3510: Screening/Surveillance Colonoscopy

https://p4qm.org/measures/3510

The American Medical Association (AMA) is writing to express our concerns that end users will not be able to make meaningful distinctions on the costs associated with screening and surveillance colonoscopies due the measure's minimal variation, the lack of any correlation of this cost measure with quality measures, and the omission of any testing to determine whether socio-economic factors should be included in the risk adjustment model. As currently defined and used, the measures could reward a clinician or group that fails to screen the highest risk patients and/or fails to identify and remove precancerous lesions from the patients they do screen.

We are also extremely concerned to see the staff preliminary assessment estimate that: "The measure's financial impact is material. For groups, if all entities performed at the third decile the reduction in total spending would be $139 million (-13.0%). For clinicians, if all entities performed at the third decile the reduction in total spending would be $106 million (-10.8%)." No rationale was provided to explain why the third decile was selected as optimal and achievable by clinicians and we are opposed to these types of arbitrary assessments being made to guide the committee's evaluation.

Problems with the Measure Definition

It should not be assumed that every clinician and group can achieve the lowest possible cost, and cost measures should provide sufficient information to enable an assessment of whether costs are reasonable and appropriate. In Section 6.2.4, the measure developer states that there is a "meaningful performance gap" because observed episode costs differ by $838, between the $730 cost at the 10th percentile and $1,568 at the 90th percentile. (The reference for this is Table 1, but the numbers in Table 1 differ from those cited in Section 6.2.4.)

Based on the data submitted for the performance gap, there is roughly an $800 difference between the 10th to 90th percentiles with only 2 individual clinicians and practices with costs above or below those deciles. We do not believe that presenting costs in this way is useful and these small differences may be reflective of patients' preferences or needs. For example, a patient's risk profile or preference could influence whether anesthesia is administered and practices may use different bowel prep types to reduce a patient's out of pocket costs that may infrequently lead to inadequate bowel prep.

Outside of patient preferences, the largest cause of variation in the cost of colonoscopy, and the largest opportunity for reducing the cost of colonoscopy, is the site of service, i.e., whether the colonoscopy is performed in a HOPD, ASC or a physician's office. As noted in Section 6.2.4, the mean episode cost for HOPD episodes was $1,446, the mean for ASC episodes was $976, and the mean for office-based episodes was $606. In sum, every episode that could be performed in a physician's office instead of an HOPD would reduce Medicare spending by more

than $800, and every episode that could be performed in an ASC instead of a hospital would reduce spending by nearly $500. However, under this measure, episodes are evaluated separately by site of service, so any savings from using a lower cost site does not result in a lower episode cost score for the clinician or group being evaluated.

In addition, the measure does not include the cost of return visits to urgent care centers or physician offices that are made to evaluate or treat complications of the colonoscopy. These visits can involve the same types of complications and services that are addressed in a subset of ED visits, so excluding them underestimates the true complication rate for colonoscopies. Patients from rural communities who travel a long distance to an ASC for a procedure will be more likely to visit an emergency department for treatment of a complication than patients who live in an urban area where they could more easily visit the physician who performed the procedure or an urgent care center. Consequently**, physicians with larger numbers of rural patients could have higher rates of ED visits on this measure and inappropriately appear to have higher complication rates.**

Furthermore, no analysis is provided showing which specific types of services and costs other than the site of service contribute to variation in costs and whether those services and costs could be reduced by physicians without reducing the effectiveness of the procedure. In the Measure Rationale (Section 1.10), the developer identifies two different factors that can cause "meaningful cost variation within the 14-day episode window" that is "attributable to clinician practice patterns" – bowel preparation quality and use of anesthesiologist services. However, there is no analysis of the extent to which those factors cause the observed variation in cost.

- The measure does nothing to directly identify the quality of bowel preparation or whether a colonoscopy was unable to be completed. The developer states that inadequate bowel preparation is associated with an increased likelihood of an incomplete procedure, but a colonoscopy billed with an incomplete procedure modifier would reduce the physician fee by 50% and thereby lower the episode cost. The developer further states that incomplete procedures due to inadequate bowel preparation require "a repeat examination within the episode window…adding directly to episode costs." However, a repeat colonoscopy could be performed more than 14 days after the incomplete procedure (if it is performed at all), and if so, it would not be counted toward the episode cost; it would instead be counted as a new episode. This creates a perverse incentive to delay a repeat colonoscopy rather than performing it as soon as possible. **The measure developer should provide an analysis of how often low episode costs are due to incomplete procedures, the proportion of episodes in which a second colonoscopy is performed, and how often a second colonoscopy occurs more than 14 days after the initial colonoscopy. The measure could be modified to extend the episode window specifically for repeat colonoscopies to avoid penalizing physicians for conducting repeat colonoscopies promptly.**

- There is no analysis of the extent to which lower and higher episode costs are due to differences in the use of anesthesiology services**. It would be better to revise the measure to tabulate costs of anesthesia services separately rather than combining them with the cost of complications in a single measure of total episode cost; separating the costs would make it easier to determine which colonoscopy teams have higher costs due to their approach to anesthesia and which have higher costs because they have higher complication rates or other factors.**

Lack of Cost and Quality Connection

The AMA strongly supports the tenet that cost must be assessed within the context of the quality of care provided; yet, the developer did not demonstrate that this measure correlates to any quality measure within the Merit-based Incentive Payment System (MIPS) program even when there is an outcome measure (GIQuIC26 Screening Colonoscopy Adenoma Detection Rate) with which this correlation would be optimal. We are very troubled that the testing did not include this assessment as it would provide more meaningful information regarding the validity of this cost measure. **Battelle guidance on evaluating cost measures includes an expectation that the developer would analyze and interpret the relationship between cost performance and quality of care in the usability criterion. We support this approach and believe that it would demonstrate how the use of this measure with GIQuIC26 Screening Colonoscopy Adenoma Detection Rate can inform quality improvements efforts and patient decision-making.** Based on our review of the submission, we do not see any discussion or analysis of this important relationship, nor does it even discuss whether there has been any effect on costs throughout the use of this measure, which has been in use for several years in MIPS.

In addition, we believe that only cost and not clinical administrative claims data were used to complete the correlation to repeat colonoscopies and adverse events. Solely relying on cost data to represent quality is inherently flawed.

Problems with Risk Adjustment

1. Excessive number of variables in the risk adjustment model

Section 5.4.2 indicates that the Gastrointestinal Disease Management Medical and Surgical Clinical Subcommittee that was consulted in the development of the measure identified 8 specific variables that they believed could increase colonoscopy episode cost in ways outside the attributed clinician's control. The risk adjustment model includes these variables, but it also includes 103 other variables without providing any clinical justification for doing so. Including hierarchical condition category variables in the model that have no clinical relationship to the cost of colonoscopy episodes can inappropriately increase or decrease the expected cost of individual episodes, which then inappropriately penalizes or rewards physicians who happen to have patients with those characteristics.

For example, according to the risk adjustment model used for the measure, the expected cost of colonoscopy episodes in hospitals and ambulatory surgery centers for patients with macular

degeneration is higher than the cost for patients without macular degeneration, even though there is no clinical reason why this should be so. The regression coefficients for the macular degeneration variables are likely driven by spurious correlations with unmeasured factors. On the other hand, the risk adjustment model predicts that the cost of colonoscopy episodes performed in physician offices should be lower for patients with macular degeneration. Inclusion of this variable in the risk adjustment model means that the cost of a colonoscopy on a patient with macular degeneration will appear to cost over $40 less in a hospital outpatient department than in a physician's office, even though the actual cost would be over $700 higher.

Including 111 variables in a risk adjustment model does not make it "better" than a model with only 8 variables if the additional variables have no clinical relationship to the cost of an episode. In fact, inclusion of so many unrelated variables makes it impossible for a physician to understand why they received the cost score they did and to know whether and how they could improve. **The developers should be required to provide an analysis of how well a risk adjustment model using only the variables identified by clinicians would perform and to provide a clinical and statistical justification for any additional variables that are included.**

2. Failure to adjust for differences in site of service mix by geography

The risk adjustment coefficients are estimated separately for colonoscopies performed in HOPDs, ASCs, and physician offices. The justification for this is that higher risk patients are treated in hospitals than other settings, but this is not necessarily true, particularly in areas where there are few or no ambulatory surgery centers. The availability of ASCs varies significantly across communities, which means the proportion of all colonoscopies performed in hospitals will also vary significantly. For example, there are fewer ASCs in states with Certificate of Need laws, and there are fewer ASCs in small communities and rural areas simply because of the smaller number of patients. As a result, hospitals in states and parts of states where there are more ASCs will have a smaller and higher-risk group of colonoscopy patients than other hospitals do, and so they will likely have higher rates of unplanned hospital visits after the procedure. **These differences will not be fully captured by the patient characteristics included in the risk model, and the measure does not adjust for the proportion of colonoscopies performed in an HOPD, so HOPDs that perform a smaller proportion of colonoscopies could inappropriately appear to be delivering lower-quality care.**

The separate models for each site of service leads to problematic results. For example, the risk adjustment model also predicts that the cost of a colonoscopy episode will be lower for a patient with severe pressure ulcers if the colonoscopy is performed in an ambulatory surgery center, but higher if it is performed in a hospital outpatient department. There is no clinical reason why this should be true, but the practical effect of including this variable and estimating separate coefficients is that a physician would be penalized for performing a colonoscopy on a patient with a history of pressure ulcers in an ASC rather than a more expensive HOPD, even if the procedure can be safely performed in both settings.

The developer should be required to estimate a single regression model with a variable based on the proportion of colonoscopies performed in HOPDs in the county of residence, and to compare that model with the current approach.

3. Failure to adjust for health related social needs

We question the lack of any testing or potential inclusion of socio-economic risk factors in the risk adjustment. The omission is inconsistent with established evidence that socio-economic status, clinical complexity, and site-of-service variables materially influence both cost and outcomes in colonoscopy.[1] The traditional approach of risk adjusting at the patient level may not be appropriate for measures where the measurement period includes care that is outside of the control of the clinician. We believe that there may be community-level variables that affect the risk of an adverse event during the days following a colonoscopy but are not currently addressed. Measures that extend beyond the procedure or outside the locus of control of the measured entity should continue to have socio-economic adjustments addressed and analyzed at different levels (e.g., patient, clinician, and community).

In addition, patients with limited access to primary care are more likely to visit EDs for problems rather than physician offices, and they are more likely to have hospital admissions for exacerbations of chronic conditions than other patients. As a result, the likelihood of repeat colonoscopies and ED visits will be higher for patients with these characteristics. Yet the risk adjustment model for this measure does not include any variables to identify these patients and adjust for their higher expected costs.

The developer says that social and environmental factors "were considered" for inclusion in the model but were not included for several reasons:

- First, it is asserted (without evidence) that "pathways through which social and environmental factors are most strongly associated with elevated healthcare costs, including emergency department utilization, readmissions for unrelated conditions, and chronic disease complications, fall outside this episode scope." Yet one of the justifications the developers use for the measure is that patients who are more likely to have poor bowel preparation are also more likely to have repeat colonoscopies within the episode window. In addition, patients with severe or poorly managed chronic conditions are more likely to have ED visits and hospital admissions for exacerbations of those conditions than other patients with the same chronic conditions, and these exacerbations and visits can occur at any time, including during the two weeks after an outpatient procedure. Consequently, it is likely that costs within the 14 day episode window will be higher for patients with higher levels of social needs.
- Second, it is asserted that the CMS HCC risk model and measure-specific clinical adjustors used in the current risk adjustment model "capture patient clinical complexity

[1] Siddique SM, Mehta SJ. Holding Gastroenterologists Accountable for Colonoscopy Through MACRA Episode-Based Cost Measure. Clin Gastroenterol Hepatol. 2019 May;17(6):1015-1018. doi: 10.1016/j.cgh.2019.01.009. Epub 2019 Jan 11. PMID: 30641144; PMCID: PMC6475480.

that is substantially collinear with social risk markers, limiting the marginal predictive contribution of adding social risk variables to the model." "Substantially collinear" is not the same as "perfectly correlated." The fact that patients with social risk characteristics are more likely to have health problems is not a justification for excluding those characteristics from the model. Moreover, **the impact of including social risk factors should be assessed after removing the majority of the HCC variables, which have no clinical relationship with colonoscopy costs.**

- Third, it is asserted that "patient selection for elective screening procedures, in combination with the clinical exclusion criteria applied to this episode group, reduces the representation of the highest social-risk beneficiaries in the attributed episode population." Even if the "highest social-risk" beneficiaries are excluded, that does not mean that the remaining beneficiaries have no social risk factors.
- Finally, it is asserted that excluding social risk factors is "consistent with prior research on MIPS episode-based cost measures, which found that adjustment for social risk factors does not meaningfully change clinician or group performance on these measures." In fact, the cited study found that the colonoscopy cost percentile for many physicians did change significantly after including social risk factors in the model. The fact that performance did not change significantly for the majority of physicians does not justify failing to make the adjustment for those who are significantly affected.

The developer should be required to provide a detailed analysis of the effects of including social risk factors.

Poor Assessment of Validity

The methodology described in Section 5.3.3 for validity testing is inaccurate and inadequate. In Section 5.3.3, the developer states that a "repeat colonoscopy … within the episode window is a direct indicator of potential overutilization; reduction through adherence to evidence-based screening intervals is an actionable improvement lever," and in Section 5.3.5, it states that "higher rates of repeat colonoscopy within the episode window [are] … inconsistent with evidence-based screening intervals." However, if a repeat colonoscopy occurs within the short 14-day episode window for this measure, it would not represent "overutilization," but rather a desirable response to poor bowel preparation or a problem that prevented completion of the initial colonoscopy, not "overutilization." In contrast, **true overutilization would be performing a screening colonoscopy every year for a patient who is not at high risk of colon cancer, but this measure cannot identify that, because it only looks at a 14 day window of time.** Even if a physician performed a colonoscopy on the same patient every month, it would not increase their average episode cost, it would merely increase the number of episodes for that physician, even though it would represent a very high level of overutilization.

The developer cites the positive correlation between the episode cost and the rate of events such as ED visits and complications as evidence of validity, but this is not an independent test of the validity of this measure; it is a tautological result of the way the measure is defined. If an ED visit or a complication occurs within the 14 day episode window, the cost of that visit or complication is included in the observed cost of the episode, so by definition, there will be a

correlation between the frequency of such events in a set of episodes and the average cost of the episodes. What the developers fail to explain is why the measured correlation is so low, i.e., what other factors are causing episode costs to be high when the rate of ED visits and complications is low? The low correlation may well indicate that the measure is being affected most heavily by factors that are not valid indicators of inefficiency or poor quality care.

An AMA analysis of the 2024 cost scores on this measure found that the distribution of the scores varies far more by state than is reasonable to expect. As shown in the attached tables, 38% of the physicians in Nebraska who were scored on the colonoscopy measure had a cost score below 5.0, 79% had a cost score below 7.5 (the national median) and only 2% had a cost score above 9.0, whereas none of the physicians in Montana who were scored on the measure had a cost score below 5.0 and 46% had a cost score above 9.0. The cost measure is only valid if it is truly the case that more one-third of the physicians scored on the colonoscopy measure in Nebraska had higher than expected costs but almost half of those in Montana had lower than expected cost. **The developer should be required to examine these distributions to determine whether they are truly due to differences in costs that are under the control of the physicians or whether they are due to problematic factors such as those described earlier.**

Colonoscopy Scores by State				
State	Median Score	% < 5	% <7.5	% > 9
NE	7.0	38.3%	79.4%	2.0%
TX	6.3	20.7%	79.8%	13.1%
CO	7.3	9.6%	51.5%	0.2%
NY	7.7	9.2%	41.0%	3.8%
SC	7.1	8.3%	89.9%	3.1%
IA	7.6	6.2%	42.6%	12.9%
NH	6.2	5.9%	82.4%	0.0%
WA	7.6	4.0%	26.5%	0.7%
MD	7.4	3.7%	81.3%	8.0%
KS	7.5	2.4%	38.9%	3.9%
WI	7.3	2.2%	99.0%	0.0%
IN	7.2	2.2%	79.5%	5.3%
LA	8.3	1.2%	21.5%	7.4%
CT	8.4	1.2%	15.3%	0.0%
ID	7.1	0.7%	96.2%	0.1%
CA	7.1	0.7%	71.3%	8.3%
NC	6.6	0.6%	91.4%	0.9%
NJ	7.5	0.5%	47.5%	1.4%
HI	7.6	0.3%	10.5%	0.0%
AL	7.7	0.2%	24.0%	7.9%
TN	7.8	0.1%	30.1%	0.7%
MN	7.5	0.1%	45.3%	0.0%
NM	7.4	0.0%	53.0%	0.1%
MI	7.2	0.0%	63.9%	1.8%
OH	7.2	0.0%	68.7%	2.7%
IL	7.4	0.0%	69.0%	2.1%

Colonoscopy Scores by State				
State	Median Score	% < 5	% <7.5	% > 9
IL	7.4	0.0%	69.0%	2.1%
PA	7.5	0.0%	35.9%	1.0%
FL	7.2	0.0%	60.1%	5.3%
AK	7.7	0.0%	13.0%	0.0%
AR	7.8	0.0%	6.0%	19.9%
AZ	7.2	0.0%	84.1%	4.7%
DE	7.7	0.0%	1.0%	3.1%
GA	7.7	0.0%	35.9%	3.9%
KY	7.4	0.0%	58.6%	0.1%
MA	6.7	0.0%	93.6%	0.5%
ME	7.0	0.0%	85.6%	0.0%
MO	7.5	0.0%	40.8%	3.6%
MS	7.6	0.0%	42.2%	0.9%
MT	7.1	0.0%	52.3%	45.9%
ND	7.9	0.0%	0.0%	0.0%
NV	10.0	0.0%	10.0%	66.7%
OK	7.2	0.0%	55.6%	0.4%
OR	7.3	0.0%	65.6%	0.1%
RI	7.2	0.0%	61.1%	11.1%
SD	6.9	0.0%	99.9%	0.0%
UT	7.4	0.0%	97.8%	0.0%
VA	7.1	0.0%	90.1%	0.6%
VT	8.5	0.0%	0.0%	9.1%
WV	6.3	0.0%	88.0%	2.8%
WY	7.7	0.0%	15.5%	2.1%

The AMA encourages the developer to reconvene the subcommittee to address the additional areas of recommended evaluation and testing including additional testing correlating quality and cost and potential addition of socio-economic risk factors into the model.