive because FFS Medicare covers a large sample of U.S. elderly population and the data are free of recall bias. Previous studies on screening colonoscopy utilization either relied on surveys that can be subject to self-reporting bias for recommended health care utilization or were limited to high-risk populations.[10-14] By using a quasi-experimental study design, our study provides an estimate of the screening colonoscopy utilization as well as quantifying the variation in the effect estimates through multiple existing classification algorithms.

Our study has several limitations. First, we did not compare all existing classification algorithms but only selected several that are applicable to claims data and representative among expert- and model-based approaches. Because the differences across the two approaches are large and consistent, we believe that including other models of the same type would not alter the key messages of our study. Second, we had to make some modifications to compare algorithms that were developed using data other than Medicare claims. Although we conducted several robustness checks, we could have altered the original intentions of the authors. Third, we did not examine the 33 percent of Medicare beneficiaries enrolled in Medicare Advantage (MA) plans as of 2018 [ref for 33 percent is https://www.kff.org/medicare/fact-sheet/medicareadvantage/]. Given the evidence that the beneficiaries in Medicare HMOs are more likely to receive appropriate preventive services than those in the traditional Medicare program[48,49] and that most MA enrollees are in HMO's, the colonoscopy patterns in the MA population deserve separate attention. Fourth, our difference-in-differences analysis assumed that the states that mandated free screening colonoscopy prior to the ACA would have a smaller percentage of beneficiaries who faced no cost-sharing following the ACA, but our data cannot test that assumption. Finally, although we adjusted for the fixed difference by states, our model cannot address any unobservable time-varying trends between the control and treatment states. Although pre-ACA screening colonoscopy trends were fairly parallel between the two groups, patient preferences for preventive services utilization or state legislative environments may evolve differently over time and disproportionately affect screening colonoscopy utilization between these two groups of states.

In summary, Medicare claims data provide evidence that screening colonoscopy among FFS Medicare beneficiaries increased following the ACA's effort to promote the utilization of evidence-based preventive services by mandating no cost-sharing, but the estimated magnitude of the increase differs substantially across approaches used to distinguish screening and therapeutic colonoscopies. Because there is unlikely to be a gold standard for identifying screening colonoscopies from claims data any time soon, these are probably the best national estimates that can be made at the moment. The differing estimates among the algorithms, however, highlight the need for further research, perhaps using nationally representative electronic medical record data in conjunction with claims, to estimate screening colonoscopy rates. In the meantime, researchers should be aware of the potential unintended model dependency when conducting claims-based research, and carefully examine the assumptions of the classification models they choose and perform extensive sensitivity analyses. In any event, the need to develop an agreed-upon reference standard for screening intent and to promote such a standard is clear.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors would like to acknowledge the valuable feedback provided by Miguel A. Hernán, MD, DrPH (Harvard T. H. Chan School of Public Health), Maggie Price (Massachusetts General Hospital), attendees of the Research-in-Progress Seminar at the Department of General Internal Medicine, Massachusetts General Hospital, and thoughtful comments provided by the anonymous reviewers.

DISCLOSURE

This research was supported by the National Institutes of Health (NIH) and the Agency for Healthcare Research and Quality (AHRQ) under award numbers R01CA164023, R01HS023128, P01AG032952, and P01CA134294. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or AHRQ. At the time this work was carried out, Dr. Newhouse was a director of and held equity in Aetna. The funding sources and potential conflict of interest played no role in study design; data collection, analysis, or interpretation; writing of the manuscript; or the decision to submit the paper for publication.

ORCID

Lina D. Song https://orcid.org/0000-0003-0971-8054

REFERENCES

1. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin*. 2017;67:177-193.
2. Centers for Disease Control and Prevention. Vital signs: colorectal cancer screening among adults aged 50-75 years – United States, 2008. *MMWR Morb Mortal Wkly Rep*. 2010;59:808-812.
3. Smith RA, Andrews K, Brooks D, et al. Cancer screening in the United States, 2016: a review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin*. 2016;66:96-114.
4. USPSTF. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2008;149:627-637.
5. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology*. 2008;134:1570-1595.
6. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2012;143:844-857.
7. Whitlock EP, Lin JS, Liles E, Beil TL, Fu R. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2008;149:638-658.
8. Joseph DA, King JB, Miller JW, Richardson LC, Centers for Disease Control and Prevention. Prevalence of colorectal cancer screening among adults–Behavioral Risk Factor Surveillance System, United States, 2010. *MMWR Suppl*. 2012;61:51-56.
9. Cooper GS, Doug Kou T. Underuse of colorectal cancer screening in a cohort of Medicare beneficiaries. *Cancer*. 2008;112:293-299.
10. Hamman MK, Kapinos KA. Affordable care act provision lowered out-of-pocket cost and increased colonoscopy rates among men in Medicare. *Health Aff (Millwood)*. 2015;34:2069-2076.
11. Fedewa SA, Goodman M, Flanders WD, et al. Elimination of cost-sharing and receipt of screening for colorectal and breast cancer. *Cancer*. 2015;121:3272-3280.
12. Cooper GS, Kou TD, Schluchter MD, Dor A, Koroukian SM. Changes in receipt of cancer screening in Medicare beneficiaries following the Affordable Care Act. *J Natl Cancer Inst*. 2016;108:5:djv374.
13. Peterse EFP, Meester RGS, Gini A, et al. Value of waiving coinsurance for colorectal cancer screening in medicare beneficiaries. *Health Aff (Millwood)*. 2017;36:2151-2159.
14. Rauscher GH, Johnson TP, Cho YI, Walk JA. Accuracy of self-reported cancer-screening histories: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2008;17:748-757.
15. Singal AG, Gupta S, Lee J, et al. Importance of determining indication for colonoscopy: implications for practice and policy original. *Clin Gastroenterol Hepatol*. 2014;12:1958-1963. e1-e3.
16. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. *JAMA*. 2004;291:2466-2470.
17. Freeman JL, Klabunde CN, Schussler N, Warren JL, Virnig BA, Cooper GS. Measuring breast, colorectal, and prostate cancer screening with medicare claims data. *Med Care*. 2002;40:IV-36-42.
18. CMS. Coding for polypectomy performed during screening colonoscopy or flexible sigmoidoscopy. MLN Matters. 2012.
19. Green BB, Coronado GD, Devoe JE, Allison J. Navigating the murky waters of colorectal cancer screening and health reform. *Am J Public Health*. 2014;104:982-986.
20. Sewitch MJ, Hilsden R, Joseph L, et al. Qualitative study of physician perspectives on classifying screening and nonscreening colonoscopy using administrative health data: adding practice does not make perfect. *Can J Gastroenterol*. 2012;26:889-893.
21. El-Serag HB, Petersen L, Hampel H, Richardson P, Cooper G. The use of screening colonoscopy for patients cared for by the Department of Veterans Affairs. *Arch Intern Med*. 2006;166:2202-2208.
22. Fassil H, Adams KF, Weinmann S, et al. Approaches for classifying the indications for colonoscopy using detailed clinical data. *BMC Cancer*. 2014;14:95.
23. Sewitch MJ, Jiang M, Joseph L, Hilsden RJ, Bitton A. Developing model-based algorithms to identify screening colonoscopies using administrative health databases. *BMC Med Inform Decis Mak*. 2013;13:45.
24. Goodwin JS, Singh A, Reddy N, Riall TS, Kuo YF. Overuse of screening colonoscopy in the Medicare population. *Arch Intern Med*. 2011;171:1335-1343.
25. Ko CW, Dominitz JA, Neradilek M, et al. Determination of colonoscopy indication from administrative claims data. *Med Care*. 2014;52:e21-e29.
26. Lee JK, Jensen CD, Lee A, et al. Development and validation of an algorithm for classifying colonoscopy indication. *Gastrointest Endosc*. 2015;81:575-582 e4.
27. Adams KF, Johnson EA, Chubak J, et al. Development of an algorithm to classify colonoscopy indication from coded health care data. *EGEMS (Wash DC)*. 2015;3:1171.
28. Cubanski J, Neuman P. Medicare doesn't work as well for younger, disabled beneficiaries as it does for older enrollees. *Health Aff (Millwood)*. 2010;29:1725-1733.
29. Na L, Streim JE, Pezzin LE, et al. Disparities in receipt of recommended care among younger versus older Medicare beneficiaries: a cohort study. *BMC Health Serv Res*. 2017;17:241.
30. Fisher DA, Grubber JM, Castor JM, Coffman CJ. Ascertainment of colonoscopy indication using administrative data. *Dig Dis Sci*. 2010;55:1721-1725.
31. Haque R, Chiu V, Mehta KR, Geiger AM. An automated data algorithm to distinguish screening and diagnostic colorectal cancer endoscopy exams. *J Natl Cancer Inst Monogr*. 2005;2005:116-118.
32. Corley DA, Jensen CD, Marks AR, et al. Variation of adenoma prevalence by age, sex, race, and colon location in a large population: implications for screening and quality programs. *Clin Gastroenterol Hepatol*. 2013;11:172-180.
33. Doubeni CA, Laiyemo AO, Klabunde CN, Young AC, Field TS, Fletcher RH. Racial and ethnic trends of colorectal cancer screening among Medicare enrollees. *Am J Prev Med*. 2010;38:184-191.
34. Doubeni CA, Laiyemo AO, Reed G, Field TS, Fletcher RH. Socioeconomic and racial patterns of colorectal cancer screening among Medicare enrollees in 2000 to 2005. *Cancer Epidemiol Biomarkers Prev*. 2009;18:2170-2175.
35. Meester RG, Doubeni CA, Lansdorp-Vogelaar I, et al. Variation in adenoma detection rate and the lifetime benefits and cost of colorectal cancer screening: a microsimulation model. *JAMA*. 2015;313:2349-2358.
36. Weiss NS, Doria-Rose VP. The effectiveness of colonoscopy in reducing mortality from colorectal cancer. *Ann Intern Med*. 2009;150:817-818; author reply 9-20.
37. CMS. Preventive and Screening Services Update – Intensive Behavioral Therapy for Obesity, Screening Digital Tomosynthesis Mammography, and Anesthesia Associated with Screening Colonoscopy. MLN Matters. 2015.
38. PPACA. Patient Protection and Affordable Care Act of 2010 (PPACA) §§4103, 4104. Pub L. 2010:111-48.
39. Colorectal Cancer Screening: What Are States Doing? http://www.ncsl.org/research/health/colorectal-cancer-screening-laws

-by-state.aspx: National Conference of State Legislatures; 2011 [updated 2011; cited 2017 Jan 1].
40. Cokkinides V, Bandi P, Shah M, Virgo K, Ward E. The association between state mandates of colorectal cancer screening coverage and colorectal cancer screening utilization among US adults aged 50 to 64 years with health insurance. *BMC Health Serv Res*. 2011;11:19.
41. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8-27.
42. Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? *Q J Econ*. 2004;119:249-275.
43. Shih YC, Zhao L, Elting LS. Does Medicare coverage of colonoscopy reduce racial/ethnic disparities in cancer screening among the elderly? *Health Aff (Millwood)*. 2006;25:1153-1162.
44. Harewood GC, Lieberman DA. Colonoscopy practice patterns since introduction of medicare coverage for average-risk screening. *Clin Gastroenterol Hepatol*. 2004;2:72-77.
45. Gross CP, Andersen MS, Krumholz HM, McAvay GJ, Proctor D, Tinetti ME. Relation between Medicare screening reimbursement and stage at diagnosis for older patients with colon cancer. *JAMA*. 2006;296:2815-2822.
46. Fenton JJ, Cai Y, Green P, Beckett LA, Franks P, Baldwin LM. Trends in colorectal cancer testing among Medicare subpopulations. *Am J Prev Med*. 2008;35:194-202.
47. Hamman MK, Kapinos KA. Colorectal cancer screening and state health insurance mandates. *Health Econ*. 2016;25:178-191.
48. Brennan N, Shepard M. Comparing quality of care in the Medicare program. *Am J Manag Care*. 2010;16:841-848.
49. Ayanian JZ, Landon BE, Zaslavsky AM, Saunders RC, Pawlson LG, Newhouse JP. Medicare beneficiaries more likely to receive appropriate ambulatory services in HMOs than in traditional medicare. *Health Aff (Millwood)*. 2013;32:1228-1235.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Song LD, Newhouse JP, Garcia-De-Albeniz X, Hsu J. Changes in screening colonoscopy following Medicare reimbursement and cost-sharing changes. *Health Serv Res*. 2019;00:1–12. https://doi.org/10.1111/1475-6773.13150