

SOUNDING BOARD

CMS Innovation Center at 10 Years — Progress and Lessons Learned

Brad Smith, M.Phil.

Ten years ago, the Center for Medicare and Medicaid Innovation (the Center) was created within the Centers for Medicare and Medicaid Services (CMS) to test innovative payment and service delivery models to transform the American health care system from one that pays for volume to one that pays for value. This highly ambitious goal was summarized in the legislation authorizing the Center to develop models that “reduce program expenditures...while preserving or enhancing the quality of care furnished to individuals.”¹ To achieve this bold goal, the Center has been given \$20 billion in funding since its inception, along with the authority to waive certain Medicare and Medicaid requirements.

Since becoming the fourth director of the Center in January 2020, I have spent substantial time analyzing the lessons learned during the past 10 years. My conclusion from this analysis is that value-based care continues to offer promise to transform American health care. However, value-based care will achieve its promise only if the federal government and stakeholders take more aggressive action to prioritize models that can truly achieve savings and improve quality.

RESULTS TO DATE

During the past decade, the Center has achieved many important successes. It has launched 54 models that address critical areas of health care, including primary care, oncology, kidney care, and cardiovascular disease. Nearly 1 million health care providers serving 26 million patients have participated in these models. Value-based care has spread rapidly across the country, with approximately 40% of Medicare fee-for-service payments, 30% of commercial payments, and 25% of Medicaid payments today being made through some form of value-based arrangement.²

The Center’s models have also delivered some positive, tangible results, including five that have resulted in substantial financial savings (Fig. 1). Several models have also produced significant improvements in quality. For example, the Comprehensive ESRD (End-Stage Renal Disease) Care model showed a decrease in emergency dialysis sessions, overall hospitalizations, readmissions, and hospitalizations for ESRD-related complications. Similarly, the Home Health Value-Based Purchasing model resulted in an average improvement of 4.6% in quality scores for home health agencies.

However, the vast majority of the Center’s models have not saved money, with several on pace to lose billions of dollars. Similarly, the majority of models do not show significant improvements in quality, although no models show a significant decrease in quality. In examining these outcomes, we have identified several key lessons.

LESSONS LEARNED

VOLUNTARY PARTICIPATION

To date, 50 of the 54 models have been voluntary for participants. Although voluntary models allow providers to choose whether to participate, the Center has found that it is very difficult for voluntary models to lower costs. Some providers will participate in voluntary models because they strongly support the move to value-based care, but most providers will participate in voluntary models on an ongoing basis only if they believe it is in their economic interest to do so. As a result, the Center has seen most participants drop out of voluntary models when they have substantial or sustained losses.

The Center has provided new up-front payments to participants in several models, in part to en-

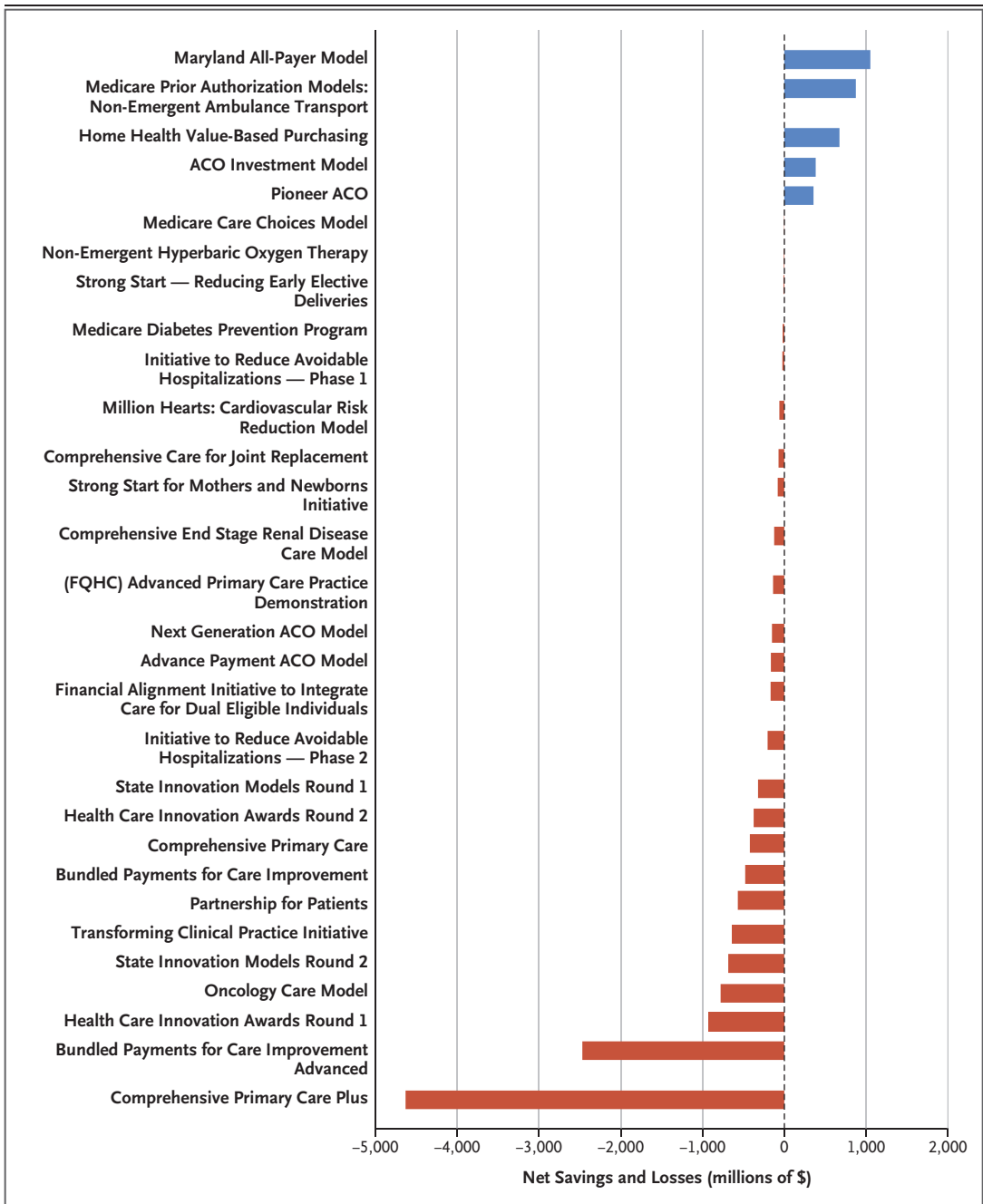


Figure 1. Net Savings and Losses Associated with Selected Models.

Shown are the net savings (in blue) and losses (in red) of 30 models that have been launched in the past decade by the Center for Medicare and Medicaid Innovation and for which results are available from independent evaluations performed by contractors on behalf of the Center as of September 2020. Some of these models are currently active, whereas others have been either discontinued or expanded since the evaluations were performed. The data reflect estimates of savings after the deduction of direct contractual costs for the performance periods. For active models, savings estimates incorporate projections through future performance periods. ACO denotes accountable care organization, and FQHC Federally Qualified Health Center.

courage participation in voluntary models as well as to provide participants with resources to invest in care coordination. For example, in the Comprehensive Primary Care Plus model, participants received a new up-front payment of \$14 to \$28 per beneficiary per month. In the Oncology Care Model, participants received a new up-front payment of \$160 per beneficiary per month. Although both these models have seen robust participation, such additional payments have been part of the reason that neither model has resulted in net savings, with the two models being among the four models with the largest net losses. For example, although the Oncology Care Model showed a modest gross reduction in Medicare expenses before accounting for these up-front payments, the model is projected to have net losses of \$781 million after accounting for the approximately \$900 million in up-front payments made to participants.

BENCHMARKING

Benchmarking is one of the most important aspects of any value-based payment model, since the benchmark determines whether participants saved or lost money. Benchmarks should be designed to fairly and accurately project future costs across an entire population or episode of care (e.g., hip replacement or heart attack). However, projecting future costs is a very difficult technical exercise, since costs can change over time and vary across geographic regions.

For example, take the Comprehensive Care for Joint Replacement (CJR) model, which is a bundled payment model for hip and knee replacements. In the CJR model, the benchmark for each procedure is set according to the average cost of each procedure during the previous 3 years. Across the country, there has been a significant year-over-year reduction in the use of care in a skilled nursing facility after hip and knee surgeries. Because the CJR model uses 3 years of historical data to set the benchmark, the benchmark has not been lowered as fast as the actual cost of these surgeries for providers who are not participating in the model. Largely because of this benchmarking issue, the model is on pace to lose \$72 million. In 2021, the Center plans to finalize a rule to improve the accuracy of the benchmarks for this model.

Small changes in coding practices can also make setting accurate benchmarks difficult. For example, in fiscal year 2017, the coding guidelines of the *International Classification of Diseases, 10th revision*, included small coding changes for both congestive heart failure and pneumonia, two of the highest-volume clinical episodes in the Bundled Payment for Care Improvement (BPCI) Advanced model. As a result of these coding changes, similar patients would probably be classified as having more serious congestive heart failure and pneumonia diagnoses in the treatment period than in the baseline period. Because benchmark prices are appropriately based on the seriousness of a patient's diagnosis, benchmark prices increased in the BPCI Advanced model, which resulted in excess payments to participants. As a result, the model is currently on pace to lose more than \$2 billion. Starting this month, the Center has corrected the BPCI Advanced benchmarks to minimize losses in future years.

Some of the challenges that are mentioned above become more acute when models use prospective rather than retrospective benchmarks. With a prospective benchmark, the Center estimates what the price will be during the following year, whereas with a retrospective benchmark, the Center determines what the price should have been in the previous year. Not surprisingly, retrospective benchmarks are much more accurate than prospective benchmarks, since by the time a retrospective benchmark is set, policymakers know what the trends were in the previous year. However, model participants prefer prospective benchmarks, since they allow participants to know in advance what price they are attempting to achieve. In order to increase participation in voluntary models, the Center has often agreed to set prospective benchmarks. However, the Center is finding that these prospective benchmarks are frequently inaccurate. For example, the Center found that retrospective benchmarks in the BPCI Advanced model would have lowered benchmarks by more than \$600 million during the first 2 years of the model.

Voluntary models also increase the possibility of adverse selection. In many cases, benchmarks are set either entirely or with some consideration for national or regional spending for the population or services in the model. When policymakers

Table 1. Gross and Net Savings and Losses in Selected Models.*

Model	Years Covered	Model Benchmark	Independent Evaluation			
			Gross Savings†	Gross Savings‡	Provider Shared Savings and Performance Payments§	Up-Front Payments¶
<i>millions of dollars</i>						
Comprehensive End-Stage Renal Disease Care	2016–2018	266	115	(172)	—	(57)
Comprehensive Care for Joint Replacement	2016–2017	237	146	(129)	—	17
Oncology Care Model	2017–2018	430	55	(51)	(266)	(262)
Next Generation Accountable Care Organization	2016–2018	499	349	(441)	(25)	(117)
Bundled Payments for Care Improvement Advanced	2018–2019	473	133	(298)	—	(164)

* Estimates of savings are based on currently available data from the Center for Medicare and Medicaid Innovation. Some of these data may be preliminary and subject to change, may cover part of the model, or may cover a partial model year. Losses are indicated in parentheses.

† This amount of gross savings was calculated on the basis of a benchmark used to calculate actual shared savings payments to participants.

‡ This amount of gross savings was calculated by independent evaluation as compared with a control group. If the benchmarks were accurate, the difference between the two gross savings amounts would be close to zero.

§ The amount of payments to providers is based on shared savings or performance, with shared savings being determined on the basis of the model benchmark.

¶ Shown are the payments that were made up-front to providers and beneficiaries regardless of performance (e.g., Monthly Enhanced Oncology Services payments in the Oncology Care Model and beneficiary payments in the Next Generation Accountable Care Organization model). Blank cells indicate that models that did not feature up-front payments.

|| Net savings or losses were calculated by independent evaluation as compared with a control group after the deduction of all payments to providers from an evaluation of gross savings. Net savings or losses do not account for the operating costs of the model (e.g., evaluation and implementation contractors), amounts that are included in the data shown in Figure 1. Values for net savings have been rounded.

set a target using an average, by definition some providers will fall below the average and others will fall above it. In a voluntary model, providers who are below the benchmark have more incentive to participate than those who are above the average. In many cases, national or regional benchmarks combined with adverse selection can make it appear as if participants have saved money even when they in fact have not.

The aggregate challenges of accurate benchmarking become apparent in comparing the results of evaluations of various models that were performed with the use of the Center's benchmarks with results from independent evaluations performed by contractors on behalf of the Center. In theory, if benchmarks were accurate, results for benchmarks and evaluations should be very similar. However, this is not the case, as shown in Table 1. For example, gross savings that are calculated with the use of model benchmarks are on average 235% higher than gross savings

calculated in independent evaluations with the use of a retrospective control group, which indicates that model benchmarks are frequently set higher than actual costs. Independent evaluations of the shared savings payments that result from these inflated benchmarks combined with the up-front payments that are made to participants resulted in net losses in five of the six models, even though the model benchmarks in each of these models showed significant gross savings.

QUALITY

Achieving quality improvements is another important goal of the Center's models. One of the main challenges with measuring quality improvement is that it is often hard to compare quality metrics between participants in the model and a control group. For model participants, the Center can collect quality metrics through claims as well as through surveys and registries that are set up specifically to collect quality data from

model participants. However, only claims-based data are typically available for the control group. As a result, the Center has data for the control group for only approximately 55% of the quality metrics that are used across its models.

Partly as a result of the difficulty of measuring quality in the control group, there is often a lack of alignment between the quality metrics for which participants receive performance-based payments in the models and the quality metrics on which the model is evaluated. For example, across a representative sample of nine models, 71 quality metrics were used to determine performance-based quality payments, but only 39 of those metrics were included in the evaluations. In addition, 99 of the 138 quality metrics that were used in the Center's evaluations were not included in the models' performance-based quality payments.

OPERATING CAPABILITIES

Implementing value-based care models requires a very different set of capabilities than those used in operating fee-for-service programs. For example, in value-based care models, the Center must align beneficiaries and providers, pay participants both capitated and shared savings payments, and provide data back to participants. The Center has learned that these tasks are not always simple, especially given the current CMS technology systems.

For example, in the BPCI Classic model, the Center had a very hard time linking physicians' National Provider Identifiers with particular participant Tax Identification Numbers because the Medicare enrollment files at the time were not up-to-date. These errors resulted in the assignment of providers to the wrong model participants, errors that were so severe that the Center decided to waive downside risk for the first 2 years of the model. This change resulted in net model losses of \$297 million, instead of net savings of \$90 million if the downside risk had not been waived.

Similarly, in the Next Generation Accountable Care Organization model, the antiquated CMS data systems did not apply the correct demonstration code to thousands of claims. As a result of this error, certain providers received payment in full rather than receiving appropriate reductions in their fee-for-service payments, a situation that resulted in losses of approximately \$50 million.

The Center will ultimately be successful only if it is able to launch and scale models that either decrease cost or improve quality. As the Center enters its second decade, there are eight important lessons to keep in mind.

First, the Center must prioritize new models that are most likely to achieve savings or improve quality. Launching new models that move large portions of risk to providers in two-sided risk arrangements and in which the government receives up-front discounts in return for giving participants greater flexibility and upside risk could help achieve this goal. Recent examples of these types of models include the Direct Contracting Global and Professional models and the Direct Contracting Duals and Geographic models.³ The Center must be especially thoughtful when launching models with new up-front payments to providers that are greater than fee-for-service payments, since these additional payments make it very difficult for models to achieve net savings. When initial data show that a model is neither reducing costs nor increasing quality, the Center must modify or terminate the model.

Second, the Center must ensure that it is setting accurate benchmarks in models. At a minimum, benchmarks need to be tested against historical data before being rolled out. When possible, the Center needs to use retrospective benchmarks or — when using prospective benchmarks — create guardrails that allow policymakers to automatically adjust the benchmark if it proves to be inaccurate by a certain amount in a retrospective analysis.

Third, the Center must work to avoid adverse selection in voluntary models. In particular, policymakers must be careful about creating too many options within models that increase the probability that the model will be gamed. In models in which adverse selection is a major risk, the Center needs to consider setting higher discounts in order to ensure savings. In situations where the risk of adverse selection is high enough to potentially make the model ineffective or where it is unlikely that the Center will have enough participants in a model for an adequate test, policymakers must consider making models mandatory.

Fourth, the Center must work to avoid gaming of risk adjustment while simultaneously keeping

in place accurate risk adjustment to ensure that participants are not discouraged from caring for the sickest patients. Ensuring appropriate risk adjustment can be achieved through implementing strategies such as risk-adjustment corridors, coding-intensity adjustments, and zero-sum risk adjustment in models.

Fifth, the Center must work to better align the quality metrics for which participants are paid with the quality metrics that are evaluated. Specifically, it must be possible to evaluate the vast majority of quality metrics for which participants are paid in a model, and the vast majority of quality metrics that are evaluated should be prioritized in the model. To minimize the burden on providers, the Center must align the quality metrics of the model with the quality metrics in other CMS programs whenever possible.

Sixth, the Center must continue to invest in centralized operations and technology infrastructure to decrease errors in model implementation, increase consistency across models, and lower the cost of model operations. To ensure that the Center is spending taxpayer dollars as efficiently as possible, the Center launched a top-to-bottom review of its \$1.4 billion annual operating budget in February 2020. Through this initiative, the Center has already identified more than \$350 million of potential cost savings from streamlining contracts, centralizing common functions, and reducing duplication across core data systems. The Center must also continue its efforts to centralize analytic, operational, and quality resources across the Center as well as continue to invest strategically in technology solutions that can be deployed across models and other CMS programs.

Seventh, the Center must continue to increase the amount of data it shares with participants. Data sharing could include providing standardized data analyses to less sophisticated participants and building application programming interfaces

(APIs) for more sophisticated participants who have the capability to analyze data themselves. When possible, data should be provided in as close to real time as possible, including by developing APIs to share claims data submitted to Medicare administrative contractors with participants in near real time.

Finally, the Center must continue to work to align its payment models with value-based care models of other payers, including Medicare Advantage, Medicaid, and commercial health plans. The more success that the Center has in lowering cost and improving quality in its models, the more likely other payers are to adopt similar models and the more rapidly the Center will be able to expand its models nationally.

CONCLUSION

During the past decade, policymakers at the Center have learned a tremendous amount about value-based care, and the Center still has the ability to deliver on its original promise of lowering costs and improving quality. However, in order to do so, the Center and its future directors must leverage the lessons learned from the past 10 years to modify existing models where necessary and when designing new models.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From the Center for Medicare and Medicaid Innovation, Department of Health and Human Services, Washington, D.C.

This article was published on January 13, 2021, at NEJM.org.

1. 42 U.S. Code § 1315a(a)(1).
2. Health Care Payment Learning & Action Network. APM measurement: progress of alternative payment models. 2019 (<http://hcp-lan.org/workproducts/apm-methodology-2019.pdf>).
3. The CMS Innovation Center home page (www.innovation.cms.gov).

DOI: 10.1056/NEJMs2031138

Copyright © 2021 Massachusetts Medical Society.