ORIGINAL ARTICLE

Check for updates

A Business Case for Artificial Intelligence Tools: The Currency of Improved Quality and Reduced Cost

Lauren Parks Golding, MD[a], *Gregory N. Nicola, MD*[b]

Abstract

For data science tools to mature and become integrated into routine clinical practice, they must add value to patient care by improving quality without increasing cost, by reducing cost without changing quality, or by both reducing cost and improving quality. Artificial intelligence (AI) algorithms have potential to augment data-driven quality improvement for radiologists. If AI tools are adopted with population health goals in mind, the structure of value-based payment models will serve as a framework for reimbursement of AI that does not exist in the fee-for-service system.

Key Words: Artificial intelligence, MACRA, MIPS, QPP, quality

J Am Coll Radiol 2019;16:1357-1361.

INTRODUCTION

Two of the most anticipated disruptors in health care in recent history are value-based care transformation and the emergence of artificial intelligence (AI) tools. Certainly, neither has sweepingly transformed health care to date. Yet both innovations will undoubtedly continue to shape the future of health care, and the intersection of the two has implications for radiologists both in terms of payment policy and data-driven quality improvement. For data science tools to mature and become integrated into routine clinical practice, they must add value to patient care by improving quality without increasing cost, by reducing cost without changing quality, or by both reducing cost and improving quality. Increasing the efficiency of radiologists is one example of value added by AI tools; however, for the purposes of this article, we will focus on the potential role of these tools specifically in the value-based payment structure of the Quality Payment Program (QPP).

FEE FOR SERVICE

Although the development of new technology in health care may be the first step in innovation, finding a way to get paid for that technology is often the rate-limiting step for market penetrance. In the fee-for-service system, this process typically begins with the creation of a Current Procedural Terminology (CPT) code [1]. CPT is a medical code set designed to describe medical, surgical, and diagnostic procedures at a granular level to enable tracking and reporting, most auspiciously for billing purposes. Once a CPT code has been created for a new technology or service, the code is assigned a value in terms of relative value units by the Relative Value Scale Update Committee. Relative value units are assigned through a complex and somewhat daunting process based on time-derived activity-based cost [2]. Two components of the valuation process are pertinent to this discussion: physician work (in terms of time and intensity) and practice expense. A detailed description of this process is beyond the scope of this article, but it has been described in this journal previously [3]. Some of the AI algorithms currently under development or theorized involve little or no

[a]Triad Radiology Associates, Winston Salem, North Carolina.

[b]Hackensack Radiology Group, PA, River Edge, New Jersey.

Corresponding author and reprints: Lauren Parks Golding, Triad Radiology Associates, 3010 Trenwest Dr, Winston Salem, NC 27103; e-mail: laurengoldingmd@gmail.com.

Gregory N. Nicola, MD, is CEO/Founder of N2 Health Insights. He is Consultant/CMO for machine learning collaboration with NeuTigers. Dr Golding states that she has no conflict of interest related to the material discussed in this article.

1546-1440/19/$36.00 ■ https://doi.org/10.1016/j.jacr.2019.05.004

physician work; indeed, the goal of many of the early tools is to increase the efficiency of radiologists by decreasing the amount of work (in terms of time) required by the physician. Early in the development process, many algorithms may actually add work for radiologists who now have to review the raw data and the AI output to verify its accuracy before rendering an interpretation. However, because of the nuances of valuing new technology in the CPT and Relative Value Scale Update Committee process, it is unlikely that reimbursable codes would be created to reflect this transient work increase, with the cost largely born by industry and research entities.

Practice expense is a subset of reimbursement designed to account for equipment and labor costs associated with providing a particular service in an office setting. This works well for concrete items such as CT scanners, technologist time, and even ultrasound gel. It works less well for other types of expenses such as add-on software packages, which have diverse sets of applications and cannot be measured as easily in per unit costs.

Physician work and practice expense are measured for each CPT code based on services provided for the most common, or typical, patient. In many cases, AI algorithms would be developed specifically for atypical patient populations. For example, an AI algorithm that identifies and classifies multiple sclerosis lesions is methodologically different than the base code for brain MRI, which is valued for the typical patient without multiple sclerosis. Creating individual CPT codes for each individual AI algorithm developed for atypical patients would be an impractical and thorny endeavor. Even if CPT codes were developed for AI tools, the valuation process would present additional obstacles. Thus, in our current fee-for-service environment, it is unlikely that AI tools would be reimbursed in the same way as a head CT.

VALUE-BASED PAYMENT MODELS

The passage of the Medicare Access and CHIP Reauthorization Act in 2015 [4] began a transition from exclusive fee-for-service reimbursement to a series of progressively more value-based payment paradigms in the QPP. The QPP includes two pathways: a modified fee-for-service payment program, the Merit-Based Incentive Payment System (MIPS) and the more risk-based alternative payment models. Data science tools have an emerging role in each of these pathways.

MIPS

The MIPS scores eligible clinicians on a 100-point final score scale designed to tie physician payments to performance in four categories: quality, cost, promoting interoperability, and improvement activities. Because most radiologists qualify for special status exemptions from promoting interoperability and many do not meet the attribution threshold for measures in the cost category, the quality category will account for the largest percentage of a clinician's final score (85% of most radiologists' final score for the 2019 performance year; Fig. 1) [5].

Positive or negative adjustments to a clinician's payments are based on the MIPS final score and applied to

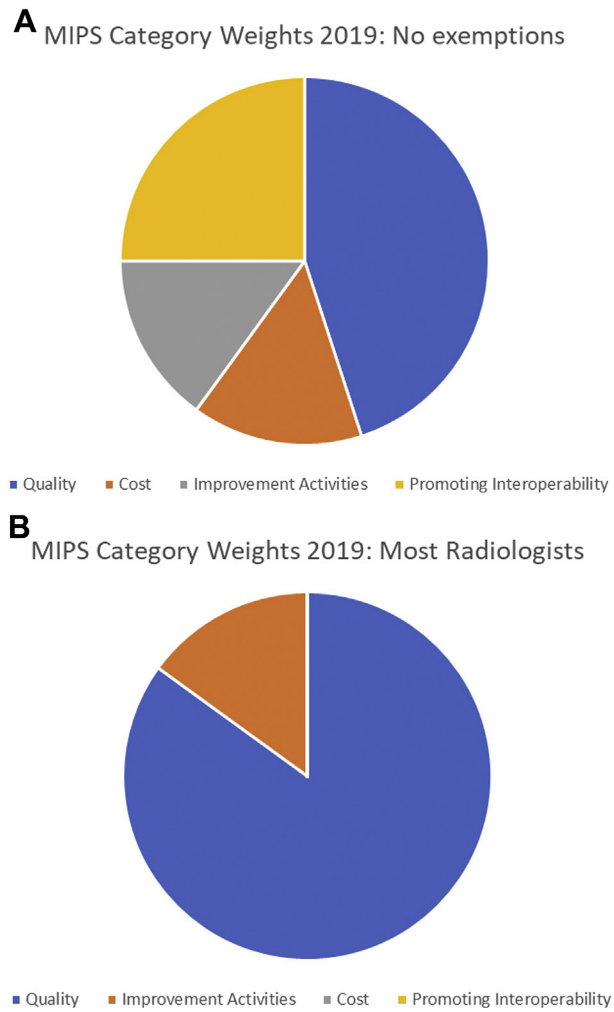

Fig 1. (A) The four performance categories of MIPS and their weighting in 2019. Artificial intelligence tools have a potential role in improving the cost and quality of care and optimizing MIPS scores in each category. (B) The quality category has a dominant impact on MIPS scores for most radiologists. MIPS = Merit-Based Incentive Payment System.

the Medicare payment for Part B claims for that clinician. Because MIPS is a budget-neutral program, the losers effectively pay the winners. The magnitude of the payment adjustments is defined by law and escalates from ±4% for the 2017 performance year, to ±5% for the 2018 performance year, to ±7% for the 2019 performance year, and ±9% for the 2020 performance year and thereafter. The actual adjustment to a clinician's income occurs 2 years after the performance year. To preserve budget neutrality, the actual dollar amount of positive or negative payment adjustments is determined by a scaling factor that is based on a performance threshold chosen on the 100-point final score scale and set by CMS. Not only is income at risk in this program, a clinician or group's reputation is also at risk in MIPS because scores in the program are published publicly on the Physician Compare website.

The four performance categories in MIPS are structured to advance the goals of value-based payment reform: improving quality without increasing cost, reducing cost without changing quality, or reducing cost and improving quality. Currently, the infrastructure to collect and report data for quality measurement is antiquated and costly. CMS has consistently favored reporting mechanisms that utilize electronic data capture, as evidenced by policies awarding bonus points for reporting quality measures using end-to-end electronic reporting and restricting the use of claims-based reporting only to those in small practices. Conceptually, AI tools should be the bedrock of data-driven quality improvement because they enable more precise and larger amounts of data to be incorporated into benchmarks, ensuring higher reliability of measures without the reporting burden of human labor inputs. For example, AI tools could optimize performance on MIPS measure 195 (stenosis measurement in carotid imaging reports) by automating standardized measurement and auto-populating results into radiology reports. Measure 195 was developed to ensure that patients are being measured according to evidence-based North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria and that surgery is appropriately performed for asymptomatic carotid stenosis. By improving performance on this measure, AI tools achieve the primary goal of improving quality of care for these patients, either at the same or reduced cost. The incentive payments associated with higher performance in MIPS is a secondary benefit to AI tools in this example.

MIPS measure 195 is on the low end of complexity regarding the overall ease of automated extraction of the necessary data elements for correct and complete reporting of the measure to CMS. The measure has very few applicable CPT codes that are required to be reported and virtually no denominator exclusions of patients who would be considered outside of the measure objective. An example of a far more difficult measure for automated extraction would be MIPS measure 405, which is an efficiency measure aimed at reducing unnecessary follow-up imaging on incidentally detected lesions in the liver, kidneys, or adrenal glands. This measure specifies strict size criteria applied to the incidental lesions above which follow-up imaging may be indicated and below which follow-up imaging may not be indicated. This size determination requires the radiologist to dictate an actual measurement into a final report for each of these lesions, a task far more time-consuming than it sounds because these incidental lesions are frequent and can be numerous. This is the first layer of complexity when trying to automate documentation and reporting of a quality measure to CMS. The second, and arguably more complex, layer arises when a measure has specific types of patients for which the measure does not apply (so called denominator exclusions). For measure 405, these excluded patients include those with cancer that has metastatic potential or immunocompromised patients with fever. This information is significantly more difficult to automatically extract and often is not included in the radiologist's report.

The barriers to automated extraction can be solved via multiple pathways using AI tools. First, a tool that documents and measures all incidental lesions could ensure that this information is available for extraction. Second, a tool that extracts relevant denominator exclusions from the patient's electronic medical records would ensure proper exclusions were accounted for before calculating measure performance. An alternative approach would be to require radiologists to document denominator exclusions in their report; however, this would require a significant education effort because radiologists would need to know all exclusions for each measure they report. Although this sounds unrealistic, it perhaps is not as daunting if highly structured radiology reports are employed including common data elements (CDEs) [6]. CDEs are discrete structured word descriptions of common findings radiologist encounter during day-to-day analysis of images. The CDE initiative is led by a joint program between RSNA-ACR-American Society of Neuroradiology (ASNR) and the RadElement.org website offers a catalog of radiology CDEs [7]. If these CDEs account for denominator exclusions, a radiologist could pick the CDE that applies to the patient's incidental lesion, properly capturing the required elements for complete reporting of a measure to CMS. Automation of the necessary data for reporting of quality measures would increase the compliance with these measures, presumably translating to higher quality of care while reducing the time and labor costs of obtaining the necessary data for adequate reporting.

The success of value-based payments is inextricably linked to the quality of measures that are used, balanced with the burden of collecting these measures. For the full value of data science tools to be realized in quality measurement, they should be utilized to expand and enrich the outcome measures available in a payment program or model. For example, AI algorithms could improve screening outcomes by increasing the predictability of findings representing cancer and linking radiology findings with pathology and genetics data. Such tools could advance quality measures for recall rate, cancer detection rate, and positive predictive value for biopsy applicable to mammography, lung cancer screening, and CT colonography screening. As the capabilities of AI evolve, new quality measures can be built around these tools based on the value they provide to patient-centered care and improved outcomes.

To have value in the new health care payment paradigm, any new AI algorithm must reduce cost with the same or improved quality, or at least not increase cost with improved quality. The cost category in MIPS will account for 15% of the final score for the 2019 performance year. Although some radiologists may not be directly accountable for cost based on the attribution methodology used in MIPS, the cost measures are intentionally structured to encourage team-based accountability for cost and resource use. This means that every member of the care team must collaborate in reducing waste and redundant services. AI tools hold promise for adding value in this environment by improving diagnostic accuracy and reducing unnecessary examinations and procedures. AI tools focused on radiomics have the potential to substantially reduce variability of our recommendations, as well as revise appropriateness criteria for recommending follow-up imaging, allowing for far greater accountability toward cost than most of the criteria currently contain. For example, an AI algorithm that increases the likelihood of malignancy for biopsies would reduce the number of procedures performed on benign lesions. Likewise, imaging follow-up for incidental findings could be avoided if the risk was low enough according to an AI-generated compendium that includes feature analysis, genetic factors, coexisting imaging findings, and comparison to enormous databases of similar abnormalities. AI tools that scour the Health Information Exchange for any prior studies that could answer the clinical question would reduce redundant imaging. These cost savings will positively impact MIPS cost measures and can be a powerful way for radiologists to demonstrate value in their hospitals, health systems, and clinically integrated networks. Analogous to the quality measure discussion, AI tools have inherent primary value in reducing health care costs and secondary benefits of optimizing incentive payments in MIPS.

ALTERNATIVE PAYMENT MODELS

In addition to MIPS, the Medicare Access and CHIP Reauthorization Act includes a second participation option, Alternative Payment Models (APMs), which incorporate risk and are farther along the continuum in the path from fee-for-service to value-based care. All of the roles for AI tools in MIPS discussed previously are relevant to APMs. The difference is that, instead of value being quantified in discrete units at the measure or performance category level, any tool that reduces cost or improves the health of a population is inherently valuable in APMs. The main focus of APMs is population health, so in turn should be the focus of corresponding AI radiology tools. These tools will certainly take more time to develop and include algorithms that can predict future disease based on current imaging so that early intervention may be deployed to improve the health of the patient while decreasing the total cost of care.

The shift toward population health creates a business case for AI tools that does not exist outside of two-sided risk arrangements in our current reimbursement structure. Data science tools and value-based payment reform are complementary, even symbiotic, in this environment. In the short term, AI has potential to optimize performance in the QPP by making it easier to capture and report quality measures and by improving cost, which will benefit AI users through bonus payments in MIPS and APMs. In the long term, those bonus payments will disappear as more and more practices adopt AI tools and the performance gap decreases. Quality measures that are automated by AI will likely become topped out, with variation in performance too small for meaningful distinctions and improvement in performance to be made. CMS has already removed highly topped-out measures from MIPS and will continue to cap scoring on these measures and remove them from the program. Incentives of MIPS are important but temporary.

AI tools that achieve the primary goals of increasing quality at the same cost, reducing cost at the same quality, or increasing quality while reducing cost will be inherently valuable when payment incentives no longer exist. Investment in AI will then become the price of doing business, particularly in capitated systems where cost savings for a population are paramount. Should AI adoption evolve this way, the structure of value-based payment models will serve as a framework for reimbursement of AI tools that does not exist in the fee-for-service system.

TAKE-HOME POINTS

- In our current fee-for-service environment, it is unlikely that AI tools would be reimbursed in the same way as traditional radiology services.
- AI tools should be the bedrock of data-driven quality improvement because they enable more precise and larger amounts of data to be incorporated into benchmarks, ensuring higher reliability of measures without the reporting burden of human labor inputs.
- As the capabilities of AI evolve, new quality measures should be built around these tools.
- AI tools have inherent primary value in reducing health care costs and secondary benefits of optimizing incentive payments in MIPS.
- The shift toward population health creates a business case for AI tools that does not exist outside of two-sided risk arrangements in our current reimbursement structure.

REFERENCES

1. AMA. The CPT code process. Available at. https://www.ama-assn.org/about/cpt-editorial-panel/cpt-code-process. Accessed November 2018.
2. AMA. RVS Update Committee (RUC). Available at. https://www.ama-assn.org/about/rvs-update-committee-ruc/rvs-update-committee-ruc. Accessed November 2018.
3. Allen B. Valuing the professional work of diagnostic radiologic services. J Am Coll Radiol 2007;4:106-14.
4. US Congress. H.R.2—Medicare Access and CHIP Reauthorization Act of 2015. Available at: https://www.congress.gov/bill/114th-congress/house-bill/2/text. Accessed November 2018.
5. Department of Health and Human Services. Centers for Medicare & Medicaid Services. 42 CFR Parts 405, 410, 411, 414, 415, 425 and 495. [CMS-1693-F, CMS-1693-IFC, CMS-5522-F3, and CMS-1701-F]. RIN 0938-AT31, 0938-AT13, & 0938-AT45. Medicare program; revisions to payment policies under the physician fee schedule and other revisions to Part B for CY 2019; Medicare shared savings program requirements; quality payment program; Medicaid Promoting Interoperability Program; Quality Payment Program—extreme and uncontrollable circumstance policy for the 2019 MIPS payment year; provisions from the Medicare Shared Savings Program—accountable care organizations—pathways to success; and expanding the use of Telehealth Services for the Treatment of Opioid Use Disorder under the Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act Available at: https://s3.amazonaws.com/public-inspection.federalregister.gov/2018-24170.pdf. Accessed November 2018.
6. Rubin D, Kahn C. Common data elements in radiology. Radiology 2017;283:837-44.
7. RadElement.org. Common data elements (CDEs) for radiology. Available at: http://www.radelement.org/. Accessed November 2018.