

# Developing and Testing an Electronic Measure of Screening Colonoscopy Overuse in a Large Integrated Healthcare System

Sameer D. Saini, MD, MS<sup>1,2,3</sup>, Adam A. Powell, PhD<sup>4,5</sup>, Jason A. Dominitz, MD, MHS<sup>6</sup>, Deborah A. Fisher, MD, MHS<sup>7,8</sup>, Joseph Francis, MD<sup>9</sup>, Linda Kinsinger, MD, MPH<sup>10</sup>, Kathleen S. Pittman, RN, MPH<sup>10</sup>, Philip Schoenfeld, MD, MEd, MSc(Epi)<sup>1,2,3</sup>, Stephanie E. Moser, PhD<sup>1,11</sup>, Sandeep Vijan, MD, MS<sup>1,2,3</sup>, and Eve A. Kerr, MD, MPH<sup>1,2,3</sup>

<sup>1</sup>Veterans Affairs Center for Clinical Management Research (CCMR), Ann Arbor, MI, USA; <sup>2</sup>Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA; <sup>3</sup>Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, USA; <sup>4</sup>Center for Chronic Disease Outcomes Research (CCDOR), Minneapolis VA Health Care System, Minneapolis, MN, USA; <sup>5</sup>Anthem, Inc, Minneapolis, MN, USA; <sup>6</sup>VA Puget Sound Health Care System, Seattle, WA, USA; <sup>7</sup>Durham Veterans Affairs Medical Center, Durham, NC, USA; <sup>8</sup>Department of Medicine, Duke University, Durham, NC, USA; <sup>9</sup>VHA Office of Analytics and Business Intelligence, Washington, DC, USA; <sup>10</sup>VHA National Center for Health Promotion and Disease Prevention, Durham, NC, USA; <sup>11</sup>Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA.

**BACKGROUND:** Most existing performance measures focus on underuse of care, but there is growing interest in identifying and reducing overuse.

**OBJECTIVE:** We aimed to develop a valid and reliable electronic performance measure of overuse of screening colonoscopy in the Veterans Affairs Health Care System (VA), and to quantify overuse in VA.

**DESIGN:** This was a cross-sectional study with multiple cross-sections.

**SUBJECTS:** U.S. Veterans who underwent screening colonoscopy between 2011 and 2013.

**MAIN MEASURES:** Overuse of screening colonoscopy, using a validated electronic measure developed by an expert workgroup.

**KEY RESULTS:** Compared to results obtained from manual record review, the electronic measure was highly specific (97 %) for overuse, but not sensitive (20 %). After exclusion of diagnostic and high-risk screening or surveillance procedures, the validated electronic measure identified 88,754 average-risk screening colonoscopies performed in VA during 2013. Of these, 20,530 (23 %) met the definition for probable (17 %) or possible (6 %) overuse. Substantial variation in colonoscopy overuse was noted between Veterans Integrated Care Networks (VISNs) and between facilities, with a nearly twofold difference between the maximum and minimum rates of overuse at the VISN level and a nearly eightfold difference at the facility level. Overuse at the VISN and facility level was relatively stable over time.

**CONCLUSIONS:** Overuse of screening colonoscopy can be measured reliably and with high specificity using electronic data, and is common in a large integrated healthcare system. Overuse measures, such as those we have specified through a consensus workgroup process, could be combined with underuse measures to improve the appropriateness of colorectal cancer screening.

**KEY WORDS:** performance measurement; colorectal cancer; screening; health services research.

J Gen Intern Med 31(Suppl 1):S53-60  
DOI: 10.1007/s11606-015-3569-y  
© Society of General Internal Medicine 2016

## INTRODUCTION

As evidence in healthcare has mounted in recent years, we have seen a growing interest in the development of performance measures to assess overuse. This is particularly true in preventive care, where most existing measures focus on underuse.<sup>1</sup> Underuse measures have played an important role in increasing the uptake of preventive care.<sup>2</sup> But as we continue to encourage prevention for patients who are likely to benefit, it is imperative that we also find ways of discouraging the use of unnecessary and potentially harmful services to improve quality of care, decrease waste, and improve access. Efforts such as the *Choosing Wisely* campaign have highlighted specific clinical areas where measurement of overuse may be warranted.<sup>3</sup> Yet, we have little practical experience with measures of overuse and their broader effects on healthcare delivery.

A key challenge in implementing performance measures is avoiding unintended effects. For example, several recent studies have shown that implementation of underuse measures can unintentionally encourage overuse of care.<sup>4-7</sup> In contrast, overuse measures have the potential to discourage utilization when care is actually indicated, promoting underuse. In light of such concerns, Mathias and Baker have recommended that the potential unintended effects of overuse measures be carefully considered during measure development, and that efforts be taken to mitigate and monitor for such effects.<sup>8</sup> As a starting point, they suggest that overuse measures be based on strong evidence and that these measures demonstrate high specificity to avoid falsely identifying overuse, thereby minimizing the risk of promoting underuse.

The purpose of this study was to develop and test an electronic measure of screening colonoscopy overuse in the Veterans Affairs Health Care System (VA), and to estimate its overuse within VA. Screening colonoscopy is an ideal candidate for an overuse measure for a number of reasons. First, appropriate use of colorectal cancer screening is supported by strong evidence.<sup>9</sup> Second, screening colonoscopy is often overused in both VA and non-VA settings.<sup>10,11</sup> Third, existing performance measures for colorectal cancer screening are focused exclusively on underuse of screening, and we have shown that implementation of these measures within VA is associated with overuse.<sup>12</sup> Additionally, colonoscopy is an invasive test with a small but real risk of complications.<sup>13</sup> Finally, because colonoscopy is a resource-limited service, overuse reduces access to necessary and appropriate colonoscopy.

It is challenging to identify inappropriate screening colonoscopies using electronic data for several reasons. First, administrative data have limited accuracy for determining whether a colonoscopy was performed for average-risk screening or for a non-screening indication (e.g., for diagnostic purposes or post-polypectomy surveillance).<sup>14,15</sup> Second, the appropriate interval between colonoscopies is dependent on knowing the results of prior colonoscopies. For example, a patient with adenomas should have a repeat (surveillance) colonoscopy sooner than a patient with a completely normal colonoscopy<sup>16</sup>; however, detailed information about the type and number of polyps resected is not captured in administrative data.<sup>17</sup> Third, many colonoscopies are appropriately repeated sooner than recommended by guidelines due to suboptimal bowel preparation, but this factor is also not captured in administrative data.<sup>18</sup>

For these reasons, working with operational partners in VA, we conducted a study with three main aims: (1) to develop a highly specific measure that can be used to identify overuse of screening colonoscopy from electronic data; (2) to assess how reliably this measure identifies facility-level overuse over time; and (3) to examine facility-level factors that may be associated with colonoscopy overuse.

## METHODS

### Overview

We performed a cross-sectional study using electronic data from the VA Health Care System for each year from 2011 to 2013. The primary analysis utilized data from the year 2013. The study proceeded in four sequential steps. First, we convened a workgroup of experts in colorectal cancer screening and performance measurement to specify an overuse measure for screening colonoscopy. Second, using the workgroup specifications, we developed an electronic measure of overuse. Third, we validated this measure in a subset of patients

by comparing it to gold-standard manual record review. Finally, we used the validated electronic measure to quantify overuse of screening colonoscopy across VA. The study was approved by the Institutional Review Board of the VA Ann Arbor Healthcare System.

### Measure Specification and Electronic Measure Construction

A performance measure of screening colonoscopy overuse was specified by an expert workgroup composed of VA experts in colorectal cancer screening (DF, PS, JD, KP) and performance measurement (EK, JF, LK). The workgroup process is described in detail in the Appendix. The final measure defined by the workgroup identified average-risk screening colonoscopies (comprising the denominator, with exclusions—Text Box 1A) that met one or more criteria for probable or possible overuse (comprising the numerator—Text Box 1B). The workgroup met three times over a 6-month period to define and review the measures and provide feedback on preliminary results of electronic data extraction.

#### Text Box 1A: Denominator and Exclusions for Screening Colonoscopy Overuse Measure

<b>Denominator:</b>	Patients who underwent colonoscopy over the selected time period (N=248,284).
<b>Exclusions:</b>	Colonoscopy performed for a non-screening indication (N=159,530):
	(1) Colonoscopy performed for diagnostic, high-risk screening, or surveillance indication (N=86,904).*
	(2) Colonoscopy performed in a patient at increased risk for colorectal cancer (N=72,212)**:
	a. Personal history of adenomatous polyps (N=68,039).
	b. Personal history of colorectal cancer (N=916).
	c. Personal history of inflammatory bowel disease (N=170).
	d. Family history of colorectal cancer (N=3087).
	(3) Colonoscopy performed during hospitalization (N=411).
	(4) Colonoscopy performed in a patient who has undergone prior total abdominal colectomy (N=3)**.

\*Using approach previously developed and validated by Fisher and colleagues (Appendix)

\*\*Using CPT and ICD-9 codes from FY00 to FY13 (Appendix)

#### Text Box 1B: Numerator for Screening Colonoscopy Overuse Measure

<b>Numerator:</b>	"Probable" overuse of screening colonoscopy:
	(1) Colonoscopy performed less than 9 years after complete colonoscopy.
	(2) Colonoscopy performed in a patient <40 or >85 years of age.
	(3) Colonoscopy performed in a patient with life expectancy <6 months.
	(4) Colonoscopy performed <6 months after negative fecal occult blood test (FOBT).
	"Possible" overuse of screening colonoscopy:
	(1) Colonoscopy performed in a patient 40 to 49 years of age.
	(2) Colonoscopy performed in a patient 76 to 85 years of age.

After the measure had been specified, we approximated measure elements from VA electronic data that were available from fiscal year (FY) 2000 to the present time in a Corporate Data Warehouse (CDW), including: (1) patient demographics; (2) diagnostic and procedure codes; (3) use and results of laboratory testing; and, (4) structured documentation of chronic and serious health conditions, including hospice enrollment. A detailed description of the methodology used to electronically approximate the measure denominator and numerator can be found in the [Appendix](#).

### Validation of the Electronic Measure Using Manual Record Review

We developed a standardized electronic health record (EHR) abstraction algorithm to identify measure elements in manual record review. This abstraction algorithm was piloted prior to use, and pilot data was used to modify the algorithm to enhance its ease of use and reliability. Once the abstraction algorithm was finalized, we selected a sample of 3,000 Veterans who had a procedure code for colonoscopy ([Appendix Table 4](#)) between April 2011 and March 2012, using a stratified random sampling strategy based on predicted overuse at each facility. First, we used our candidate electronic measure to estimate the proportion of screening colonoscopies at each VA facility that met a definition for overuse in FY11. Next, we used these data to stratify facilities into three groups: bottom quartile (low expected overuse); middle two quartiles; and top quartile (high expected overuse). Finally, we randomly sampled 3000 Veterans with electronic documentation of colonoscopy from each of these three groups (1000 from low overuse facilities, 1000 from average overuse facilities, and 1000 from high overuse facilities). Manual record review was performed by West Virginia Medical Institute, a professional chart abstraction group that performs large-scale, national chart reviews for VA performance measurement programs on an ongoing basis. Further detail about the manual record review process is provided in the [Appendix](#).

### Independent Variables

We obtained or constructed several potential facility-level predictors of screening overuse: (1) proportion of screen-eligible patients in FY13 who were “up to date” for screening according to current guidelines (assessed by chart review conducted as part of VA’s ongoing External Peer Review Program); (2) median number of days between positive fecal occult blood test (FOBT) and colonoscopy (“FOBT wait time”) (see below); (3) complexity score (a measure of facility complexity based on factors such as patient risk, academic affiliation, and patient volume);<sup>19</sup> (4) academic affiliation (obtained from the VA Office of Academic Affiliations); (5) number of colonoscopies performed in FY13 (obtained from CDW); and, (6) proportion of colonoscopies “outsourced” to non-VA facilities in FY10 (also known as “fee basis” colonoscopies) (obtained from CDW). To quantify the FOBT wait time, we identified all

positive FOBTs in FY12 ( $N=55,494$ ) and all colonoscopies performed within 12 months after positive FOBT ( $N=22,728$ ). We then calculated the median time in days from positive FOBT to colonoscopy for each facility. Data were also extracted from CDW on several patient-level variables, including age, gender, and Charlson comorbidity index.<sup>20</sup>

### Statistical Analysis

We compared assessments of overuse constructed from manual record review to those constructed from electronic data. Specifically, we examined diagnostic test characteristics (sensitivity and specificity) and simple agreement for probable overuse, possible overuse, and overall overuse. These test characteristics were examined for both national VA data and for each stratum of predicted overuse, as specified in our stratified sampling strategy. Additionally, underlying reasons for overuse were tabulated (corresponding to the individual elements comprising the numerator of our overuse measure). VA facilities are organized into 21 regional Veterans Integrated Service Networks (VISNs), each of which has substantial autonomy over budgetary decisions and approaches to delivery of care. We therefore examined variation at the facility level and also the VISN level.

Multilevel, multivariable logistic regression was performed to identify independent facility-level predictors of screening overuse. We also examined stability in overuse over time at the national, VISN, and facility level (from FY11 to FY13). Data were missing for less than 1 % of the sample for descriptive analysis and for approximately 15 % of predictor variables for multivariable analysis. These records were excluded from multivariable analysis. Analyses were performed using the Stata 12 statistical package (StataCorp, College Station, Texas).

## RESULTS

### Validation of Electronic Measure

We identified 3000 Veterans who had electronic documentation of a colonoscopy in FY11. Of these, 2976 (99 %) had documentation of a colonoscopy in manual record review; 2915 (97 %) had documentation of colonoscopy indication and were used for validation of the electronic measure. Of these 2915 colonoscopies, 2675 were identified as either non-screening (2171) or appropriate screening (504) in manual record review (i.e., did not meet a measure definition of overuse), and 240 met a measure definition of probable or possible overuse. As per our a priori study design, the electronic measure was highly specific, but not sensitive, for overuse compared to manual record review. Out of the 2675 colonoscopies that were non-screening or appropriate screening, 2585 (specificity=97 %) were correctly identified as appropriate by the electronic measure. Out of the 240 colonoscopies that met a definition of overuse by manual record

review, 47 (sensitivity=20 %) were correctly identified as overuse by the measure. Low sensitivity of the electronic measure for overuse was largely due to low sensitivity of the electronic measure for screening indication (sensitivity=36 %). Specificity and sensitivity for overuse were similar in high-performing facilities (i.e., those with low rates of overuse) (specificity=97 %, sensitivity=19 %) and in low-performing facilities (specificity=95 %, sensitivity=27 %).

**Electronic Measurement of Overuse**

Baseline characteristics of facilities and patients are shown in Tables 1 and 2, respectively. In FY13, 88,754 screening colonoscopies were identified across 122 VA facilities (Tables 1 and 2). Of these, 20,530 (23 %) met the definition for probable (17 %) or for possible (6 %) overuse (Table 3). The most common reasons for overuse were performance of colonoscopy less than 6 months after negative FOBT (35 %) and performance of colonoscopy less than 9 years after prior colonoscopy (31 %) (Table 3). Substantial and significant variation in colonoscopy overuse was noted between facilities and between VISNs, with a nearly eightfold difference between the maximum and minimum rates of overuse at the facility level and a nearly twofold difference at the VISN level (Figs. 1 and 2). Furthermore, overuse at the VISN and facility level was relatively stable over time. Between FY11 and FY13, all VISNs in the bottom quartile of performance (high overuse) remained in the same quartile. Similarly, all VISNs in the top quartile of performance (low overuse) remained there. Likewise, during this time period, 70 of 122 facilities remained in the same quartile of performance for all 3 years, and 103 of 122 facilities improved or worsened by no more than one quartile. Of the 27 facilities with high overuse, all remained in the bottom quartile for the entire 3-year period.

In multivariable analysis adjusting for patient characteristics (age, gender, and Charlson comorbidity index), we examined several facility-level variables as potential predictors of overuse, including: (1) proportion of screen-eligible patients who were “up to date” for screening according to current guidelines; (2) median number of days between positive fecal occult blood test (FOBT) and colonoscopy (“FOBT wait time”); (3) complexity score; (4) academic affiliation; (5) number of colonoscopies performed in FY13; and, (6) proportion of colonoscopies “out-

**Table 2. Baseline Characteristics of Patients (N=88,754)**

Patient Characteristic	Participants, n (%)
Age (years)	
< 40	1717 (1.9)
40–49	3794 (4.3)
50–75	80,709 (90.9)
76–85	2360 (2.7)
> 85	172 (0.2)
Median age (IQR)	62 (54–66)
Gender	
Male	82,383 (92.8)
Female	6361 (7.2)
Charlson comorbidity index	
0	41,297 (46.5)
1–3	40,420 (45.5)
≥ 4	7037 (7.9)

sourced” to non-VA facilities. None of these factors was independently associated with overuse of screening colonoscopy.

**DISCUSSION**

Colorectal cancer screening is a widely recommended preventive service that has traditionally been underused. But in large integrated healthcare systems like VA, systematic efforts to increase screening rates have been successful.<sup>21</sup> In this context, we used an expert workgroup process to develop and test an electronic measure of screening colonoscopy overuse. Of nearly 90,000 screening colonoscopies performed in VA in FY13, 23 % met a consensus definition for overuse. Common reasons for overuse were colonoscopy being performed soon after negative FOBT (35 %), colonoscopy being repeated less than 9 years after prior negative screening colonoscopy (31 %), and colonoscopy being performed in a patient under 50 years of age (17 %).

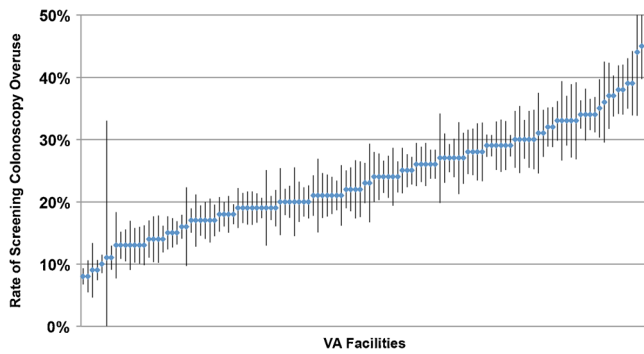
While our electronic measure had low sensitivity (reflecting limitations in our ability to electronically ascertain screening indication), the measure had very high specificity. Thus, while the measure failed to identify all possible overuse, the likelihood of “false positive” categorization of overuse was low. Furthermore, rates of overuse were stable over time at the facility and VISN levels, suggesting that our electronic measure reliably reflects underlying organizational characteristics. Nonetheless, it is possible that some of the colonoscopies that the measure categorized as “overuse” were, in fact, appropriate. For example,

**Table 1. Baseline Characteristics of VA Facilities (N=122)**

Facility Characteristic	% or Median (IQR)
Academic affiliation (%)	93
Complexity score (%)	
High	59
Medium	14
Low	27
Rate of overuse (%)	23 (18–29)
Screening colonoscopies performed (#)	1,693 (1,150–2,509)
Patients “up to date” for screening (%)	82 (80–85)
Time between (+) FOBT and colonoscopy (days)	54 (42–75)
Colonoscopies “outsourced” to non-VA sites (%)	10 (0–38)

**Table 3. Reasons for Overuse of Screening Colonoscopy Identified by Electronic Measure (N=20,530)**

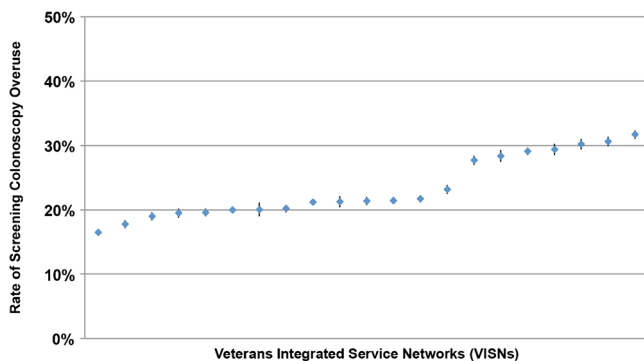
	Definition of Overuse	Number of Colonoscopies (%)
Probable Overuse	Less than 6 Months After Negative FOBT	6904 (35)
	Less than 9 Years After Negative Colonoscopy	6439 (31)
	Age < 40	1579 (8)
Possible Overuse	Age > 85	141(1)
	Life Expectancy < 6 Months	142 (1)
	Age 40–49	3462 (17)
	Age 76–85	1,863 (9)



**Figure 1. Overuse of screening colonoscopy across 122 VA facilities (N=88,754). Each marker represents a single VA facility, with error bars indicating 95% confidence intervals (median overuse = 23%, interquartile range = 18% to 29%).**

some procedures could have been indicated based on signs or symptoms that were not adequately documented in the CDW, or based upon prior incomplete colonoscopy (i.e., failure to reach the cecum or achieve adequate bowel preparation). Additionally, colonoscopies performed in individuals under 50 years of age could be considered appropriate in African Americans, who may be at increased risk for colorectal cancer, though this recommendation is not included in USPSTF or VA screening guidelines on this topic. Despite these potential limitations, our electronic measure provides a tool that, coupled with existing underuse measures, could be used to assess and improve the appropriateness of colonoscopy utilization. In current practice, measures of underuse are frequently “unopposed” by corresponding measures of overuse. Such lack of balance can inadvertently promote overutilization. As meaningful use of electronic health records becomes more widespread, limitations in ascertaining colonoscopy indication will improve, thereby improving the sensitivity of electronic measures of overuse. Along these lines, VA is working to develop a standardized, structured system for colonoscopy documentation, including procedure indication and findings.

Existing and proposed measures of colorectal cancer screening overuse have focused on two key areas: (1) screening colonoscopy performed less than 10 years after a negative



**Figure 2. Overuse of screening colonoscopy across 21 Veterans Integrated Service Networks (VISNs) (N=88,754). Each marker represents a single VISN, with error bars indicating 95% confidence intervals (median overuse = 21%, interquartile range = 20% to 28%).**

colonoscopy (a CMS Physician Quality Reporting System [PQRS] measure adopted in 2013),<sup>22</sup> and, (2) screening colonoscopy performed in an individual over 85 years of age (proposed by both gastroenterology professional societies and the National Committee for Quality Assurance).<sup>23</sup> As our work demonstrates, these individual measures (which are components of our composite measure) have the potential to be operationalized electronically using administrative data with acceptable reliability. Our work also demonstrates that within VA, screening colonoscopy is rarely overused in the very elderly, with only 1% of overuse being due to screening in those over age 85. Thus, focusing on this population is likely to have limited value in improving quality of care.

Our study has several strengths that should be highlighted. First, we successfully used an expert workgroup process to develop a high-specificity electronic measure of overuse, one that demonstrated stability over time across facilities and VISNs. Because this measure is electronically implementable and stable over time, it could be used to identify facilities with high rates of overuse and subsequently be used to monitor these facilities for performance improvement. At the same time, existing measures of colorectal cancer screening underuse could be monitored to ensure that efforts to reduce overuse are not being accompanied by underuse of screening. Second, we used data from the largest integrated healthcare system in the United States, analyzing data on nearly 90,000 patients per year across 122 facilities. The use of VA data allowed us to examine overuse in a healthcare system with a long-standing, comprehensive, and highly effective implementation of underuse measures of colorectal cancer screening. Because performance of colorectal cancer screening is now a standard measure in many healthcare systems, it is likely that similar patterns of overuse can be found in other high-performing healthcare systems. Moreover, our measure could be used to identify and monitor overuse in other healthcare systems. For example, in healthcare systems with greater standardization of electronic data (e.g., Kaiser Permanente), our measure is likely to be more accurate than in VA.

Our study also has limitations. Accurate coding of colonoscopy indication is critical to any effort to assess colonoscopy quality. Because VA electronic data have limited accuracy for colonoscopy screening indication, we were unable to develop an electronic measure that was highly sensitive and specific for overuse. A less specific but more sensitive measure could have been developed with existing data, potentially providing more accurate estimates of the absolute volume of screening colonoscopy overuse across the healthcare system, but such a measure could promote screening underuse in the context of performance measurement. It is also possible that coding practices differ systematically between VA sites, a phenomenon that could at least partly explain the variation noted across facilities. Prior work suggests that colonoscopy indication may be more accurately coded in fee-for-service Medicare.<sup>15</sup> Some sites may also already have systems in place to minimize overuse (e.g., electronic assessment of prior colonoscopy or FOBT as part of the colonoscopy ordering template), another potential cause of

variation. Future work will be needed to better understand the underlying causes of overuse. It is also important to note that Veterans may undergo procedures outside VA, meaning that our data (which did not capture use of non-VA care) may underestimate true rates of overuse among Veterans. As well, some of our findings may be specific to the VA healthcare context. In non-VA settings, for instance, overuse of colonoscopy within 6 months of a negative FOBT (the most common reason for overuse in our study) is less likely, since FOBT is used less frequently for screening. Finally, our measure was relatively conservative in its use of health status and life expectancy, requiring a life expectancy of less than 6 months to meet a definition of overuse. While this strict definition maximizes specificity, it does so at the expense of sensitivity. For example, nearly 8 % of patients in our cohort had a Charlson comorbidity index  $\geq 4$ , indicating poor health and limited life expectancy. In such patients, screening is unlikely to be of benefit and may even be harmful, indicating overuse.<sup>24</sup> Yet, the vast majority of these patients have a life expectancy of more than 6 months, meaning that they would not have met our measure definition of overuse. The development of more clinically meaningful and patient-centered measures that take into account health status and patient preference is a topic of ongoing work.

In summary, our results suggest that screening colonoscopy is overused in the VA Health Care System, and that rates of overuse vary widely across VA facilities and VISNs. The electronic measure we developed and tested demonstrated high specificity and reliability for facility-level measurement, successfully stratifying facilities according to performance. The stability of facility performance over time supports the hypothesis that the measure is capturing true organizational characteristics and/or provider practices. In the future, this measure, coupled with the existing measure of colorectal cancer screening underuse, could be used to guide quality improvement in large health systems.

---

**Corresponding Author:** Sameer D. Saini, MD, MS; Veterans Affairs Center for Clinical Management Research (CCMR), 2215 Fuller Road, 111D, Ann Arbor, MI 48105, USA (e-mail: sdsaini@umich.edu).

**Prior Presentations:** This work was presented in oral form at the 2015 VA HSR&D/QUERI National Conference in Philadelphia, PA, on 9 July 2015.

#### Compliance with Ethical Standards:

**Funding Source:** This work was funded by VA Health Services Research and Development (CDA 09-213-2 and RRP 12-184). Partial funding for this work was provided by the Veterans Health Administration (VHA) Office of Informatics and Analytics (OIA). These funding agencies had no role in the design and conduct of the study. VHA OIA reviewed the manuscript prior to submission. The opinions expressed in this paper are of the authors and do not necessarily reflect those of the Department of Veterans Affairs.

**Conflicts of Interest:** Sameer D. Saini – None; Adam A. Powell – None; Jason A. Dominitz – None; Deborah A. Fisher – None; Joseph Francis – None; Linda Kinsinger – None; Kathleen S. Pittman – None; Philip Schoenfeld – None; Stephanie E. Moser – None; Sandeep Vijan – None; Eve A. Kerr – None

## REFERENCES

1. Chan KS, Chang E, Nassery N, Chang H-Y, Segal JB. The state of overuse measurement: a critical review. *Med Care Res Rev MCRR*. 2013;70(5):473–96.
2. Kerr EA, Fleming B. Making performance indicators work: experiences of US Veterans Health Administration. *BMJ*. 2007;335(7627):971–3.
3. Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *Jama*. 2012;307(17):1801–2.
4. Powell AA, White KM, Partin MR, et al. Unintended consequences of implementing a national performance measurement system into local practice. *J Gen Intern Med*. 2012;27(4):405–12.
5. Kerr EA, Lucatorro MA, Holleman R, et al. Monitoring performance for blood pressure management among patients with diabetes mellitus: too much of a good thing? *Arch Intern Med*. 2012;1–8.
6. Beard AJ, Hofer TP, Downs JR, et al. Assessing appropriateness of lipid management among patients with diabetes mellitus: moving from target to treatment. *Circ Cardiovasc Qual Outcomes*. 2013;6(1):66–74.
7. 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(20):2466–70.
8. Mathias JS, Baker DW. Developing quality measures to address overuse. *JAMA*. 2013;309(18):1897–8.
9. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2008;149(9):627–37.
10. Sheffield KM, Han Y, Kuo YF, Riall TS, Goodwin JS. Potentially inappropriate screening colonoscopy in Medicare patients: variation by physician and geographic region. *JAMA Intern Med*. 2013;1–9.
11. Powell AA, Saini SD, Breitenstein MK, et al. Rates and correlates of potentially inappropriate colorectal cancer screening in the veterans health administration. *J Gen Intern Med*. 2015;30(6):732–741.
12. Saini SD, Vijan S, Schoenfeld P, Powell AA, Moser S, Kerr EA. Role of quality measurement in inappropriate use of screening for colorectal cancer: retrospective cohort study. *BMJ*. 2014;348:g1247.
13. Warren JL, Klabunde CN, Mariotto AB, et al. Adverse events after outpatient colonoscopy in the Medicare population. *Ann Intern Med*. 2009;150(12):849–57. W152.
14. Fisher DA, Grubber JM, Castor JM, Coffman CJ. Ascertainment of colonoscopy indication using administrative data. *Dig Sci*. 2010;55(6):1721–5.
15. Ko CW, Dominitz JA, Neradilek M, et al. Determination of colonoscopy indication from administrative claims data. *Med Care*. 2014;52(4):e21–9.
16. 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(3):844–57.
17. Ko CW, Dominitz JA, Green P, Kreuter W, Baldwin L-M. Accuracy of Medicare claims for identifying findings and procedures performed during colonoscopy. *Gastrointest Endosc*. 2011;73(3):447–53.
18. Johnson DA, Barkun AN, Cohen LB, et al. Optimizing adequacy of bowel cleansing for colonoscopy: recommendations from the US multi-society task force on colorectal cancer. *Gastroenterology*. 2014;147(4):903–24.
19. Carney BT, West P, Neily J, Mills PD, Bagian JP. The effect of facility complexity on perceptions of safety climate in the operating room: size matters. *Am J Med Qual Off J Am Coll Med Qual*. 2010;25(6):457–61.
20. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol*. 1994;47(11):1245–51.
21. Trivedi AN, Grebla RC. Quality and equity of care in the Veterans Affairs Healthcare System and in Medicare Advantage health plans. *Med Care*. 2011;49(6):560–8.
22. 2013 Physician Quality Reporting System (PQRS) Measures List. [http://www.cms.gov/apps/ama/license.asp?file=/PQRS/downloads/2013\\_PQRS\\_MeasuresList\\_ImplementationGuide\\_12192012.zip](http://www.cms.gov/apps/ama/license.asp?file=/PQRS/downloads/2013_PQRS_MeasuresList_ImplementationGuide_12192012.zip). Accessed June 15, 2015.
23. Proposed New Measures for HEDIS 2015: Colorectal and Prostate Cancer Appropriateness / Overuse Measures. [http://www.ncca.gov/Portals/0/HomePage/Overuse\\_Materials.pdf](http://www.ncca.gov/Portals/0/HomePage/Overuse_Materials.pdf). Accessed June 15, 2015.
24. Gross CP, Soulos PR, Ross JS, et al. Assessing the impact of screening colonoscopy on mortality in the Medicare population. *J Gen Intern Med*. 2011;26(12):1441–9.

APPENDIX

Expert Workgroup Process

A performance measure of screening colonoscopy overuse was specified by an expert workgroup comprising VA experts in colorectal cancer screening (DF, PS, JD, KP) and performance measurement (EK, JF, LK). Prior to the initial workgroup meeting, members were provided with relevant literature for review, including the 2008 U.S. Preventive Services Task Force (USPSTF) guideline for colorectal cancer screening and the 2007 Veterans Health Administration (VHA) Colorectal Cancer Screening Directive (VHA Directive 2007-004). The members were charged with developing a measure that: (1) was based on high-quality evidence (i.e., the evidence summarized in the 2008 USPSTF guideline); (2) maximized specificity at the expense of sensitivity; and, (3) could be implemented electronically.

The workgroup began by broadly defining overuse of screening colonoscopy as a screening colonoscopy performed at an inappropriately short interval (e.g., screening colonoscopy performed 5 years after a prior negative screening colonoscopy in an average-risk patient) or in a patient for whom the benefit of screening is low (e.g., screening colonoscopy in an 86 year-old). The workgroup was then asked to more clearly specify the data elements that comprised the measure denominator (the eligible population—Text Box 1A) and the numerator (those meeting the measure—Text Box 1B). For the numerator, discussion focused on the appropriate cutoffs for age, time interval, and life expectancy. The workgroup then classified each item comprising the numerator as “probable” or “possible” overuse based on the strength of the guideline recommendation and the likelihood of misclassification. The workgroup also specified exclusions from the measure denominator (Text Box 1A)—factors indicating that the colonoscopy was not an average-risk screening procedure—such as increased risk for colorectal cancer (e.g., screening colonoscopy in a patient with a family history of colorectal cancer) or ineligibility for screening (e.g., prior total abdominal colectomy).

The workgroup met three times over a 6-month period to define the measures and review and provide feedback on preliminary results of electronic data extraction. Rather than using a formal Delphi process, any potential disagreement was resolved through discussion. Ultimately, consensus was achieved on all elements of the measure specification. The final measure defined by the workgroup identified average-risk screening colonoscopies (comprising the denominator, with exclusions—Text Box 1A) that met one or more criteria for probable or possible overuse (comprising the numerator—Text Box 1B).

Electronic Health Record Abstraction

Manual record review was performed by West Virginia Medical Institute (WVMI), a professional chart abstraction group that performs large-scale, national chart reviews for the VA performance measurement program (External Peer Review

Program, or EPRP) on an ongoing basis. EPRP uses quality control processes to maximize consistency and completeness of data collected from VA records. These processes include: (1) internal quality control (IQC) question and mnemonic level analysis; and (2) item-level inter-rater reliability assessment. For the existing VA colorectal cancer screening underuse measure, WVMI record review has been shown to have high reliability. For example, in 2015, IQC question and mnemonic level analysis did not identify any outliers associated with this measure. During this period, the median question level agreement rate was 95 %.

Using a process similar to that used for the colorectal cancer screening underuse measure, we developed a standardized electronic health record (EHR) abstraction algorithm to identify overuse measure elements in manual record review. This abstraction algorithm was developed in collaboration with WVMI and EPRP. We first outlined the data elements that would be needed from manual record review to calculate the measure. We then determined which potential data sources could be accessed to retrieve each of these elements (e.g., endoscopy report, primary care clinic note, laboratory data). We also developed a lexicon of potential findings in each of these data sources (e.g., for colonoscopy indication, findings could include average-risk screening, high-risk screening, surveillance, or diagnostic). This process was iterative and collaborative, conducted through a combination of electronic communication and two conference calls. The a priori data elements, data sources, and potential findings were combined into a standardized algorithm for record reviewers. The algorithm was refined and then piloted on a sample of 20 patients to identify areas for further refinement. The final algorithm was used to review the records of 3000 patients.

Electronic Approximation of Measure Denominator

To approximate the measure denominator (Text Box 1A), we first identified all colonoscopies performed in FY13 using Current Procedural Terminology (CPT) and Healthcare Common Procedure Coding System (HCPCS) codes (Appendix Table 4). We then excluded procedures that may have been performed for diagnostic or high-risk screening or surveillance indications, using an approach previously developed and val-

Table 4. Codes Used to Identify Colorectal Cancer Screening Tests

Colonoscopy	
CPT codes	44388, 44389, 44390, 44391, 44392, 44393, 44394, 44397, 45378, 45379, 45380, 45381, 45382, 45383, 45384, 45385
HCPCS codes	G0105, G0121
FOBT	
LOINC codes	2335-8, 12503-9, 12504-7, 14563-1, 14564-9, 14565-6, 27396-1, 27401-9, 27925-7, 27926-5, 29771-3, 56490-6, 56491-4, 57905-2, 58453-2

CPT = Current Procedural Terminology; HCPCS = Healthcare Common Procedure Coding System; FOBT = fecal occult blood test; LOINC = Logical Observation Identifiers Names and Codes

idated by Fisher and colleagues.<sup>1</sup> Specifically, we excluded patients who had an ICD-9 code for specific gastrointestinal symptoms or for colorectal neoplasia within 12 months of the FY13 colonoscopy (Appendix Table 5). We also excluded all individuals who had a colonoscopy performed within 12 months prior to the FY13 colonoscopy (Appendix Table 4).

To further increase specificity of the electronic measure, we also excluded individuals with codes indicating high risk for colorectal cancer or prior total abdominal colectomy at any time in the prior 14 years (from FY00 to FY13) (Appendix Table 6). Specifically, patients were excluded if CPT or ICD-9 codes revealed any of the following diagnoses between FY00

**Table 5. Codes Used to Exclude Colonoscopies Performed for Non-Screening Indications<sup>1</sup>**

ICD-9	Description
787.3, 789.0, 789.3	Abdominal distention, pain, swelling
280.x	Iron-deficiency anemia
285.1, 285.9	Anemia
783.0	Anorexia
560.9	Bowel obstruction
787.9	Change in bowel habits
564.0	Constipation
558.9, 564.5	Diarrhea
787.6	Fecal incontinence
578.x	Gastrointestinal bleed
792.1	Heme-positive stool
569.3	Hemorrhage of rectum/anus
787.0	Nausea/vomiting
783.2	Weight loss
555.x	Crohn's disease
556.x	Ulcerative colitis
V100.5, V100.6	History of colorectal cancer
V127.2	History of colon polyps
562.12	Diverticulitis with hemorrhage
562.13	Diverticulosis with hemorrhage
569.85	Angiodysplasia with hemorrhage
558.1	Colitis - radiation
560.0	Intussusception
560.1	Paralytic ileus
564.7	Megacolon
564.1	Irritable colon

1. Fisher DA, Grubber JM, Castor JM, Coffman CJ. Ascertainment of colonoscopy indication using administrative data. *Dig Sci.* 2010;55(6):1721-1725. doi:10.1007/s10620-010-1200-y.

**Table 6. Additional Codes Used to Identify Patients at Increased Risk for Colorectal Cancer**

Diagnosis	Code
Colectomy (CPT)	44150 - 44156, 44210 - 44212
Colorectal cancer (ICD-9)	153.x, 154.0, 154.1, 154.8, V10.0, V10.05, V10.06
Colorectal cancer (HCPCS)	G0105, G0213, G0214, G0215, G0231
Colon polyps (ICD-9)	211.3, 211.4, 230.3, 230.4, V12.72
Inflammatory bowel disease (ICD-9)	555.x, 556.x
Family history of colorectal cancer (ICD-9)	V16.0, V18.51

CPT = Current Procedural Terminology; ICD-9 = International Classification of Disease, Ninth Revision; HCPCS = Healthcare Common Procedure Coding System

and the qualifying FY13 colonoscopy (Appendix Table 6): (1) prior colectomy; (2) history of colorectal cancer; (3) history of colon polyps; (4) history of inflammatory bowel disease; or (5) family history of colorectal cancer. These additional exclusion criteria were selected to ensure that the cohort comprised individuals who were at average (rather than increased) risk for CRC. Finally, we excluded individuals who underwent their FY13 colonoscopy during a hospitalization (since such colonoscopies are unlikely to be performed for screening).

### Electronic Approximation of Measure Numerator

Specification of electronic elements comprising the measure numerator (probable and possible overuse—Text Box 1B) was less complicated than for the denominator, since these elements were primarily based on factors such as patient age and the time interval between colonoscopies. To identify fecal occult blood tests (FOBTs), we used Logical Observation Identifiers Names and Codes (LOINC) (Appendix Table 4). To identify patients with life expectancy < 6 months, we used structured data from CDW that is used to indicate limited life expectancy for clinical purposes.