CBE 3510: Screening/Surveillance Colonoscopy Episode-Based Cost Measure
Response to American Medical Association Public Comment
Spring 2026 Maintenance of Endorsement

We appreciate the American Medical Association's (AMA) detailed engagement with the Screening/Surveillance Colonoscopy episode-based cost measure and thank the AMA for its comments on measure definition, cost-quality connection, risk adjustment, and validity assessment. Our response addresses each area in turn.

Measure Definition

We understand the AMA's concern that the measure may be interpreted as assuming every clinician and group can achieve the lowest possible cost and may not enable an assessment of whether costs are reasonable and appropriate. The measure is not designed to reward the lowest possible cost per episode, nor is it intended to determine whether a given episode's cost is "reasonable and appropriate" in absolute terms. Rather, for each attributed episode the measure calculates the ratio of the episode's observed cost to the cost expected given the patient's clinical case-mix, then averages that per-episode observed-to-expected ratio across all of the clinician's or group's attributed episodes to produce the measure score. Variation in patient risk profile, including factors the AMA identifies as preference-driven, such as anesthesia use or bowel-preparation type, is addressed through the risk-adjustment model, which accounts for legitimate differences in case complexity. We acknowledge the AMA's observation that the cross-reference from Section 6.2.4 to Table 1 could be clarified: the narrative cites episode-level percentile cutoffs of observed cost while Table 1 presents entity-level decile means of the risk-adjusted measure score and risk-adjusted cost.

We understand the AMA's concern that the measure could reward clinicians who fail to screen high-risk patients or identify precancerous lesions. This measure focuses on the cost of screening and surveillance colonoscopies performed by attributed clinicians and groups; it does not attribute or score clinicians based on decisions not to perform colonoscopies. Clinical performance dimensions such as patient selection for screening and lesion detection are directly assessed through quality measures, including Merit-based Incentive Payment System (MIPS) Quality ID 320 (Appropriate Follow-Up Interval for Normal Colonoscopy in Average Risk Patients) and the Qualified Clinical Data Registry (QCDR) GIQuIC26 Screening Colonoscopy Adenoma Detection Rate, and cost information from this measure is intended to be paired with quality measures in reporting and interpretation to provide a complete view of clinician performance.

We acknowledge the AMA's observation that evaluating episodes separately by site of service means that shifting to a lower-cost setting does not lower a clinician's episode cost score. Estimating risk-adjustment coefficients separately by site of service was an intentional design decision informed by the Gastrointestinal Disease Management Medical and Surgical Clinical Subcommittee that developed the measure. The Subcommittee recognized that (i) not all clinicians and groups practice in areas with a full range of setting options, and (ii) patient clinical

factors can drive setting selection. Estimating a single set of risk-adjustment coefficients across all settings would attribute setting-mix differences that are often outside a clinician's control to the clinician's performance. This site-specific structure directly addresses the AMA's concern about variation in the availability of care settings across geographies: because a clinician's score is not penalized for the setting mix available in their area, the design avoids attributing setting-availability differences to the clinician's performance. From a statistical perspective, pooling forces a compromise fit that is biased toward the dominant group, and can cause systematic over-prediction for lower-mean subgroups and under-prediction for high-mean subgroups. Running separate models is usually the better approach when the subgroup means are expected to be different, such as with the differences in payment policies for colonoscopies performed in different settings.

We appreciate the AMA's attention to the completeness of complication-related services captured in the episode, the potential for rural clinicians to disproportionately rely on Emergency Departments (EDs) for complication management, and the availability of information on cost drivers beyond site of service. Outpatient clinician services performed during the 14-day post-trigger episode window, including office and urgent care evaluation and management (E/M) visits, are assigned to the episode when they meet the service assignment rules listed in the "Service_Assignment" tab of the Screening/Surveillance Colonoscopy Measure Codes List file, which specify the codes and clinical themes (cardiopulmonary complications, lower gastrointestinal hemorrhage, pathology, perforation or peritonitis, and repeat colonoscopy or flexible sigmoidoscopy) that define clinically related services. We have not observed evidence that clinicians and groups practicing in rural areas systematically perform worse on this measure than those in other settings. Service-level detail, including the contribution of anesthesia services and other clinical themes to episode cost, is already available to clinicians and groups through the MIPS performance feedback reports, and additional analyses of the services driving episode cost are presented in the validity testing section.

We appreciate the AMA's attention to several design elements, including bowel-preparation quality, the incomplete-procedure modifier, and the 14-day episode window's treatment of repeat colonoscopies, as well as its proposals for design changes including an extended window for repeats and separate tabulation of anesthesia costs. The measure specifications reflect intentional decisions by the Clinical Subcommittee that developed the measure and feedback received during the 2017 National Field Test. The specifications have remained consistent since development, receiving National Quality Forum (NQF) Cost and Efficiency Standing Committee endorsement in 2019 through the NQF Consensus Development Process. Additionally, public comments received during the 2022 Wave 1 Reevaluation did not raise concerns about these aspects of the measure specifications. We note the AMA's proposed design changes for future consideration.

Cost and Quality Connection

We thank the AMA for its emphasis on the importance of connecting cost performance to quality of care, and we share the AMA's view that examining this relationship is a valuable component of the Usability assessment, consistent with the Battelle guidance the AMA references. We agree that a widely reported quality outcome measure such as the GIQuIC26 Screening Colonoscopy Adenoma Detection Rate would in principle provide useful quality context; however, that measure is rarely reported within MIPS, which precluded a meaningful correlation analysis for this evaluation. In place of a direct cost-to-quality-measure correlation, the validity testing analyzes the relationship between measure performance and clinician-modifiable, claims-based events (namely, repeat colonoscopy within the episode window and post-procedure adverse events), which are identified using administrative claims data, not cost data alone, and which are what could be reliably measured across the attributed population.

We appreciate the AMA's interest in understanding the measure's effect on Medicare spending since its introduction in MIPS. The measure is intended to assess clinician and group performance relative to their peers within a given performance period, rather than to assess changes in cost over time. Section 6.2.4 documents the sustained performance gap the measure continues to produce under current scoring, indicating continued opportunity for cost efficiency across the attributed population. Any evaluation of the measure's broader effect on Medicare spending would need to be interpreted relative to the specific MIPS scoring environment in each performance period, which has changed repeatedly since the measure's implementation, including the reweighting of the MIPS cost performance category to 0% for the 2020 and 2021 performance periods due to the COVID-19 public health emergency, and the change beginning with the 2024 performance period in the cost-measure achievement-point methodology, which replaced decile-based percentile benchmarks with a distribution anchored on the median cost (set at 10% of the performance threshold) with benchmark cut-points defined by standard deviations from the median.

Risk Adjustment

We appreciate the AMA's detailed review of the risk-adjustment model and its questions about the inclusion of variables beyond the eight clinically identified adjustors, including examples such as macular degeneration. The measure's risk-adjustment model uses the Centers for Medicare & Medicaid Services (CMS) Hierarchical Condition Category (HCC) framework, which is the standard risk-adjustment approach applied across MIPS episode-based cost measures. The HCC framework is combined with the eight measure-specific clinical adjustors identified by the Gastrointestinal Disease Management Medical and Surgical Clinical Subcommittee during measure development. The model is calibrated at the population level and is evaluated based on aggregate predictive performance; as documented in Section 5.4, predictive ratios near 1.0 across risk deciles show that the risk adjustment model is well-calibrated. Consistent with this population-level calibration, individual HCC variable coefficients can appear counterintuitive when examined in isolation. This is an expected property of risk-adjustment models estimated

empirically from claims, in which the full set of CMS-HCC and other adjustors is fit jointly; individual coefficients represent marginal statistical associations rather than standalone predictions of the cost of a service or of a site of care. Interpreting a single coefficient in isolation, or as a prediction of the cost differential between sites of service, does not reflect how the adjustment operates within the model. A reduced eight-variable-only model would depart from the standard MIPS episode-based cost measure risk-adjustment methodology. In addition, coefficients should be interpreted with caution: in an ordinary least squares (OLS) model, each coefficient represents the expected change in cost for a one-unit increase in that risk adjustor, holding all other independent variables constant, but only on average across the model's population, so the same coefficient may not hold for any given subgroup. This is a reason separate models are performed: coefficients, like means, are not expected to be constant across subgroups.

We acknowledge the AMA's concern about how differences in site-of-service mix across geographic areas are addressed by the risk-adjustment model. The decision to estimate separate coefficients for episodes performed in Hospital Outpatient Departments (HOPDs), Ambulatory Surgery Centers (ASCs), and office settings was informed by the Clinical Subcommittee that developed the measure and reflects genuine differences in cost drivers across settings. The same population-level calibration point applies to counterintuitive coefficients observed in the setting-specific models, including the pressure-ulcer example the AMA cites. Separate models by site of service ensure clinicians are not unfairly penalized when their patients, often due to access barriers rather than clinical necessity, receive care in a higher-cost setting such as an HOPD. Each setting's model reflects its own cost structure and patient population, so performance is judged relative to episodes in the same setting, not against a blended benchmark that assumes uniform access to ASCs and office settings.

We appreciate the AMA's attention to health-related social needs and their potential relationship to episode cost. We note that Closing Care Gaps remains optional for maintenance of endorsement. Additional testing of the effect of including social risk factors in the risk-adjustment model is not available at this time.

Validity Assessment

The measure's validity testing analyzes whether higher rates of clinically meaningful events at the clinician or group level (i.e., repeat colonoscopy and post-procedure adverse events within the episode window) are associated with higher risk-adjusted episode costs. Testing demonstrates the expected positive association across the attributed population, providing evidence that the measure captures meaningful sources of cost variation that can be under the reasonable influence of the attributed clinician or group. The modest, but generally statistically significant, magnitude of the correlation reflects the low base rates of these events across the attributed population.

We acknowledge that the AMA has conducted analyses of MIPS performance information showing variation in state-level distributions of Performance Year 2024 (PY2024) MIPS cost scores. We would like to note that this data only reflects those clinicians and groups who ultimately had the Screening/Surveillance Colonoscopy measure contribute to their MIPS final score, and may not be representative of the full population eligible to be assessed on this measure. Additionally, the measure's risk adjustment is designed to account for patient case-mix at the episode level, not to standardize the distribution of clinician-level scores across geographic aggregates. Some differences in performance distributions across states are therefore expected, either by chance in the reported subset or due to actual performance differences.

Conclusion

We thank the AMA again for its detailed comments and thoughtful engagement with the Screening/Surveillance Colonoscopy episode-based cost measure. The measure's current specification reflects the guidance of the Gastrointestinal Disease Management Medical and Surgical Clinical Subcommittee that developed it, the 2019 NQF Cost and Efficiency Standing Committee endorsement, sustained CMS use in MIPS since PY2019, and stakeholder input received through the 2022 Wave 1 Reevaluation public comment period. We appreciate the AMA's recommendations regarding additional testing of cost-quality correlation, further analysis of social risk factor adjustment, and refinements to the measure specification, and note these recommendations for future consideration.