ASC Quality Collaboration

January 28, 2015

**VIA ELECTRONIC SUBMISSION**

Board of Directors
National Quality Forum
1030 15th Street NW
Suite 800
Washington, DC 20005

**Re: Appeal of Endorsement of NQF #2539, Facility Seven-Day Risk-Standardized Hospital Visit Rate After Colonoscopy**

Dear Members of the Board of Directors:

On behalf of the ASC Quality Collaboration (ASC QC), a cooperative effort of organizations and companies interested in ensuring ambulatory surgical center (ASC) quality data is appropriately developed and reported, please accept the following remarks in appeal of the recent National Quality Forum (NQF) endorsement of the *Facility Seven-Day Risk-Standardized Hospital Visit Rate After Colonoscopy* measure (NQF #2539). The ASC QC's stakeholders include ASC corporations, ASC industry associations, physician and nursing professional societies, and accrediting bodies with oversight of ASCs. Please see Appendix A for a list of these organizations, which represent over 1,500 ASCs.

The ASC QC's strong commitment to advancement of quality is reflected in the steps we have taken independently to facilitate quality reporting by ASCs –including the development of facility-level measures for ASCs; participation in a broad range of quality measurement projects with partners including federal agencies such as AHRQ, CDC and CMS; and providing education regarding the ASC industry and environment of care with the goal of improving quality measurement efforts.

We appreciate the ongoing efforts of the NQF to set high standards for healthcare quality measurement. Unfortunately, there are serious technical issues with the *Facility Seven-Day Risk-Standardized Hospital Visit Rate After Colonoscopy* measure that cause it to fall short of meeting the normally rigorous standards set by NQF. We have been following this measure closely throughout the All-Cause Admissions and Readmissions Measures Project, and though we have commented on several important issues related to this measure, the measure developer has not satisfactorily addressed all of them. We do not believe the measure should remain endorsed without resolution of these outstanding issues.

### A. Lack of Validity Testing in the Settings of Care Measured

Validity testing for this measure relies primarily on work the developer did in the past related to a different measure for the inpatient setting. This testing relied on the use of inpatient hospital claims. For reasons that are unclear, the developer has cited this work with inpatient claims as a basis for the validity of this measure, which is based on outpatient claims.

As the Board is well aware, claims are generated for purposes of reimbursement, and Medicare's inpatient and outpatient payment systems are governed by entirely different sets of rules. These rules have a significant impact on the coded information that is submitted on claim forms. Inpatient reimbursement is primarily diagnosis driven, and the coding practices surrounding inpatient claim submission reflects this, with a strong emphasis on fully characterizing the patient's diagnoses and comorbid conditions. In contrast, outpatient reimbursement is service driven, and coding practices are focused on accurately describing the services rendered. Diagnosis coding in ASCs is focused on establishing the medical necessity of those services; the ASC billing format does not support the reporting of diagnosis codes that are not explicitly associated with the services provided to the patient during the encounter.

Because this measure relies *entirely* on that coded information for risk adjustment, it is essential to establish that the codes submitted are, *in fact and not in supposition*, reflective of the clinical aspects of care that the measure purports to measure. The developer went to the effort of testing and establishing this for the inpatient measure it developed, but did not evaluate it for this outpatient measure. Given the different claims structures for inpatient versus outpatient claims, the inpatient results cannot be assumed to apply to the outpatient setting. The sensitivity and specificity of using administrative claims data following outpatient colonoscopy must be determined in order to establish the validity of the measure.

Further, this measure intends to evaluate both Hospital Outpatient Department (HOPD) and ASC performance, but these two settings do not use the same claim format. HOPDs submit claims using the UB-04; ASCs submit claims using the CMS-1500. Among their differences, the two forms vary in the total number of fields available for the submission of diagnosis codes, and in the types of fields associated with diagnosis coding. Without testing, one cannot claim that HOPD results can be fairly and appropriately compared to ASC results.

These issues were brought to the attention of the measure developer and NQF through our comments on the draft report issued by the All-Cause Admissions and Readmissions Standing Committee. The measure developer responded, "CMS will take these points into consideration in field testing and implementing the measure." Yet we see no indication that CMS plans any kind of field-testing at all; rather, the agency is proposing to move directly to implementation. We object in the strongest possible terms to *post-endorsement* validity testing.

The impact of these differences – outpatient versus inpatient, and HOPD versus ASC - must be systematically assessed to assure the measure results are attributable to differences in quality rather than differences in claims architecture and coding practices. The measure score should be directly validated against outpatient medical records and measure results across settings must be assessed to ensure that any cross-setting comparisons are valid.

**B. The Measure Is Not Valid Due to Systematic Undercounting of HOPD Events**

As noted above, the *Facility Seven-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy* measure purports to be specified for both the ASC and HOPD settings. It has been asserted that the measure is "well-defined and precisely specified for consistent implementation within and between organizations that will allow for comparability." This is not true. Although the measure is likely to do a good job of counting hospital visits following ASC care, it would *systematically undercount hospital visit rates following HOPD care* occurring in the seven-day period following outpatient colonoscopy. As the measure is currently specified, comparisons across the two settings cannot be made on equal footing. The following explains why near-term events following care in the HOPD setting would not be counted accurately using this measure's algorithm.

This measure's reliance on the use of administrative claims has a significant and adverse impact on the validity of the measure results. As a result of Medicare's three-day payment window policy, there are major challenges in identifying index HOPD visits, and therefore subsequent "hospital visits" related to HOPD care. The three-day payment window policy requires that outpatient services provided by a hospital, or any Part B entity wholly owned or wholly operated by a hospital, *on the date of a beneficiary's inpatient admission* must be billed with the inpatient stay. In addition, outpatient services provided by a hospital, or any Part B entity wholly owned or wholly operated by the hospital, *on the first, second, and third calendar days preceding the date of a beneficiary's inpatient admission* are also deemed related to the admission, and must be billed with the inpatient stay. Part B entities affected by this policy include hospital outpatient departments, hospital emergency departments and wholly owned physician practices. The three-day payment policy applies to all non-diagnostic services provided during the payment window unless the hospital attests that the services are clinically unrelated. Diagnostic services are always subject to the payment window policy, irrespective of whether they are considered clinically related.

Simply stated, CMS does not permit HOPDs to generate a claim when there is an inpatient admission during the three-day window, except in cases where the service was therapeutic and the hospital attests that the subsequent admission was unrelated. Claims that do not exist cannot be counted. As a result, this measure cannot identify inpatient admissions that may have resulted from colonoscopies performed in the HOPD setting when those unplanned admissions occur on the date of the colonoscopy, or during the three days subsequent to the procedure. The measure would only identify hospital visits occurring on days 4, 5, 6 and 7 following the index HOPD visit; index claims for days 0, 1, 2 and 3 would not be created or counted. This missing data skews the measure results by undercounting the number of inpatient admissions attributed to the HOPD. As a result, measure scores cannot be compared across settings.

In the time since this measure was first brought before the MAP and we pointed out this flaw in the measure's design, the developer has looked for ways to work around the three-day payment window policy. As things stand now, the measure algorithm identifies the index colonoscopy using claims for colonoscopies in the Part B carrier file. ASC facility claims would be identified directly, using ASC facility claims. However, HOPD claims would be identified *indirectly, using physician claims*. Specifically, the measure algorithm would look for physician claims for colonoscopy indicating an HOPD place of service (POS) that had an inpatient admission

within 3 days *and* lacking a corresponding HOPD claim. It would then count such physician claims as HOPD "claims".

The problem with this approach is that POS coding has a long history of inaccuracy. Over a period of more than a decade, the Department of Health and Human Services Office of Inspector General (HHS OIG) has performed repeated audits of physician POS coding that *consistently demonstrate high error rates*. These errors result in physician claims that indicate the service was performed in the physician office, when in fact the service was actually performed in a hospital outpatient department or ASC. See, as some examples of many such OIG reports over the years: A-02-04-01010 (error rate 88%), A-05-04-00025 (error rate 79%), A-06-04-00046 (error rate 76%), A-01-06-0052 (error rate 81%), A-01-09-00503 (error rate 90%) and, most recently, A-01-10-00516 (error rate 83%) of September 2011. *Errors in POS coding are not an isolated, infrequent or insignificant problem*. In light of this information, it's not possible to attribute any credibility to a measure algorithm that would use POS coding on physician claims as a means of identifying HOPD claims.

The practical impact on the measure is this: its plan to rely on POS coding to identify HOPD claims that are missing due to the three-day payment window policy means that a significant number of these missing index HOPD claims *will never be identified*. Any algorithm that relies on using POS coding on physician claims would systematically undercount HOPD events by failing to identify a significant number of the index HOPD visits. As a result, not only would the HOPD rates reported be inaccurate (too low), they would also not be comparable to ASC results.

Unless HOPD claims during the three-day payment window can be accurately identified, the measure routinely disadvantages ASCs by unfairly reporting complete rates for ASCs and incomplete rates for HOPDs. We object in the most strenuous terms to the endorsement of such a fundamentally flawed measure as an appropriate means for the comparison of ASC and HOPD hospital visit rates following colonoscopy.

**C. Measure Rationale and the Three-Day Payment Window Policy**

In order to be endorsed, measures must demonstrate meaningful performance gaps. We remain concerned that the three-day payment window policy may also have impacted the data used in the analyses performed to establish the performance gap for this measure. These analyses estimated the measure score for both ASCs and HOPDs using 2010 Healthcare Cost and Utilization Project (HCUP) data, and then separately calculated the measure score for HOPDs alone using 2010 data from the Chronic Conditions Data Warehouse (CCW). Both analyses found provider variability. It is unclear how much of this variability may have been a reflection of the three-day payment window policy, which was implemented for dates of service on or after June 25, 2010. Those Medicare claims before June 25 would have included index HOPD visits that occurred within the three-day window; Medicare claims on or after June 25 would not have included index HOPD visits that occurred within the three-day window. It is possible that conclusions reached regarding variability in performance - based entirely on these analyses - are incorrect, and that the variability observed was actually a function of the change in CMS payment policy in the middle of the period analyzed.

**D. Issues with the Reliability of the Measure**

The measure developer has acknowledged that the number of outcome events for this measure is already low. To manage this, the measure has been specified in ways that generate large case volumes (for example, the inclusion of physician office claims for colonoscopy in the measure denominator, despite its characterization as a "facility-level measure"). Despite these steps, the results of the reliability testing for this measure were quite low. With two years of data, the intra-class correlation coefficient (ICC) was, on average, 0.335, which according to conventional interpretation is only "fair."

Of note, this subpar result was only obtained *after excluding low volume facilities* from the calculation. As the measure developer explained to NQF, "[b]ecause we expect facilities with relatively few cases to have less reliable estimates, we only included scores for facilities with at least 400 cases in the reliability calculation (i.e., with 200 cases in each of the split samples, about 100 cases/year). This approach is consistent with a reporting strategy that includes smaller facilities in the measure calculation but does not publicly release the measure score for smaller facilities (i.e., labels them in public reporting as having "too few cases" to support a reliable estimate)." Unfortunately, the measure specifications do not make it explicit that the measure is not appropriate for use in facilities with low volumes of colonoscopies.

After originally submitting the measure to NQF for consideration, the developer recalculated the reliability testing score using the Spearman-Brown prophecy formula, in order to approximate the ICC for a three-year sample. This resulted in a higher ICC of 0.43, which according to conventional interpretation is "moderate" (though on the very low end of moderate, which ranges from 0.41 to 0.60). Though we believe this is a non-standard application of the Spearman-Brown formula, the NQF's All-Cause Admissions and Readmissions Standing Committee has accepted this ICC score, which provides a reliability estimate for three years of data, as sufficient to meet reliability endorsement criteria.

Because these reliability issues have not been addressed during the NQF endorsement process, the measure can be implemented without addressing either the need to exclude low volume facilities or the need to include multiple years of data collection. For example, although the measure developer has suggested a three-year data collection period, CMS is planning to use a one-year period. We believe the NQF must take responsibility for ensuring that all conditions that could affect the reliability of the measure results be explicitly stated in the measure specifications in order to ensure that when the measure is implemented, it can be reasonably assumed that the measure will generate reliable results. Leaving key aspects of data collection to the discretion of the implementer, especially when this impacts the reliability of publicly reported data, is not a situation NQF should allow. Endorsed measures should be adequately specified so that the circumstances under which the measure score is reliable are abundantly clear.

In addition, we would like to continue to express our opinion that the reliability of a measure intended for public reporting and accountability purposes should be substantial given what is at stake. NQF requires other types of measures to have high standards of reliability, yet this category of measures has been permitted to systematically fall short of the usually much higher expectations. We believe that if facilities are to be publicly judged based on the results calculated for this measure, the reliability of those measure scores should be "substantial" (0.61 to 0.80 per convention), at a minimum.

**E. The Measure Score Suffers from a Lack of Actionability in ASCs**

The rates of the outcomes the measure seeks to identify are low. As a result the measure has been specified in ways that generate large case volumes, but that diminish its usefulness. Specifically, the need for volume prevents stratification, meaning the measure score would be reported as a single rate for each facility. This presents challenges for actionability because the measure score provides no insight other than how the facility's rate of hospital visits compared to the expected rate. Because the data used to generate the measure score are not accessible to the facility, it would be impossible for the ASC to determine even basic information, such as which patients were affected, the numbers of ED visits, observation stays or inpatient admissions that occurred, or why any subsequent visit occurred. The measure developer has stated, "CMS agrees that the measure score alone provides limited information for quality improvement since the outcome combines ED, observation stays, and admissions, and that more detailed information on patient outcomes would assist facilities with quality improvement. CMS plans to report patient-level data confidentially to facilities that indicates whether the patient had a hospital visit, the type of visit (admission, ED, or observation stays) if any, and the facility to which the patient is admitted."

In essence, the measure developer has not built usability into the measure, but is relying on those who implement the measure to come up with some means to provide actionable data. Although CMS provides facility-specific reports to hospitals regarding inpatient re-admissions - and we are cautiously optimistic that CMS would include the principal discharge diagnosis for the hospital visit - these steps would not be sufficient to support meaningful quality improvement *in ASCs* because the ASC would have such limited insight into why the patient's hospital visit occurred.

The measure developer has pointed to the improvements in hospital readmission rates as evidence that this measure would produce comparable improvements in colonoscopy outcomes. While it is true that selected hospitals have been successful in reducing readmission rates, these reductions have resulted from analyses that extend well beyond the receipt of a CMS benchmarking report with limited data about their readmissions. Reviews of hospital industry guidance and the case reports of successful hospitals indicate that these improvements have resulted when systems have been put in place that allow the hospital to identify readmissions to their own facility in real time. These systems flag patients that are readmitted, allowing staff to review the patient's record in detail and perform root cause analysis to determine what led to the patient's readmission. Many interview selected patients and their families at the bedside during the subsequent admission to learn more about why the patient was readmitted and what, if anything, could have been done to prevent it. This allows the hospital to identify and prioritize improvement opportunities. These hospitals are successful precisely because they have access not only to the patient's records from the index admission, but also to the patient's records (and the patient) at the time of readmission.

What the measure developer has not recognized is that the ASC setting, as a result of legislation and regulation, is markedly different from the hospital setting in ways that would significantly impact the ability to develop a performance improvement initiative around the results of this measure. In accordance with Federal regulation, ASCs are a unique supplier type that serves *solely* as the site for outpatient surgery and is involved with the care of the patient only immediately before, during and immediately after a surgical procedure. Unlike other outpatient surgical settings,

such as clinician offices, ambulatory clinics or hospital outpatient departments, ASCs *may not provide post-operative follow-up care* after patient discharge. ASCs, as distinct entities that operate in an entirely separate capacity from physician offices, emergency departments, and hospitals *do not have direct access to the records* of these other providers. (Please see the Code of Federal Regulations at 42 CFR 416 and the CMS State Operations Manual Appendix L - Guidance for Surveyors: Ambulatory Surgical Centers if documentation of the above is needed.) As a result of this mandated isolation, ASCs must either interview the patient over the phone or obtain permission from the patient to obtain their medical records from the treating emergency department or admitting hospital in order to obtain the information needed to perform the analyses required for effective improvement activities surrounding this measure. It is certainly possible that selected patients would cooperate with this effort, but in our experience, this willingness diminishes rapidly over time. If an ASC were to contact patients a year or more after their ASC care (which is when the ASC would receive a CMS report on their performance for this measure) we expect the number of patients agreeing to requests for interviews or medical record releases would be very small indeed.

In short, the measure itself does not provide usable results, and the processes outside the measure that the developer points to as filling this gap are not sufficient to provide actionable data for the ASC setting. NQF should not endorse measures that do not provide actionable results.

**F. Very Limited Ability to Make Distinctions Among Facilities**

While administrative claims do not impose additional data collection or submission burdens on providers, they are blunt instruments for assessing quality. This measure suffers from very limited discriminatory power. Using the standard 95 percent interval estimate to report the measure score, the developers indicate 99.5% of facilities would be classified as no different than expected, 0.4% of facilities as worse than expected and 0.1% of facilities as better than expected. The overwhelming majority of facilities, 99.6%, would receive a measure score indicating their performance was at, or better than, the expected level. The number of underperforming facilities would be very small. If we extrapolate 0.4% to the entire universe of approximately 5300 Medicare-certified ASCs (*not all of which perform colonoscopies, so this is clearly an overestimate*), the number of underperforming ASCs would be *very generously estimated* at twenty-one facilities. This also means it would be equally unusual for a consumer to be able to discriminate among facilities using the results of the measure.

While the developers state there is variability in performance, as a practical matter the risk standardized results indicate little room for improvement. In fact, one could legitimately conclude this measure, based on the data presented during the endorsement process, is *already "topped out"*.

**G. Extremely Long Timeframe Means the Measure Score Could Mislead**

As a result of the already low rate of the outcomes for this measure, a very long data collection period (3 years) is required in order to generate measure scores that are even moderately reliable. Even if we set aside the issue of the significant lag time from the generation of claims to the reporting of measure results, the measure's extended data collection timeframe means that past performance would continue to impact the measure score for each facility for a long time. The publicly reported measure score would not be a reflection of current, or even recent, performance.

In fact, the score would obscure either significant improvement or significant deterioration in recent performance. As a result, consumers could be misled by the lack of timely data. They could mistakenly believe a facility is no different from others, when in fact it has made significant recent improvements and would be a superior choice. Or they could be led to believe a facility is no different from others when in fact the facility has had a recent (and even steep) decline in performance and would be an inferior choice.

**H. Incomplete Adaptation to the Outpatient Setting**

As noted above, the measure developer has relied heavily on previous work performed to develop a measure of inpatient readmissions in constructing key elements of this outpatient measure, such as the planned admission algorithm and risk adjustment model. Unfortunately, the measure that has been put forth to the NQF retains elements of the inpatient measure that are not appropriate, an indication that the measure has not been thoroughly reviewed and fully adapted for outpatient use. As just one example, certain condition categories (CCs) are not included in risk adjustment if they are only recorded at the time of the colonoscopy, as they are considered to be possible adverse outcomes. Although end stage renal disease (ESRD) would not be a complication of colonoscopy diagnosed and recorded at the time of the procedure, it was included on the list of CCs and was not removed until we pointed it out. There are others that have not been addressed.

The measure developer indicated it would review the list of CCs with their technical experts, though this group is said to have already reviewed the measure details previously. We see no evidence that this review happened prior to endorsement and know that inappropriate vestiges of the original inpatient measure remain embedded in the outpatient specifications. NQF should not endorse measures that are not fully appropriate to the setting of care they purport to measure, and should not rely on post-endorsement fixes when there are pre-endorsement indications that changes are needed.

***

Thank you for considering our concerns regarding the endorsement of this measure. We look forward to continuing our long association with the National Quality Forum and would be happy to assist with questions or provide additional information at your request.

Sincerely,

Donna Slosburg, BSN, LHRM, CASC
Executive Director, ASC Quality Collaboration
727-367-0072
donnaslosburg@ascquality.org

## Appendix A
## Current Participants in the Activities of the ASC Quality Collaboration

Accreditation Association for Ambulatory HealthCare
Ambulatory Surgery Foundation
Ambulatory Surgical Centers of America
American College of Surgeons
American Osteopathic Association, Healthcare Facilities Accreditation Program
AmSurg
ASD Management
Association of periOperative Registered Nurses
Covenant Surgical Partners
Florida Society of Ambulatory Surgical Centers
Hospital Corporation of America, Ambulatory Surgery Division
Outpatient Ophthalmic Surgery Society
Regent Surgical Health
Surgery Partners
Surgical Care Affiliates
The Joint Commission
United Surgical Partners International
Visionary Enterprises, Inc.