The Potential for Cost Savings through Bundled Episode Payments

David M. Cutler, Ph.D., and Kaushik Ghosh, Ph.D.

In the quest to manage the spiraling cost of U.S. health care, one approach has generated great interest. The philosophy behind much current policy — including the Affordable Care Act (ACA) — is that aggregating fee-for-service reimbursement into payments for broader bundles of care will lead to greater efficiency in the provision of care and thus lower costs.

Under the accountable care organization model, perhaps the best-known example of this strategy, medical reimbursements are aggregated to the person-year level. Other programs aggregate reimbursement for episodes of care — for example, care for a particular cardiovascular or orthopedic condition. The Episode of Care Payment Demonstration project, which is authorized by the ACA, requires the Centers for Medicare and Medicaid Services to experiment with bundling Medicare Part A and Part B payments for inpatient care. It resembles the Acute Care Episode payment program, an ongoing demonstration program that bundles Part A and Part B payments for select types of inpatient care episodes.

The central issue in any proposal for aggregating payments is determining at what level services should be bundled together. Episode-based bundled payments are easier for individual physicians or small physician groups to manage, since a given physician is often involved in the full course of a care episode. In contrast, accepting global payments for all of a particular patient’s care generally requires a high degree of integration among multiple physicians. It may be for this reason that episode-based bundled payments seem to save more money than patient-based bundled payments.1 On the other hand, bundling payments for care episodes does not provide incentives to reduce the number of episodes. If limiting the number of episodes of care is a major consideration in reducing costs, bundling care at the patient level would be preferred.

A central issue, then, is the tradeoff between the relative ease of bundling at the episode level and the additional savings incentives from bundling at the patient level. Are a good share of potential savings given up by using only episodes for setting bundled payments?

We set out to estimate the cost savings associated with episode-based and patient-based bundled payments. For episode-based bundled payment, we selected a ran-
dom 5% of the elderly population in fee-for-service Medicare in 2007, assigning the principal diagnosis code of every hospital admission to one of 285 categories in the Clinical Classification System, as determined by the Agency for Healthcare Research and Quality. We considered a new episode to have started when a claim was not preceded by a hospitalization related to the same organ system within 180 days. For example, a hospital admission in March 2007 for myocardial infarction would count as a new episode if there were no related hospitalizations during the prior 6 months. A readmission for heart failure 2 months later would count as part of the same episode. In all cases, rehabilitation services would not qualify as a new episode.

We then assigned to each episode under consideration all inpatient and outpatient spending for the same organ system occurring within 180 days of the episode’s onset (going into 2008 as necessary). Under this definition, a person could have multiple overlapping episodes (for example, a myocardial infarction and a hip fracture within a 6-month period). Some claims, however, were not assigned to an episode — for instance, physician visits not related to a prior hospitalization. We adjusted for geographic differences in prices using data from the Dartmouth Atlas of Health Care for 2007 Medicare reimbursements in each hospital referral region.²

Using these criteria, we arrived at 245 types of episodes. In total, spending on these episodes accounted for just over half of Medicare spending for this sample in 2007. The remainder of spending was for physician and outpatient claims not associated with a hospitalization. Spending was not uniform across episodes but was instead concentrated within a relatively small number. The top 5 and top 17 episode types (see table) accounted for one fourth and one half, respectively, of the costs of the 245 episode types, and three fourths of the spending on these episodes was accounted for by the top 43 conditions.

The most expensive condition was osteoarthritis, and most episodes of care for osteoarthritis involved elective hip or knee replacement. Other high-cost episodes included those for cardiovascular conditions and muscu-
loskeletal conditions (hip and leg fractures). This is not surprising, since the usual candidates for bundled payments are episodes of care for hip fractures and joint replacements and for cardiac care. Individual cancers did not make the top 17 (colon cancer being the most costly, at number 28), though cancer as a whole is very costly.

A cost-distribution analysis for the 17 most expensive episodes shows that the initial inpatient admission accounted for 60% of spending. Subsequent admissions (i.e., readmissions related to the same organ system) accounted for 23%, with separately billed physician services (i.e., procedures and evaluation and management) totaling another 10% and non-inpatient imaging representing 1%. The remainder of spending was for durable medical equipment, home health care, hospice care, and other items. By averaging costs for each condition in the 306 hospital referral regions, we also found that spending varies widely from region to region. For each condition, some regions are high-cost, whereas others are low-cost. To translate this variation into potential dollars saved, we simulated what the spending would be if costs in each area in which the average was above the 25th percentile were brought down to the 25th-percentile level — a feat that might be accomplished if, for example, Medicare paid a bundled rate for episodes and capped it at the level of the 25th-percentile areas. If it did so, the reduction in costs for these 17 conditions would be $10 billion annually. If the cap were set at the 50th-percentile level, the savings would be $4.7 billion. If such caps were applied to all 245 episode types, the annual savings gained from reducing spending to the 25th-percentile level would total $29 billion; reducing it to the 50th-percentile level would save $15 billion.

To estimate the amounts of patient-based global payments, we determined the average costs per patient in each of the 306 hospital referral regions, again adjusting for regional price differences. Setting the patient-based global payment at the level of average spending in the 25th-percentile regions would save $35 billion nationally. If spending were set at the 50th-percentile level, the savings would be $18.2 billion nationally.

These results suggest that an episode-based bundled-payment system could save nearly as much as a patient-based bundled-payment system. Episode-based bundled payments would save 83% of the amount that would be saved with the use of a patient-based bundling system if the 25th-percentile standard were used and 82% if the 50th-percentile standard were used. The reason for this similarity is that there can be substantial heterogeneity in spending for different types of episodes within a given region. Some regions have high spending for certain conditions even if they have low spending overall. Episode-based payment encourages efficiency in treating the conditions on which spending is high, regardless of whether the region as a whole is low-cost. Patient-based payment, by contrast, achieves no additional savings if the region as a whole is not high-cost.

In summary, our results suggest that it is possible to achieve very substantial health care savings by moving from a fee-for-service model to bundled payments for episodes of care, whether in a stand-alone program or as a component of an overall global-payment model. The primary limitation of our analysis is that we haven’t accounted for heterogeneity in the complexity of disease within episode types that may affect the average costs in a region. We suspect that such variations in average complexity are small, however, so that our results are an accurate estimate of cost savings from episode-based bundled payments.

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