

Managing the Measurement of Colonoscopy Quality

Jason A. Dominitz, MD, MHS¹ and Cynthia W. Ko, MD, MS²

Abstract: The adenoma detection rate (ADR) is our current best colonoscopy quality indicator, but it is not without limitations. In this issue of the Journal, novel ADR benchmarks are proposed based on historical local colonoscopy results. These minimally acceptable, standard of care, and aspirational benchmarks may encourage continuous quality improvement through the explicit determination of notably higher but proven achievable ADR targets, although validation in clinical practice is needed. Ultimately, we must transition from ADR measurement to the implementation of robust quality improvement processes that assure the best outcomes for our patients.

Am J Gastroenterol 2019;114:1199–1201. <https://doi.org/10.14309/ajg.0000000000000307>

Management guru Peter Drucker's saying that "what gets measured, gets managed" applies not only to the business world but also to health care overall and, especially for gastroenterologists, to colonoscopy. In 2002, the US Multi-Society Task Force on Colorectal Cancer created a new quality metric called the adenoma detection rate (ADR) and established a benchmark ADR of $\geq 20\%$ for men and women undergoing average-risk screening colonoscopy (1). Although the concept of an ADR had face validity as a quality metric, it was not until 2010 that it was demonstrated that physicians' ADR was associated with their patients' risk of developing postcolonoscopy colorectal cancer (2). In 2014, Corley et al. (3) studied 136 endoscopists whose ADR ranged from 7% to 53%. Each 1% increase in ADR was associated with a 3% decrease in cancer incidence and a 5% decrease in cancer mortality, with a hazard ratio for interval cancer of 0.52 when comparing highest to lowest ADR quintile endoscopists. Importantly, this retrospective study did not find a clear "ceiling effect" between the ADR and cancer outcomes. Thus, in 2015, the American Society for Gastrointestinal Endoscopy/American College of Gastroenterology Task Force on Quality in Endoscopy (Task Force) increased the ADR benchmark to 25% (4), and in this issue of the Journal, Hilsden et al. (5) proposed ADR benchmark refinements to promote quality improvement beyond a minimally acceptable threshold.

True to Drucker, the measurement of the ADR has led to the management of the ADR. Numerous studies have tested techniques and technologies to improve the ADR (6–8), and the ADR is now a formal quality measure for Medicare's Quality Payment Program (9). However, what we do not know is what constitutes the minimally acceptable ADR. As a profession, we are expected to police ourselves and assure that only competent professionals are permitted to care for patients. So where should we "draw the line" on competence? Is the Task Force ADR benchmark of $\geq 25\%$ the correct threshold? Are variations in the

ADR due to differences in the underlying population risk expected and, therefore, acceptable?

In an effort to address some of these questions, Hilsden et al. (5) proposed a novel approach to develop ADR benchmarks based on historical colonoscopy results in the local population. Their approach begins with classifying endoscopists into quartiles of performance based on their ADR in a baseline year (year 0). They then propose benchmarks for (i) a minimally acceptable ADR, (ii) a standard of care ADR, and (iii) an aspirational ADR during the subsequent year (year 1). The minimally acceptable benchmark was defined as the mean ADR found in year 1 for those endoscopists in the lowest 2 quartiles in year 0. The standard of care benchmark was similarly defined using the average year 1 ADR for those in quartiles 2 and 3, whereas the aspirational benchmark was based on the average year 1 ADR of those in the fourth quartile. For the population of Calgary, Canada, the recommendations were a minimally acceptable ADR of 25% (coincidentally identical to the Task Force benchmark), a standard of care ADR of 30%, and an aspirational ADR of 39%. To formally account for random variation, 95% confidence intervals are calculated for each physician's ADR. Of the 29 physicians studied, 1 (3%), 2 (7%), and 9 (29%) failed to reach the minimally acceptable, standard of care, and aspirational benchmarks, respectively.

This benchmarking approach is appealing because it uses local practice data and encourages continuous quality improvement through the explicit determination of notably higher but proven achievable ADR targets. However, there are some important considerations. First, the authors' definitions of the minimally acceptable, standard of care, and aspirational benchmarks were somewhat arbitrary and not fully validated. However, for many, the phrase "minimally acceptable" would define the standard of care. Whether alternative definitions of the ADR, such as those proposed by Hilsden et al., lead to actual quality improvement requires prospective evaluation. Second, this approach assumes

¹VA Puget Sound Health Care System, University of Washington School of Medicine, Seattle, Washington, USA; ²University of Washington School of Medicine, Seattle, Washington, USA. **Correspondence:** Jason A. Dominitz, MD, MHS. E-mail: jason.dominitz@va.gov.

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

Received March 18, 2019; accepted May 15, 2019; published online June 25, 2019

no meaningful differences in physicians' patient mix even within the local population. However, case mix does vary between endoscopists, with differences in patient age, sex, and likely other clinically meaningful variables as well. It would be feasible to stratify or adjust the chosen ADR benchmarks by such easily measured variables, as is currently performed for patient sex. In addition, their approach requires data from many endoscopists (the authors recommend at least 20) and performance variability to produce useful benchmarks. Such calculations seem feasible for large practices or health care systems, but challenges remain for smaller practices. Without significant performance variation, the calculated aspirational, standard of care, and minimally acceptable ADRs will be similar. Thus, this approach will not identify physicians who perform low-quality colonoscopy if all endoscopists underperform similarly. Finally, evaluating individual low-volume endoscopists may be challenging because of wide confidence intervals around their ADR estimates. Therefore, this tiered ADR approach should not be used alone, and other quality metrics (e.g., complications and interval cancers) should be used to the fullest extent possible to assess the overall colonoscopy competence.

The flipside of "what gets measured, gets managed" is the idea (attributed to Albert Einstein) that "not everything that can be counted counts, and not everything that counts can be counted." This concept applies to the ADR in that detection of an adenoma is but one small piece of colorectal cancer prevention. Other key components include the detection and complete resection of *all* significant neoplasia and providing appropriate surveillance. Alternatives to the ADR have been proposed to address the first part, such as the ADR Plus, the number of adenomas per colonoscopy, and the adenoma miss rate (10,11), although quality indicators are currently lacking for assessing complete polypectomy (12).

Apart from these issues, the greatest challenges for our profession may be in the implementation and use of such metrics to effect actual quality improvement. First, practices must be able to accurately implement quality metrics without undue burden. Measuring the ADR can be challenging because it requires linkage of the endoscopy and pathology reports. Although natural language processing techniques can assist (13), for many practices, this requires manual chart review. Second, it is unclear how often the local benchmarks would be recalibrated. If an iterative benchmarking process is undertaken, it would be expected that the ADR would improve over time, especially for the lower quartiles (assuming low adenoma detectors either improve or stop performing colonoscopy). Although this is the goal of quality improvement, there may be a risk of misclassifying competent endoscopists because of low colonoscopy volumes, differences in case mix, or other random variation. The potential implications of such misclassifications for the individual physician's practice reinforce the need to assess colonoscopy competency using multiple measures. Most importantly, what should be done about physicians who fail to achieve a minimally acceptable ADR? Previous studies have demonstrated that the ADR can be improved with the use of accessory devices or focused training (6,14,15). However, few mechanisms exist for practicing physicians to undergo this training. Artificial intelligence systems may be an important adjunctive technology for improving the ADR (16), although more clinical trials are needed. Finally, who is responsible for monitoring individual physician's ADR?

Does this responsibility rest with local facilities (who may have a conflict of interest) or at the level of certification boards or governmental bodies?

Our professional societies may be best positioned to take on a greater role in assisting practices with implementation of quality metrics, including the development of effective training resources for practicing endoscopists. However, until such a time as we can assure that individual physicians meet minimum ADR standards, what is to be done? Is there any duty to inform those patients who had undergone screening colonoscopy by a low-performing physician? Should these patients be offered repeat colonoscopy or a shortened interval to their next colorectal cancer screening or surveillance than would otherwise have been recommended? These questions do not have easy answers.

In summary, given its association with colorectal cancer outcomes and amenability to improvement, the ADR is our current best colonoscopy quality indicator. The approach of Hilsden et al. may help tailor this quality metric to the local populations and set the standard of care and aspirational targets for quality improvement, although it requires further validation. Ultimately, our greatest challenge may lie in transitioning from measurement of the ADR to the implementation and management of robust quality improvement processes that assure the best outcomes for our patients.

ACKNOWLEDGMENTS

This material is the result of work supported in part by resources from the Veterans Health Administration. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

CONFLICTS OF INTEREST

Guarantor of the article: Jason A. Dominitz, MD, MHS.

Specific author contributions: J.A.D and C.W.K. contributed to all aspects of this editorial.

Financial support: None.

Potential competing interests: None.

REFERENCES

1. Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: Recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2002;97:1296-308.
2. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010;362:1795-803.
3. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014;370:1298-306.
4. Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. *Am J Gastroenterol* 2015;110:72-90.
5. Hilsden RJ, Rose SM, Dube C, et al. Defining and applying locally relevant benchmarks for the adenoma detection rate. *Am J Gastroenterol* 2019;114:1315-21.
6. ASGE Technology Committee; Konda V, Konda V, Chauhan SS, et al. Endoscopes and devices to improve colon polyp detection. *Gastrointest Endosc* 2015;81:1122-9.
7. Rex DK. Polyp detection at colonoscopy: Endoscopist and technical factors. *Best Pract Res Clin Gastroenterol* 2017;31:425-33.
8. Jia H, Pan Y, Guo X, et al. Water exchange method significantly improves adenoma detection rate: A multicenter, randomized controlled trial. *Am J Gastroenterol* 2017;112:568-76.

9. Centers for Medicare and Medicaid Services. Quality Payment Program. https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2019_Measure_343_MIPSCQM.pdf. Accessed on June 10, 2019.
10. Wang HS, Pisegna J, Modi R, et al. Adenoma detection rate is necessary but insufficient for distinguishing high versus low endoscopist performance. *Gastrointest Endosc* 2013;77:71–8.
11. Aniwan S, Orkoonsawat P, Viriyautsahakul V, et al. The secondary quality indicator to improve prediction of adenoma miss rate apart from adenoma detection rate. *Am J Gastroenterol* 2016;111:723–9.
12. Pohl H, Srivastava A, Bensen SP, et al. Incomplete polyp resection during colonoscopy—results of the complete adenoma resection (CARE) study. *Gastroenterology* 2013;144:74–80 e1.
13. Imler TD, Morea J, Kahi C, et al. Multi-center colonoscopy quality measurement utilizing natural language processing. *Am J Gastroenterol* 2015;110:543–52.
14. Coe SG, Crook JE, Diehl NN, et al. An endoscopic quality improvement program improves detection of colorectal adenomas. *Am J Gastroenterol* 2013;108:219–26; quiz 227.
15. Kaminski MF, Anderson J, Valori R, et al. Leadership training to improve adenoma detection rate in screening colonoscopy: A randomised trial. *Gut* 2016;65:616–24.
16. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. *Gut* 2019. [Epub ahead of print February 27, 2019.]