

CHAPTER 14

---

**Mandated report:  
Designing a value incentive  
program for post-acute care**

---



# Mandated report: Designing a value incentive program for post-acute care

## Chapter summary

The Consolidated Appropriations Act, 2021, requires the Commission to report on a prototype value-based payment program under a unified prospective payment system (PPS) for post-acute care (PAC) services and analyze the impacts of the prototype's design. The report is due March 15, 2022. Although this chapter does not make formal recommendations, it has a strong foundation in the Commission's past work and recommendations on value incentive programs.

Building on the Commission's past work, we present key design elements for a PAC value incentive program (VIP). For each of the following elements, policymakers would need to make decisions to develop and implement a PAC VIP.

- **Small set of performance measures.** The PAC VIP would adjust payments based on provider performance on a small set of measures tied to clinical outcomes, patient experience, and resource use. Policymakers would need to decide whether all providers should be scored on the same set of measures and which measures should be scored.
- **Strategies to ensure reliable measure results.** The PAC VIP's measure results would reflect true differences in performance and not random

## In this chapter

- Elements of a value incentive program for post-acute care
- Results of our illustrative model of a PAC VIP design
- Other ways to encourage providers to improve performance
- Implementing a PAC VIP is a complex undertaking

variation. Policymakers would need to define the reliability standard for measure results and determine which strategies will ensure reliable results for as many providers as possible.

- **System to distribute rewards with minimal “cliff” effects.** The PAC VIP would use a simple scoring approach that awards points for every level of performance achieved. Policymakers would need to decide whether a provider should meet some minimum performance standard before it earns performance points that translate into a reward.
- **Approach to account for differences in patients’ social risk factors using a peer-grouping mechanism, if necessary.** If higher social risk is tied to poorer outcomes, the PAC VIP would stratify providers into peer groups based on the social risk of their patient populations. Under this grouping mechanism, providers in peer groups with patient populations at high social risk would receive larger adjustments for attainments in quality compared with other providers. Policymakers would need to decide how to define and measure patient populations’ social risk to establish the peer groups, as well as how many peer groups would be needed to meaningfully differentiate providers.
- **Method to distribute the entire provider-funded pool of dollars.** The PAC VIP would redistribute all withheld funds to providers based on their performance. Policymakers would need to determine the size of rewards and penalties needed to motivate providers to improve performance.

For illustrative purposes, we modeled a PAC VIP design that includes these design elements and adjusts each provider’s payments based on its performance. Approaches taken for four of the elements could be readily incorporated into a design—a starter set of performance measures, the reliability standard, a scoring methodology, and the distribution of incentive payments. However, questions remain about an approach to account for the social risk of a provider’s patient population. Although there is a conceptual relationship between the share of fully dual-eligible beneficiaries (beneficiaries eligible for both Medicare and Medicaid, a proxy for low income) a provider treats and its outcomes, we did not find an empirical association in each of the four settings. More work is needed to define a measure of social risk that considers multiple dimensions before concluding whether adjusting performance results for social risk is always needed.

Implementing a PAC VIP would involve many steps and would be a multiyear endeavor. First, a PAC PPS would need to be implemented so that setting-specific practice patterns (such as length of stay) begin to converge.

Concurrently, CMS would need to begin aligning regulatory requirements for PAC providers. Until this process is completed, providers' performance would likely be compared only within each setting because current practice patterns reflect current regulatory requirements and the payment incentives embedded in the various PPSs. Setting-specific comparisons of performance would be phased out over time, leading up to comparisons of performance regardless of setting.

CMS would need to select a set of performance measures that captures differences across providers. There will be trade-offs between using common measures and using patient population-specific measures. In addition, the measure set should evolve to include accurate measures of the maintenance and improvement in patients' functional status and of patient experience. CMS would need to test a measure of social risk that has both a conceptual relationship and an empirical association with outcomes. CMS should explore the use of geographic area-level measures of social risk and whether they are accurate proxies for the social risk of individual patients.

Finally, CMS would need to design a methodology that scores providers' performance, ensures reliable measure results, distributes rewards with minimal cliff effects, accounts for differences in the social risks of a provider's patient population through peer grouping if necessary, and fully redistributes provider-financed incentive payments to providers. The Commission's PAC VIP model would be a good starting point for CMS's deliberations. ■



---

## Background

---

Post-acute care (PAC) providers—skilled nursing facilities (SNFs), home health agencies (HHAs), inpatient rehabilitation facilities (IRFs), and long-term care hospitals (LTCHs)—offer Medicare beneficiaries a wide array of services, ranging from recuperation and rehabilitation services to hospital-level services. The Commission and others have documented overlap of the many types of patients treated in the four settings, though the amount of overlap differs by clinical condition (Gage 2012, Medicare Payment Advisory Commission 2019a). For example, the treatment of patients recovering from stroke is relatively broadly distributed across the four settings (though some of this would be explained by differences in the severity of the stroke), whereas the treatment of patients who require ventilator care is concentrated in LTCHs. Several factors account for the overlap in treatment settings: The supply and use of PAC varies across the country; there are no clear criteria identifying which patients need PAC (and how much); and there is a dearth of evidence-based guidelines to direct beneficiaries to the setting with the best outcomes (Medicare Payment Advisory Commission 2014). Reflecting these ambiguities, Medicare per capita spending for PAC varies geographically more than for any other type of service (Institute of Medicine 2013, Medicare Payment Advisory Commission 2017b).

Two recent trends may illustrate the potential overlap in care furnished by PAC providers. First, providers participating in alternative payment models (such as CMS's bundled payment initiatives and accountable care organizations) have shifted patients away from institutional PAC settings (IRF and SNF) and increased the share of patients treated in HHAs, without eroding quality of care (Agarwal et al. 2020, Marrufo et al. 2021, Navathe et al. 2020). Second, during the coronavirus public health emergency (PHE) in 2020, beneficiaries avoided SNFs and were treated elsewhere. Between 2019 and 2020, of the top conditions discharged from hospitals and referred to PAC, the shares treated in SNFs dropped, while the shares going to HHAs and, to a lesser extent, IRFs rose.

### **A unified payment system for PAC providers**

Despite the overlap in patients, Medicare uses separate prospective payment systems (PPSs) for each setting,

which results in considerably different payments for similar patients. To establish site-neutral payments based on patient characteristics rather than setting, the Congress requested that the Commission and the Secretary of Health and Human Services develop prototypes for a unified PAC payment system for all PAC providers. To meet the mandate, in 2016 the Commission recommended design features for a unified PPS for paying PAC providers and concluded that a unified payment system was feasible (Medicare Payment Advisory Commission 2016). The Commission is required to submit a second report once the Secretary has issued its report on a prototype design. The Commission has work underway to update its analyses and currently plans to submit the second mandated report in 2023, assuming the Secretary's report is issued by the end of 2022.

The recommended design elements include a uniform unit of service (a stay), outlier policies for unusually short or unusually high-cost stays, and a common risk-adjustment method that would raise or lower payments depending on the patient's condition, comorbidities, and other factors. The Commission noted one design feature that cannot be uniform: The base payment rate for home health care needs to be lower to reflect this setting's considerably lower cost. Otherwise, HHAs would be substantially overpaid and institution-based care would be substantially underpaid. An adjustment would help ensure that placement decisions are based on a patient's care needs, not payment incentives. Subsequent to the 2016 report, the Commission recommended that a PAC PPS be phased in over multiple years and that aggregate payments be lowered to more closely align payments with costs (Medicare Payment Advisory Commission 2017a). In addition, our analysis of consistency in recording functional assessment data raised questions about the use of this information in establishing payments (Medicare Payment Advisory Commission 2019a).

Because a unified PAC PPS would establish a common payment system, Medicare's existing setting-specific regulations would need to be aligned before the new PPS is fully phased in. Otherwise, PAC providers would continue to face different staffing and licensing requirements—and the associated costs—for treating similar patients. The Commission suggests a two-tiered regulatory approach (Medicare Payment Advisory

## Mandate to establish a prototype value-based payment program under a unified prospective payment system for post-acute care services

### Consolidated Appropriations Act, 2021

Not later than March 15, 2022, the Medicare Payment Advisory Commission shall submit to Congress a report on establishing a prototype value-based payment program under a unified prospective payment system for post-acute care services under the Medicare program under title XVIII of the Social Security Act (42 U.S.C. 1395 et seq.). Such report—

(1) shall—

(A) consider design elements such as—(i) measures that are important to the Medicare program and to beneficiaries under such program; (ii) methodologies for scoring provider performance and effects on payment; and (iii)

other elements determined appropriate by the Commission; and

(B) analyze the effects of implementing such prototype program; and

(2) may—

(A) discuss the possible effects, with respect to the Medicare program, on program spending, post-acute care providers, patient outcomes, and other effects determined appropriate by the Commission; and

(B) include recommendations with respect to such prototype program, as determined appropriate by the Commission, to Congress and the Secretary of Health and Human Services. ■

Commission 2019a). PAC providers would be required to meet a common set of requirements that would establish the basic provider competencies to treat the average PAC patient. Providers opting to treat patients with specialized or very high care needs—such as those who require ventilator support or high-cost wound care—would need to meet a second tier of requirements that would vary by the specialized care need. The basis of provider requirements would thus shift from the setting of care to the care needs of the patients a provider opts to treat.

Current practice patterns (most notably, lengths of stay) differ considerably across PAC settings, reflecting differences in setting-specific regulations and payment systems. Until requirements are aligned, provider performances under a PAC value incentive program (VIP) would need to be compared within each setting. However, we expect that lengths of stay would begin to converge as providers face the same regulations and payment incentives under a unified PPS. Over time, the PAC VIP could transition from comparing providers within a setting to looking across all PAC providers.

A unified PAC PPS is not the end point for payment reform; rather, it would represent a necessary first step in a longer-term restructuring of how Medicare should pay providers. The Commission believes that, ultimately, Medicare needs to move away from fee-for-service (FFS) payment systems and toward alternative payment models and population-based payments. These arrangements would put providers at risk for all health care spending and outcomes. In the interim, it is essential that Medicare payments are accurate, based on patient (and not setting) characteristics, and tied to provider performance. A VIP would reward providers for achieving good outcomes for their patients and penalize providers with worse performance.

### A value incentive program for PAC providers

A VIP is an essential complement to the implementation of a PAC PPS because payments would continue to be paid on an FFS basis. FFS payment does not include incentives to furnish high-quality care (when doing so raises a provider's costs) and



encourages unnecessary utilization. By tying a portion of payments to measures of quality and resource use, a VIP would create incentives for providers to furnish efficient (low-cost, high-quality) care to FFS beneficiaries. When providers are subject to a common payment system and similar regulatory requirements, a single VIP should accompany it. Performance can then be compared across all providers.

Recognizing the importance of a companion VIP for a unified PPS, the Congress, in the Consolidated Appropriations Act, 2021, required the Commission to report on a prototype value-based payment program under a unified PPS for PAC (see text box on the mandate).

Typically, VIPs adjust a provider's Medicare FFS payments based on performance on measures tied to clinical quality, patient experience, and resource use. Providers with good performance receive higher payments, while providers with poor performance receive lower payments. A provider's performance during an assessment period is compared with that of other providers or with some performance scale and then converted to a provider-specific payment adjustment. This adjustment is then applied to all Medicare FFS payments for that provider in a subsequent fiscal year. Because the payment adjustments affect Medicare FFS payments, the measures generally do not consider performance for other patients.

Ideally, in a uniform PAC VIP, performance would be compared across settings using the same measures for at least a core set of measures. However, because the payment systems and regulatory requirements are distinct for each setting, current practice patterns vary considerably across settings. Therefore, at least initially, performances under a PAC VIP would need to be compared within each setting using a uniform set of measures. Once practice patterns (such as length of stay) converge, comparisons across settings could be made.

In this chapter, we discuss design elements of a PAC VIP that are consistent with the Commission's principle that Medicare payments should not be made without considering quality of care (Medicare Payment Advisory Commission 2018). Although this chapter does not make formal recommendations, it

has a strong foundation in the Commission's past work and recommendations on value incentive programs. The Commission has previously applied its principles for Medicare quality incentive programs to the design of programs for hospitals, Medicare Advantage plans, and SNFs (Medicare Payment Advisory Commission 2021b, Medicare Payment Advisory Commission 2020, Medicare Payment Advisory Commission 2019b).

---

## Elements of a value incentive program for post-acute care

---

Relying on the Commission's principles for quality measurement and our previous work on redesigning Medicare quality incentive programs, we discuss key design elements of a PAC VIP. The design elements include:

- a small set of performance measures;
- strategies to ensure reliable measure results;
- a system to distribute rewards with minimal "cliff" effects;
- an approach to account for differences in patients' social risk factors using a peer-grouping mechanism, if necessary; and
- a method to distribute the entire provider-funded pool of dollars.

For each PAC VIP element, policymakers would need to make development and implementation decisions.

### Small set of performance measures

Medicare quality programs should include a small set of performance measures tied to outcomes, patient experience, and resource use. In developing the PAC VIP, policymakers would need to decide whether all providers should be scored on the same set of measures and which measures should be scored.

### Should all providers be scored on the same set of measures?

The PAC VIP could score all providers on the same set of performance measures, such as hospitalizations during the stay, successful discharge to the community,

and Medicare spending per beneficiary (MSPB). In our illustrative model, we chose to score PAC providers on these three performance measures.

Alternatively, the PAC VIP could include a combination of common measures and measures tailored to specific patient populations. For example, in addition to the common measures, the performance of providers that treat a sufficient number of patients on ventilators could be gauged using a measure of successful ventilator weaning. Including both broad and patient population-specific measures could capture a more comprehensive view of a provider's care. However, using different measures for different providers would limit the ability to compare performance across providers. Some providers would be scored only on the common measures, while others would be scored on a combination of common measures and measures specific to a subgroup of patients treated by a provider. Measures that evaluate a subgroup of a provider's populations could have sample size and reliability issues. CMS could also choose to score the same common measures in the PAC VIP and publicly report additional patient population-specific measures that are relevant for each provider.

### **What measures should be used to gauge provider performance?**

In our illustrative PAC VIP model, we evaluated performance using three measures: all-condition hospitalizations within stay, successful discharge to the community, and MSPB. These measures are relevant to all PAC providers and are important to beneficiaries, the Medicare program, and entities such as accountable care organizations (ACOs) and health systems interested in establishing networks of high-performing providers. The measures capture different dimensions of PAC: hospitalizations during the PAC stay; admissions and deaths in the period following a PAC stay; and Medicare spending during and after the PAC stay. The measures also hold providers jointly accountable for good outcomes. For example, if a PAC provider refers a beneficiary to another PAC provider for additional care, the successful discharge to the community measure creates incentives for the first provider to refer beneficiaries to subsequent providers that have low hospitalization, mortality, and MSPB rates. These measures capture provider behavior in response to broad environmental factors such as

participation in alternative payment models (e.g., ACOs) and increasing Medicare Advantage (MA) penetration, which may encourage use of lower-cost PAC.

We developed measures that use uniform definitions and risk adjustment across the PAC settings. We can calculate the measures using already reported claims data. All three measures have considerable variation in performance within each setting, suggesting opportunities for providers to improve and the ability to differentiate performance among providers.

We did not include process measures in our illustrative model because of the Commission's established principle that quality payment programs should use measures tied to outcomes, patient experience, and value. Process and other more granular measures may be important for public reporting, but they are not outcome measures that should be tied to payment. CMS should continue to use other quality measures and compliance standards to monitor PAC provider performance and publicly report this information.

CMS needs to fill in gaps in the availability of key performance measures so they can be included in a PAC VIP—most notably, the maintenance or improvement in function and patient experience. Therefore, we expect a PAC VIP measure set would evolve as other data and measures became available. Past work by the Commission raised serious questions about the current state of the functional assessment data. Because this information affects payments for HHAs, SNFs, and IRFs and the calculation of certain quality metrics, providers have an incentive to report the information in ways that raise payments and appear to improve performance (Medicare Payment Advisory Commission 2019a). Evaluations of the home health value-based purchasing (VBP) program also raised questions about the recording of patient assessment information (Pozniak et. al. 2021). In the Commission's June 2019 report to the Congress, we discuss strategies to improve assessment data reporting, the importance of auditing and monitoring the reporting of these data, and alternative measures of function that do not rely on provider-completed assessments (Medicare Payment Advisory Commission 2019a).

Another strategy to gather functional assessment information would be to require hospitals to assess PAC-bound patients at discharge from the hospital.

## Measures of patient experience should be developed for post-acute care users

Medicare's quality payment programs should include measures of patient experience. Across the health care system, research finds that improving patient experience translates to better health. Patients who feel heard and have positive care experiences report better health outcomes and are more likely to adhere to treatment plans (Agency for Healthcare Research and Quality 2020a).

Implementation of patient experience surveys across post-acute care (PAC) settings is limited. CMS has implemented a Home Health Consumer Assessment of Healthcare Providers and Systems (CAHPS®) survey to capture the experiences of beneficiaries receiving home health care.<sup>1</sup> The Agency for Healthcare Research and Quality and CMS have developed CAHPS surveys for short-stay skilled nursing facility, inpatient rehabilitation facility, and long-term care hospital patients, but CMS does not require these providers to administer and report survey results from the beneficiaries they treat. CMS could explore using these existing surveys as the basis for a uniform PAC patient experience survey.

A PAC patient experience survey could include core questions concerning all types of PAC providers (such as asking for an overall rating) and separate supplemental questions for institutional providers (such as rating the facility's cleanliness) and home health agencies (such as discussion of home safety). Given the high level of comorbidities and cognitive impairments among PAC patients, collecting patient experience surveys could often require the use of proxies. To implement a PAC patient experience survey, CMS would need to finalize surveys and develop patient experience measures based on survey responses that are adjusted for respondent characteristics (e.g., sex, age, education, whether a proxy completed the survey). CMS would also need to implement a process for third-party survey vendors to collect survey results from patients (or their proxies). Collecting patient experience information would add burden to providers and CMS, but the Commission contends that these are valuable measures to assess a provider's quality of care. ■

While all SNF users and most IRF and LTCH users have a prior hospital stay, the majority of home health users do not, making this a less viable option for a PAC VIP. Alternatively, CMS could gather patient-reported outcomes, although none are currently collected in PAC settings or included in PAC quality reporting programs. Further, many PAC patients have a high severity of illness and cognitive impairments that would affect the ability to collect accurate patient-reported information. The use of proxies to gather this information would need to be an integral part of developing this option.

Another important measure of provider performance is patient experience. However, currently there are no uniform patient experience surveys for PAC users

(see text box). In addition, Medicare does not collect medical record or electronic clinical data (e.g., lab results) that would allow calculation of some clinical outcome measures.

### **All-condition hospitalizations within stay**

Hospitalizations (admissions and readmissions) are outcomes that are disruptive to patients and caregivers, costly to the health care system, and put patients at additional risk of hospital-acquired infections and complications. Hospitalizations are also a major source of patient and family stress and may contribute substantially to the loss of function, particularly in older patients.

For our illustrative PAC VIP model, we calculated uniform, risk-adjusted hospitalization within-stay

rates for PAC providers, using three years of claims data (2015 to 2017). This measure holds providers accountable for hospitalizations that occur during the entire PAC stay. In addition to counting inpatient readmissions, the measure includes returns to the hospital for outpatient observation stays. The risk-adjustment model includes the following information: the beneficiary's primary reason for treatment; severity of illness, comorbidities, age, sex, and original reason for Medicare entitlement; whether the beneficiary received special care (ventilator use or wound care) during the preceding hospital stay or during the PAC stay; the length of the preceding hospital stay and the number of intensive care unit days (if any); and the number of hospitalizations during the past year. Beneficiaries who died during their PAC stay are excluded from the measure calculation, while beneficiaries who are enrolled in hospice are included.

Until there are aligned regulatory requirements and a uniform payment system (with uniform incentives), the hospitalization rates will differ across settings. Not only are IRFs and LTCHs licensed as hospitals (so we would expect them to transfer fewer beneficiaries to acute care hospitals), but lengths of stay (during which a beneficiary could be hospitalized) currently vary more than two-fold across settings. For example, in 2017, the average lengths of stay were 25 days in SNFs and 12.7 day in IRFs.

**Successful discharge to the community** Discharge to a community setting is an important health care outcome for many patients for whom the overall goals of PAC include optimizing functional status and returning home. However, providers should not discharge patients who are not medically ready to return to the community because doing so may result in hospital events. Also, as noted above, when patients need additional care following a PAC stay, providers should have an incentive to refer them to subsequent providers that have low hospitalization and mortality rates. Unlike the hospitalization within-stay measure, successful discharge to the community captures a patient's outcomes after discharge from the PAC setting.

Building on CMS's specification, we developed a risk-adjusted measure of successful discharge to the community that uses a uniform approach to adjust for differences in the mix of patients treated by a provider.

Successful discharge to the community is defined as having been discharged from the PAC provider to the community and having no unplanned hospitalizations or mortality in the next 30 days. "Community" is defined as home/self-care, with or without home health services, and includes nursing home residents who return to the same facility.<sup>2</sup> Discharges to hospice or resident stays with a hospice benefit in the postdischarge window are excluded from the calculation.

For our illustrative model, we calculated uniform, risk-adjusted results on the measure of discharge to the community for the four types of PAC providers using three years of claims data (2015 to 2017). The risk-adjustment model included the following factors: the beneficiary's primary diagnosis, comorbidities, age, sex, and original reason for entitlement; the length of the preceding hospital stay (if there was one) and the special care (ventilator or dialysis) provided during it; and the number of acute hospital stays during the past year.

**Medicare spending per beneficiary** The MSPB for PAC (MSPB-PAC) is a provider-level measure of resource use that captures Part A and Part B Medicare spending during a patient's PAC stay and the following 30 days. Low MSPB-PAC is considered desirable. To keep per beneficiary spending low, the PAC provider has an incentive to furnish high-quality care (e.g., avoid hospitalizations), make referrals for the necessary level and amount of subsequent care, ensure safe transitions between care settings, and discharge beneficiaries to high-quality PAC providers. The measure helps create incentives for providers not participating in broad delivery reforms (such as accountable care organizations and bundled payment programs) to focus on an episode of care that begins with admission and extends for a period after discharge. For beneficiaries who are hospitalized and then use SNF services, the measure overlaps with the MSPB measure for hospitals (which holds hospitals accountable for spending during the hospital stay and 30 days after discharge). By having overlapping measures, PAC providers and hospitals have the same incentive to keep resource use low. Paired with outcome measures, the MSPB-PAC measure could also detect stinting on care by identifying providers with consistently low spending per beneficiary and low quality.



Building on CMS's specification, we developed a risk-adjusted measure of spending that uses a uniform approach to adjust for differences in the mix of patients treated by a provider. Using three years of claims data (2015 to 2017), we calculated the risk-adjusted MSPB for each PAC provider relative to the setting average. The risk-adjustment model includes the following factors: the beneficiary's primary diagnosis, comorbidities, age, sex, and original reason for entitlement; whether the beneficiary had ESRD, was in a long-term care institution, or was enrolled in hospice; the timing of the stay (e.g., whether it immediately followed a prior hospital stay or followed a prior PAC stay); and the length of stay in an intensive care or coronary care unit during a prior hospital stay.

### **Strategies to ensure reliable measure results**

The measure results used in the PAC VIP should be reliable, meaning that they should reflect true differences in performance and not be attributable to random variation. Key decisions for policymakers include defining the reliability standard for measure results and selecting the strategies to ensure reliable measure results for as many providers as possible.

#### **What reliability standard for measure results should be used?**

A high reliability standard should be used to determine the minimum number of stays required for a provider's performance to be scored in a PAC VIP. For providers with low patient volume, establishing reliable measure results is problematic because they do not have enough observations to ensure that the measure detects signal (actual performance) rather than noise (random variation). Unreliable measure results can lead to the wrong conclusions about a provider's performance; a low-volume provider can appear to have unusually good or poor performance when in fact its performance is not statistically different from the average (Garrett et al. 2021). Low-volume providers are also more likely to have performance that varies from year to year, which could result in a provider incurring penalties one year and receiving a reward the next.

In our illustrative PAC VIP model, we used a minimum case count that resulted in an acceptable reliability for each measure (i.e., 0.7, meaning that 70 percent of the variance in a measure's results was attributable to

actual performance differences and that providers can be differentiated).<sup>3</sup> This level of reliability required a minimum of 60 stays (for each measure). Because there are many small SNFs and HHAs, this requirement had more effect on them than on IRFs and LTCHs, which tend to be larger.<sup>4</sup>

#### **What strategies will ensure reliable measure results for as many providers as possible?**

Setting a minimum case count to ensure reliability inevitably means excluding some providers from the quality measurement program. One way to include as many providers as possible is to pool data across years, allowing a performance measure to be calculated for many small providers that would otherwise be excluded. Such pooling is consistent with other quality payment program designs and measures. For example, Medicare's Hospital Readmissions Reduction Program uses three years of performance data to calculate readmission results. In our illustrative PAC VIP model, we pooled three years of claims data to increase the number of observations for each provider. Blending performance across years also encourages sustained high quality. However, pooling data across years could dampen a provider's drive to improve if their recent better results are blended with older, poorer performance. In such a case, the provider's improved performance would not be fully recognized in its payment incentive payment for several years. To counter this disincentive, policymakers could weight the more recent years more heavily. Policymakers could also pool data across years only for low-volume providers, while scoring just the most recent year's performance for providers that meet a minimum count in a single year.

#### **System to distribute rewards with minimal "cliff" effects**

Consistent with the Commission's principles, a PAC VIP should reward or penalize a provider using a continuous, prospectively set scale for each measure. By recognizing every level of performance, providers are always better off improving quality to achieve a higher level of quality—thus negating the need to separately score improvement. Further, the approach enables providers with similar performances to earn, all else being equal, similar payment adjustments. In contrast, a scoring approach that includes "cliffs" (preset numeric thresholds) can result in providers

with similar performances receiving markedly different payment adjustments because of where a performance falls in relation to a preset “cut point.” A provider’s performance that is scored just above the cut point could receive a sizably larger payment adjustment compared with the payment adjustment for another provider’s performance that is scored just below the cut point.

The performance scale for each measure should be set nationally, because as a national program, Medicare should apply the same performance scale to all providers. Medicare should not have different expectations for quality based on a provider’s location. The scale should be prospectively set so providers know how their performance on a measure translates to points before the payment year. Knowing the scale ahead of time allows providers to set their improvement goals and activities.

In our illustrative PAC VIP model, we established a continuous, prospectively set scale for each measure. We scored each PAC provider on its performance on each performance measure against national, setting-specific scales. We compared providers within a setting because the considerable variation in performances across settings reflects, in part, setting-specific requirements and payment policies. For example, some providers are licensed as hospitals, so they are less likely to have admissions to acute care hospitals during the PAC stay. Also, CMS criteria for IRFs and LTCHs currently limit the types of patients they admit. After implementation of a unified PAC PPS with consistent payment rates and regulations, we expect differences in practice patterns and costs to narrow. At that point, transitioning to common performance targets could be appropriate.

### **Should a provider meet some minimum performance standard before it earns a reward?**

A key decision for policymakers in developing the PAC VIP is whether a provider should meet some minimum performance standard before it earns performance points that translate into a reward. This criterion would prevent providers with relatively poor performance from earning a reward. One way to accomplish this goal would be to set a performance-to-points scale so that no points are assigned below a minimum threshold. A minimum threshold could be set based on clinical judgment where there is an applicable clinical standard.

For example, clinical definitions of “controlled diabetes” could be used to set a threshold for a measure gauging a provider’s success at managing diabetes. However, for some outcome measures, there may be no clinical standards. For example, even with a goal to keep beneficiaries out of the hospital, some beneficiaries need to be rehospitalized to receive appropriate care. For such measures, policymakers could use a relative minimum threshold—for example, the worst quartile of performers—so that providers in that cohort would not receive points.

Setting a minimum performance threshold would help meet beneficiaries’ and the program’s reasonable expectations that providers furnish some minimum level of quality. It would also prevent the worst-performing providers from earning performance points that could translate into a reward (or, more likely, a smaller penalty).

Although a minimum threshold would, in principle, avoid rewarding the poorest performers, there are several reasons not to include one in a scoring design. First, it would create a cliff, or numeric threshold, between providers whose performance falls just below and those just above the threshold. In addition, a minimum threshold would disproportionately penalize providers who treat a high share of patients at high social risk because they are more likely to have lower performance on quality measures. Under the PAC VIP, the lowest-performing providers would always be penalized, regardless of their share of beneficiaries at high social risk, because the design establishes “winners” and “losers” within each peer group. Finally, a threshold would undercut the purpose of a peer-group strategy that is designed to counter the disadvantages these providers face in achieving good performance. Preventing the lowest-performing providers from earning any points would create even larger disparities between the lowest-performing and other providers. The disparity would result from the dollars withheld from the lowest-performing providers being redistributed to the other providers, raising these other providers’ incentive payments (or reducing their penalties).

In designing a PAC VIP, the Commission aims to increase the equity across providers when tying performance to value incentive payments. Therefore, despite the merits of including a minimum

performance threshold, the Commission comes to a different conclusion and supports an approach that counters the challenges that providers treating high shares of patients at high risk have in achieving good performance. We did not include a minimum standard in our illustrative PAC VIP model.

### **Approach to account for differences in patients' social risk factors using a peer-grouping mechanism, if necessary**

Providers that treat a large share of patients with social risk factors may be relatively disadvantaged in a quality payment program because it may be harder for them to achieve good outcomes for their patients. Thus, a quality payment program should account for differences in the providers' patient populations to counter the disadvantages they could face in achieving good outcomes.

Rather than adjusting performance measures for patients' social risk factors, which can mask disparities in performance, Medicare should make adjustments to payments based on a provider's performance compared with its peers (Medicare Payment Advisory Commission 2018). With peer grouping, each provider's performance is compared with providers with similar mixes of patients at high social risk (that is, its "peers") to determine rewards or penalties based on performance. A provider would earn points based on its performance relative to setting-specific national performance scales, but how those points are converted to incentive payments would vary by peer group, with larger multipliers (i.e., the payment adjustment per point) for peer groups with higher shares of beneficiaries at high social risk. Providers would know the performance scales, their peer-group assignment, and peer-group multipliers before the payment year so that they would have time to set their improvement goals and activities. Key decisions for policymakers when implementing peer grouping include how to define and measure the social risk of patient populations to set the peer groups and how many peer groups would be needed to differentiate providers.

### **How should the social risk of a provider's patient population be defined and measured?**

Social factors such as income, housing, social support, transportation, and nutrition affect access to health services and desired health outcomes (Office of

Disease Prevention and Health Promotion 2021). The National Academies of Sciences, Engineering, and Medicine (NASEM) outlined considerations to determine whether a social risk factor (measure) should be accounted for in a Medicare quality payment program (National Academies of Sciences 2016b). The social risk factor should have a *conceptual relationship* with the outcome of interest (that is, there should be a reasonable hypothesis positing how the social risk factors could affect a Medicare beneficiary's health outcome) and *empirical association* (that is, there should be verifiable evidence of an association between the social risk factor and the outcome of interest). This consideration is consistent with the Commission's principle that the Medicare program should take into account, as necessary, differences in a provider's patient population, including social risk factors. NASEM notes that research and experience would inform whether there is a reasonable conceptual basis for expecting a systematic relationship. It acknowledges that a conceptual relationship may not be consistent over time or across settings. An empirical association confirms the conceptual relationship; policymakers would need to decide how strong the association needs to be before peer grouping is undertaken.

Medicare beneficiaries who are disabled or low income are eligible to enroll in Medicaid. In our illustrative PAC VIP model, we tested a share of a provider's patients who were fully dual eligible for Medicare and Medicaid as a measure of social risk because there is a conceptual relationship between dual eligibility and our outcomes of interest. There is a clear and established relationship between poverty, socioeconomic status, and health outcomes—including increased risk for disease and premature death (Office of Disease Prevention and Health Promotion 2021). Compared with other beneficiaries, dual-eligible beneficiaries are more likely to report being in poor health and having limitations in performing activities of daily living (Medicare Payment Advisory Commission 2021a). Fully dual-eligible beneficiaries are more likely to be readmitted to the hospital, so providers that treat a disproportionate share of fully dual-eligible beneficiaries may have worse results for the measure of hospitalization within the stay (Bennett and Probst 2016). Dual-eligible beneficiaries also have higher mortality rates, so their providers can have worse results on the discharge to community

## Area-level measures of social risk

**S**ocial factors such as income, housing, social support, transportation, and nutrition can affect access to health services or desired health outcomes. However, such social risk information for individual beneficiaries is not routinely or systematically collected across the health care system or is not currently available to Medicare.

Research indicates that residents of impoverished neighborhoods or communities are at higher risk for mental illness, chronic disease, mortality, and lower life expectancy (Office of Disease Prevention and Health Promotion 2021). Because some health outcomes are tied to communities, there may be potential to use area-level measures to capture a broader range of a patient population's social risks. More development and testing of area-level measures of social risk is needed, and that work is outside of the scope of our mission. Here, we summarize some of the work by others on using area-level measures of social risk, describe some currently available measures, and call for CMS to explore the use of area-level measures of social risk.

As directed by the Congress in the Improving Medicare Post-Acute Care Transformation Act of 2014, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation and

the National Academies of Sciences, Engineering, and Medicare (NASEM) have done extensive research and deliberated on how to account for social risk factors in Medicare quality measurement and payment (Assistant Secretary for Planning and Evaluation 2020a, Assistant Secretary for Planning and Evaluation 2016, National Academies of Sciences 2016a, National Academies of Sciences 2016b). The NASEM committee concluded that a measure of neighborhood deprivation at the census tract level is likely to be a good proxy for a range of both individual and true area-level constructs relevant to performance indicators used in quality payment programs. The committee called for testing composite measures (i.e., a composite measure of neighborhood characteristics such as share of families below the poverty level and unemployment rate) and a simple single-indicator item (such as median household income). One test the committee suggested was to assess the performance of any given variable (single or composite) across multiple geographic areas and, in particular, rural areas.

Several composite area-level deprivation indicators have been developed by researchers and other government agencies. They generally use U.S. Census data from previous years, which are updated every few years. Three such indicators include:

*(continued next page)*

measure (Wadhera et al. 2020). Moreover, dual-eligible beneficiaries make up disproportionately high shares of Medicare spending, so providers treating a high share of dual-eligible beneficiaries can have higher MSPB rates (Keohane et al. 2018).

In its work on social risk factors and Medicare value-based payment programs, the Department of Health and Human Services' Assistant Secretary for Planning and Evaluation concluded that dual eligibility for Medicare and Medicaid remains a powerful predictor of

poor outcomes in Medicare's value-based purchasing (VBP) programs (Assistant Secretary for Planning and Evaluation 2020b). Beneficiary eligibility for Medicare and Medicaid status is readily available in administrative data. Using a provider's share of fully dual-eligible beneficiaries is consistent with prior Commission work on other VIPs (hospital, Medicare Advantage, and SNF).

Although there are many reasons to use dual eligibility as a proxy for beneficiary social risk, we recognize



## Area-level measures of social risk (cont.)

- **Area Deprivation Index (ADI):** The ADI ranks neighborhood socioeconomic disadvantages using American Community Survey (ACS) data collected by the U.S. Census (University of Wisconsin School of Medicine and Public Health 2021). Each neighborhood is defined as the census block group.<sup>5</sup> The ADI combines 17 social risk factors for each geographic unit, including two measures of poverty, unemployment rate, income level, income disparity, educational attainment, percent employed in white-collar occupations, median home value, crowding, housing units without complete plumbing, vehicle access, telephone access, owner-occupied housing, median gross rent, median monthly mortgage, and share of single-parent households.
- **Social Vulnerability Index (SVI):** The SVI measures the potential negative effects of external stresses on specific communities' health (Agency for Toxic Substances and Disease Registry 2021). The Centers for Disease Control and Prevention (CDC) reports this measure to help local officials identify communities that may need support before, during, or after disasters. The SVI evaluates each census tract on 15 social factors from the ACS data that fit into 4 categories (socioeconomic status, household

composition and disability, minority status and language, and housing type and transportation).

- **PLACES:** PLACES is intended to help local health departments and jurisdictions better understand the burden and geographic distribution of health-related outcomes in their areas and help them plan public health interventions (Centers for Disease Control and Prevention 2021). PLACES provides small-area estimates for counties, places, census tracts, and ZIP Code Tabulation Areas to help state and local health officials focus their efforts on improving health. PLACES includes 27 health measures, including unhealthy behaviors (5), health outcomes (13), and prevention practices (9). The 27 measures are based on CDC Behavioral Risk Factor Surveillance System and ACS data.

CMS should test various area-level measures for their potential to account for differences accurately across providers in the social risk of their patient populations. More research is needed to understand the accuracy of any area-level measure for Medicare beneficiaries compared with the gold standard of person-reported information. For example, the reliability of an area-level index should be assessed by comparing individual responses from a patient survey with their associated area-level index. ■

that it is an imperfect measure. One drawback is that Medicaid eligibility requirements and benefits vary across states. Also, dual eligibility may be too narrow because it reflects a beneficiary's income but does not directly reflect other social risks, like food insecurity and limited access to transportation.

One approach to capturing beneficiary social risk more comprehensively would be to use area-level measures of social risk (see text box). Policymakers and researchers should explore what existing or new area-level measures are used and whether the measures

accurately account for the social risk of individual patients.

In our illustrative PAC VIP model, we tested the use of the Area Deprivation Index (ADI), a geographically based measure of various social risk factors (such as the area's poverty rate, average educational attainment, and access to an automobile or telephone) of the communities where a provider's patients live. We chose this measure because it is publicly available at the nine-digit ZIP code level, which may, for any given beneficiary, be more accurate for a broader geographic

**TABLE  
14-1**

**For each setting, better performance on measures earns more points under illustrative PAC VIP model**

	All-condition hospitalization rate within stay (lower is better)	Medicare spending per beneficiary (<1 is better)	Successful discharge to the community rate (higher is better)	All-condition hospitalization rate within stay (lower is better)	Medicare spending per beneficiary (<1 is better)	Successful discharge to the community rate (higher is better)
<b>Points</b>	<b>SNF</b>			<b>IRF</b>		
0	23%	1.4	23%	11%	1.2	54%
2	19	1.2	31	10	1.1	59
4	16	1.1	38	8	1.0	63
6	13	1.0	44	7	1.0	66
8	11	0.8	52	6	0.9	70
10	8	0.7	62	4	0.9	74
<b>Points</b>	<b>LTCH</b>			<b>HHA</b>		
0	11%	1.2	16%	35%	1.2	35%
2	8	1.1	20	26	1.1	52
4	6	1.0	23	22	1.0	64
6	4	1.0	27	19	1.0	73
8	3	0.9	31	16	0.9	82
10	1	0.8	38	12	0.8	91

Note: PAC (post-acute care), VIP (value incentive program), SNF (skilled nursing facility), IRF (inpatient rehabilitation facility), LTCH (long-term care hospital), HHA (home health agency). Each of the three risk-adjusted measures in the PAC VIP model is continuously scored from 0 to 10 points; only a subset of points is displayed here. The performance-to-points scale is set using the range of risk-adjusted performances for all providers in the setting. Because performance on Medicare spending per beneficiary is rounded to the tenth place, some performance values appear to be associated with different points. The SNF performance scores are based on 12,937 providers. The IRF scores are based on 991 providers. The LTCH scores are based on 387 providers. The HHA scores are based on 8,618 providers.

Source: MedPAC analysis of claims data, 2015–2017.

area, such as a county or five-digit ZIP code. Although we present a summary of that modeling (see text box on ADI results, pp. 542–543), we focus discussion of our modeling results on dual eligibility as the measure of social risk because it is the most-tested proxy for patient social risk that we have at this time.

In our illustrative PAC VIP model, we examined the *empirical association* (correlation) between performance and the share of fully dual-eligible beneficiaries treated to determine whether peer grouping is needed to account for differences in providers’ patient populations for each setting. When performance was systematically worse for providers treating high shares of patients at high social risk (i.e.,

there was an inverse relationship between the two), peer grouping was used to determine the payment adjustments for that setting. If peer grouping had not been used and a provider’s performance was worse due to the high social risk of its patient population, there would be no mechanism to reward the performance it achieved. Conversely, if performance was better for providers treating more patients at high social risk, we concluded that peer grouping was not needed for that setting.

**How many peer groups should be used to differentiate providers?**

In setting the number of peer groups, policymakers would need to balance two goals. A sufficient number

of groups is needed to differentiate the social risk of patient populations so that the providers within each group are relatively homogeneous in their shares of social risk. At the same time, the number of groups should be small enough that distinctions between providers are meaningful. CMS should evaluate the number of groups that best differentiates providers with similar shares of patients with social risk. For each setting, the number of peer groups will be a function of the number of providers in a setting and the range of social risk across providers. In our model, the average share of fully dual-eligible beneficiaries differed by setting. The lowest share was in IRFs (18 percent) and the highest share was in SNFs (43 percent). The share in HHAs was 33 percent, and the share in LTCHs was 38 percent.

One approach would be to group providers using natural breaks in the distribution of the social risks, if there are any. This approach could result in an unequal number of providers in each peer group, but it could more accurately reflect “like” providers. When the distribution does not have natural breaks, CMS could create groups with equal numbers of providers in each.

In our model, we did not find any natural breaks in the distribution of social risk that would indicate peer groups. Therefore, we used equal-size peer groups and based the number of groups on the number of providers in a setting.

### **Method to distribute the entire provider-funded pool of dollars**

Medicare quality programs should not attempt to achieve Medicare savings but rather should fully distribute the provider-financed pool of incentive payments as rewards and penalties. A PAC VIP would distribute the entire provider-funded pool of dollars within each peer group based on providers’ quality performance during the performance period. A key decision for policymakers is how large potential rewards and penalties need to be to motivate providers to improve performance and avoid poor performance.

### **How large should the rewards and penalties be?**

Policymakers could consider a program that begins with a 2 percent withhold and scales up to a larger withhold amount (e.g., 5 percent) over two or three years. A graduated approach is used in Medicare’s

home health VBP demonstration (run by the Center for Medicare & Medicaid Innovation), which started with a 3 percent withhold but increased to 8 percent in 2022. In 2023, CMS is implementing the home health VBP program nationally using a 5 percent withhold. Alternatively, policymakers could opt to begin immediately with the higher withhold amount (e.g., 5 percent).

We modeled an illustrative PAC VIP that uses 5 percent of provider payments to fund the pool of dollars. Our model includes seven steps to convert performance points to payment adjustments and, within each peer group, entirely distribute the pool of dollars as rewards (see text box describing the process to convert points to a quality-based payment adjustment, pp. 536–537).

---

## **Results of our illustrative model of a PAC VIP design**

---

Although a conceptual relationship exists between a provider’s share of fully dual-eligible beneficiaries and its outcomes, we did not find consistent empirical relationships across the four PAC settings. Using a provider’s share of dual-eligible beneficiaries treated as the measure of social risk, we found that adjusting for social risk was needed for SNFs and IRFs, and peer groups would counter the disadvantages they face in earning performance points. In contrast, higher social risk was associated with better performances for HHAs and LTCHs. More work is needed to confirm this finding and to disentangle the various factors that shape provider performance. For example, a broader measure of social risk that captures its multiple dimensions could more uniformly tie social risk to performance.

In our illustrative PAC VIP model, providers gain more points for better performance on the three performance measures. For each measure, points are assigned on a performance-to-points scale from 0 to 10 based on the continuous and setting-specific national distributions of providers’ scores.<sup>6</sup> Providers earn more points for lower within-stay hospitalization rates, lower MSPB, and higher rates of successful discharge to the community. Table 14-1 illustrates how the three measures are converted into PAC VIP points by PAC setting (only a subset of points is shown).

**TABLE  
14-2**

**Providers' share of fully dual-eligible beneficiaries was not consistently related to performance under illustrative PAC VIP model**

	SNF	IRF	LTCH	HHA
Correlation between share of fully dual-eligible beneficiaries treated and total performance points	-0.60	-0.18	0.11	0.14

Note: PAC (post-acute care), VIP (value incentive program), SNF (skilled nursing facility), IRF (inpatient rehabilitation facility), LTCH (long-term care hospital), HHA (home health agency). Total performance points are the sum of the average points a provider in a setting earns on three measures: hospitalizations within the post-acute care stay, Medicare spending per beneficiary, and successful discharge to the community. This analysis included 12,937 SNFs, 991 IRFs, 387 LTCHs, and 8,618 HHAs.

Source: MedPAC analysis of claims data, 2015–2017.

The best-performing SNFs, with a hospitalization rate of about 8 percent, would earn 10 points for that measure, while the worst-performing SNFs (hospitalization rate of about 23 percent) would not earn points for that measure. SNF and HHA settings had the most variable performances for all three measures, so the 10-point scale spans larger differences in performance compared with the range in points for IRFs and LTCHs. For every PAC provider, after the points for each quality measure are determined, the total PAC VIP points are calculated by averaging the points for each measure (0 to 10 points). This calculation effectively weights each measure equally, although policymakers could weight them differently.

**Across PAC settings, the share of fully dual-eligible beneficiaries was not uniformly related to performance**

Using a provider's share of fully dual-eligible beneficiaries treated as the measure of social risk, we found that SNFs and IRFs with high shares of fully dual-eligible beneficiaries had worse performance than those with low shares. The association between performance and this measure of social risk was strong for SNFs and relatively weak for IRFs, as indicated by the size of the negative correlations (Table 14-2).<sup>7</sup> Especially for SNFs, a peer-grouping approach would counter some of the disadvantages that providers with high shares of fully dual-eligible beneficiaries face in achieving good performance. As the average share of fully dual-eligible beneficiaries increased, providers

would have the potential to earn larger rewards for better performance.

In contrast, for HHAs and LTCHs, as a provider's share of fully dual-eligible beneficiaries increased, average performance improved, though the relationships were relatively weak. We confirmed these results with regressions predicting performance based on the share of fully dual-eligible beneficiaries treated and, in separate models, other provider characteristics. Based on these correlations, we modeled VIP payments for LTCHs and HHAs without peer groups. However, more work should be done before policymakers conclude that peer grouping would not be needed for these providers.

The model's results with peer groups for SNFs and IRFs and no peer groups for HHAs and LTCHs are instructive. For SNFs and IRFs, we scaled the number of peer groups by setting so that each group was sufficiently large to calculate the effects of the peer grouping. We used 20 groups for SNFs (about 650 SNFs in each group) and 5 groups for IRFs (about 200 IRFs in each group). We did not use peer groups for HHAs and LTCHs.

**Skilled nursing facilities**

Across the 20 peer groups, the average share of fully dual-eligible beneficiaries ranged from 3 percent for providers in Peer Group 1 (the providers treating beneficiaries with the lowest social risk) to 91 percent

**TABLE  
14-3**

**In the illustrative PAC VIP model, rewards for the best-performing SNFs were larger for those with higher shares of fully dual-eligible beneficiaries**

Peer group	Average share of fully dual-eligible beneficiaries	Average PAC VIP points earned	Range of performance points (25th–75th percentiles)	Multiplier (converts points to payment)	Net payment adjustment (after 5% withhold)
1 (lowest share)	3%	7.1	6.2 to 8.2	0.70%	–4.9% to 2.0%
2	10	7.1	6.1 to 8.2	0.71	–3.3 to 2.1
3	15	6.8	5.8 to 8.2	0.74	–4.5 to 2.3
4	19	6.6	5.5 to 7.8	0.78	–3.9 to 2.7
5	23	6.3	5.1 to 7.6	0.82	–4.3 to 3.0
6	27	6.1	5.0 to 7.3	0.85	–4.3 to 3.4
7	30	5.9	4.7 to 7.1	0.86	–4.4 to 3.3
8	34	5.7	4.5 to 7.1	0.89	–4.9 to 3.7
9	37	5.5	4.2 to 6.9	0.90	–4.8 to 4.0
10	40	5.2	3.9 to 6.5	0.98	–4.7 to 4.5
11	44	5.1	3.8 to 6.4	1.00	–4.9 to 4.9
12	47	4.9	3.6 to 6.1	1.06	–4.5 to 5.6
13	51	4.5	3.1 to 5.9	1.13	–5.0 to 5.5
14	54	4.3	2.9 to 5.7	1.21	–4.7 to 6.3
15	58	4.0	2.4 to 5.4	1.28	–5.0 to 7.4
16	62	3.9	2.6 to 5.2	1.33	–4.9 to 8.0
17	67	3.7	2.1 to 5.1	1.42	–4.9 to 7.5
18	73	3.3	1.7 to 4.7	1.61	–4.9 to 10.2
19	80	2.9	1.4 to 4.1	1.81	–4.9 to 12.0
20 (highest share)	91	2.6	1.3 to 3.7	2.12	–5.0 to 15.0

Note: PAC (post-acute care), VIP (value incentive program), SNF (skilled nursing facility). Peer groups are based on the share of fully dual-eligible beneficiaries. There are about 650 SNFs in each of the 20 peer groups. Peer groups are assigned based on the share of the SNF’s Medicare patients who were fully eligible for Medicare and Medicaid benefits for at least one month of the year. The incentive pool of dollars for each peer group includes 5 percent of Medicare payments for each SNF in the peer group. The multiplier is the percentage adjustment to payments per performance point. Negative payment adjustments are penalties; positive adjustments are rewards. The analysis included 12,937 SNFs.

Source: MedPAC analysis of claims data, 2015–2017.

for providers in Peer Group 20 (the providers treating beneficiaries with the highest social risk) (Table 14-3).<sup>8</sup> Average performance points earned by providers steadily declined across the peer groups, reflecting the difficulty providers face in achieving good performance for patients at high social risk. To avoid penalizing providers with high shares of beneficiaries at high social risk, the multiplier that converts a SNF’s performance points to a payment adjustment increases with each peer group

(see text box on the methodology to convert points to a payment adjustment using SNFs as an example, pp. 536–537). As a result, SNFs in the group with highest social risk (Peer Group 20) had the potential to earn larger rewards for higher quality compared with SNFs in the peer group with lowest social risk (Peer Group 1).

About an equal number of SNFs earned a reward or were assessed a penalty, and the average net adjustment to payments (net of the 5 percent withhold)



**TABLE  
14-4**

**In the illustrative PAC VIP model, payment adjustments varied by SNF characteristics**

SNF characteristics	Number of providers	Average net payment adjustment (after 5% withhold)
All SNFs	12,937	0.13%
Ownership		
Nonprofit	2,741	0.37
For profit	9,359	0.07
Government	828	0.13
Location		
Urban	9,714	0.18
Rural	3,217	0.01
Facility type		
Hospital based	503	1.94
Freestanding	12,425	0.07

Note: PAC (post-acute care), VIP (value incentive program), SNF (skilled nursing facility). The table shows unweighted average net payment adjustments. Although rewards are financed entirely by the pool of withheld payments, net payment adjustments do not necessarily average to 0 percent because larger providers, which contribute more dollars to the pool, have their net payment adjustments weighted the same as smaller providers, which contribute fewer dollars to the pool in the average. The subgroups of providers do not always sum to 12,937 due to missing data on provider characteristics.

Source: MedPAC analysis of claims data, 2015–2017.

was 0.1 percent.<sup>9</sup> Comparing SNFs across all the peer groups, the largest reward was a 15 percent increase to payments and the largest penalty was a 5 percent reduction. While the differences were small across provider groups, nonprofit, urban, and hospital-based SNFs had higher average net payment adjustments compared with for-profit, rural, and freestanding SNFs (Table 14-4). Regression analysis that included share of dual-eligible beneficiaries, ownership, provider type, and size to explain differences in performance confirmed these results (facility size was not a significant factor).

We examined the performance of hospital-based SNFs because they had notably higher average payment adjustments than freestanding SNFs. This result reflects better performance on all three measures. Compared with freestanding facilities, hospital-based providers on average had hospitalization rates during

the stay that were 45 percent lower, MSPB that was 42 percent lower, and successful discharge to community rates that were 27 percent higher. Hospital-based SNFs typically have lower readmission rates (which affects the results for the measure of hospitalization during the stay and MSPB), which may be due to their higher staffing levels and physician presence as well as more timely lab results for patients. We also examined the relationships between staffing levels (total nurse hours per resident per day) and VIP points; for each quality measure, we found statistically significant relationships. Higher staffing levels were positively related to VIP points, higher rates of successful discharge home, and lower hospitalization rates. These findings are consistent with those of a study of nursing home quality measures, which found that better performance was associated with higher staffing levels and lower shares of Medicaid patients (Saliba et al. 2018). This study also found that hospital-based

**TABLE  
14-5**

**In the illustrative PAC VIP model, IRFs with higher shares of fully dual-eligible beneficiaries would have the potential to earn higher rewards**

Peer group	Average share of fully dual-eligible beneficiaries	Average PAC VIP points earned	Range of performance points (25th–75th percentiles)	Multiplier (converts points to payment)	Net payment adjustment (after 5% withhold)
1 (lowest share)	7%	5.5	4.1 to 6.9	0.98%	–4.7% to 4.3%
2	12	5.0	3.8 to 6.1	1.10	–4.3 to 5.2
3	16	5.2	3.8 to 6.6	1.02	–4.7 to 4.7
4	21	5.0	3.2 to 6.5	1.09	–4.4 to 5.6
5 (highest share)	37	4.5	2.8 to 6.0	1.16	–5.0 to 6.1

Note: PAC (post-acute care), VIP (value incentive program), IRF (inpatient rehabilitation facility). There are about 198 IRFs in each of the 5 peer groups. Peer groups are assigned based on the share of the hospital’s Medicare patients who are fully eligible for Medicare and Medicaid benefits for at least one month of the year. The incentive pool of dollars for each peer group includes 5 percent of Medicare payments for each IRF in the peer group. The multiplier is the percentage adjustment to payments per performance point. Negative payment adjustments are penalties; positive adjustments are rewards. The analysis included 991 IRFs.

Source: MedPAC analysis of 2015–2017 fee-for-service claims and payment data.

providers had lower readmission rates and higher rates of discharge to community and that higher Medicaid shares worsened performance on both measures.

**Inpatient rehabilitation facilities**

The peer-grouping approach may help counteract the disadvantages that IRFs that treat high shares of fully dual-eligible beneficiaries face in achieving high performance, but the effect of peer grouping in our model was small (Table 14-5). Compared with IRFs in the lowest peer group (Peer Group 1), providers with the highest share of fully dual-eligible beneficiaries (Peer Group 5) earned fewer performance points (4.5 points compared with 5.5 points), but under peer grouping, they have the potential to earn larger rewards for higher quality (a point-to-payment adjustment multiplier of 1.16 percent compared with a multiplier of 0.98 percent for IRFs in Peer Group 1). The small effects of peer grouping in IRFs compared with SNFs may be partly due to the narrow range of shares of fully dual-eligible beneficiaries that may adversely affect performance. The average shares ranged from 7 percent for IRFs in Peer Group 1 (the lowest quintile) to 37 percent for IRFs in Peer Group 5 (the top quintile). In contrast, the average shares in SNFs ranged from 12 percent for SNFs in the lowest quintile to 78 percent for SNFs in the top quintile.

It is also possible that Medicare’s criteria for payment may restrict the complexity of the beneficiaries treated in IRFs.<sup>10</sup> The criteria reduce the differences in the patients admitted and, in turn, may narrow differences in performances across providers (even after risk adjusting results for clinical factors). There was a single average performance-point difference between the lowest and highest IRF peer groups. In contrast, the range in average performance points between the lowest and highest SNF peer groups was much wider (a spread of 4.5 points).

The results for the middle three peer groups indicate that although the providers’ shares of fully dual-eligible beneficiaries steadily increased, their average performance was similar. We contemplated collapsing these middle groups into one group but found that the multiplier would still be around 1 percent, which is not surprising given the weak relationship between duals’ share and performance. We opted to retain equal-size peer groups, consistent with the other VIP designs the Commission has proposed.

Across all the peer groups, a slightly larger share of IRFs earned rewards rather than penalties (56 percent compared with 44 percent). The average adjustment to

## Using peer groups to convert points in the post-acute care value incentive program to rewards and penalties

The Commission's illustrative model of the post-acute care (PAC) value incentive program (VIP) distributes quality-based payments to providers. A provider's performances on three quality measures are compared with a setting-specific performance scale. For skilled nursing facilities (SNFs) and inpatient rehabilitation facilities (IRFs), we found that providers treating a large share of fully dual-eligible beneficiaries had worse performance (dual-eligible beneficiaries receive both Medicare and Medicaid benefits, and this is used as a proxy for low income). Thus, for SNFs and IRFs, providers are assigned to peer groups based on their share of fully dual-eligible beneficiaries. Each peer group has about the same number of providers and a pool of dollars based on a 5 percent payment withhold from each of the respective group's providers. For long-term care hospitals (LTCHs) and home health agencies (HHAs), we did not find that providers treating a large share of fully dual-eligible beneficiaries had worse performance. For LTCHs and HHAs, there is a pool of dollars based on a 5 percent payment withhold from these settings' providers.

We follow seven steps to convert each provider's quality measure performance to a payment adjustment for calculating rewards and penalties.

**Step 1:** Calculate each provider's performance on each of the three risk-adjusted quality measures using beneficiary-level administrative data.

**Step 2:** Convert each provider's performance on each of the three quality measures to points based on a continuous performance-to-points scale (PAC setting-specific scales). With a continuous scale, any difference in performance is translated to a difference in payment.

**Step 3:** Average each provider's points on the three measures to determine the provider's PAC VIP total points.

**Step 4:** For each SNF and IRF, calculate the share of Medicare admissions who are fully eligible for Medicaid. Assign each SNF into 20 equal-size peer groups based on their share of fully dual-eligible patients. Assign each IRF into five equal-size peer groups based on their share of fully dual-eligible patients. LTCHs and HHAs have one peer group.

**Step 5:** For each group, create a pool of dollars of expected PAC VIP payments based on 5 percent of the peer-group providers' total Medicare payments.

**Step 6:** For each peer group, calculate the multiplier (the percentage adjustment to payment per PAC VIP point) that converts PAC VIP total points to dollars and results in spending the group's pool of dollars defined in Step 5.

$$\text{multiplier} = \text{PAC VIP pool for peer group} / (\text{sum (each facility's total Medicare payments} \times \text{its total PAC VIP points)})$$

**Step 7:** Compute each provider's adjustment for the coming year based on past performance and its peer group's multiplier.

$$\text{provider's SNF VIP-based adjustment} = \text{multiplier} \times \text{provider's PAC VIP total points}$$

These steps illustrate the conversion of PAC VIP points to payment adjustments using peer grouping. Table 14-6 considers the example of two SNFs, SNF A and SNF B. For each of the SNFs, we calculate performance results based on administrative data for each of the three quality measures (Step 1). Using the continuous performance-to-points scales, we convert quality performance to points (Step 2). We average each provider's performance on the three measures to determine PAC VIP total points (Step 3). SNF A has higher total VIP performance (10.0) than SNF B (7.5).

Though SNF A is smaller than SNF B, with 2,400 Medicare days per year compared with 4,400 for SNF B, they have similar shares of admissions who

*(continued next page)*



## Using peer groups to convert points in the post-acute care value incentive program to rewards and penalties (cont.)

**TABLE  
14-6**

### An example of converting points to payment adjustments for two SNFs within a peer group

Step	Peer Group 1 (Step 4)	
	SNF A	SNF B
Medicare days [facility beds x 365 days x occupancy rate x Medicare share of days]	2,400	4,400
PAC VIP total points (Steps 1-3)	10.0	7.5
Total base facility Medicare payments	\$1,000,000	\$2,000,000
5 percent of facility Medicare payments (withhold)	\$50,000	\$100,000
Pool of dollars for peer group (Step 4-5)	\$150,000	
Percentage adjustment to payment per PAC VIP point (the multiplier) for peer group (Step 6) [group's pool / sum of (provider's payments x points)]	0.60 percent adjustment per point	
PAC VIP payment adjustments (Step 7) [points x multiplier]	6.00%	4.50%
PAC VIP payments [PAC VIP payment adjustment x total payments]	\$60,000	\$90,000
Net payments after 5 percent provider contribution to the pool	+ \$10,000	- \$10,000

Note: SNF (skilled nursing facility), PAC (post-acute care), VIP (value incentive program). This example assumes a peer group of two SNFs with a similar share of fully dual-eligible beneficiaries (Step 4).

are fully dual eligible for Medicare and Medicaid, which places them in the same peer group (Step 4). We next determine 5 percent of each facility's total Medicare payments (Step 5). Since SNF A has fewer Medicare days, its contribution to the pool of dollars is less (\$50,000) than SNF B's contribution (\$100,000). The total SNF VIP pool of dollars to be redistributed for this peer group is equivalent to 5 percent of combined payments to the two SNFs (\$150,000). The multiplier for the peer group is then calculated (Step 6), which sets the payment adjustment per point for the peer group. For Peer

Group 1, the multiplier is 0.6 percent; thus, each PAC VIP point earned results in a payment adjustment of 0.6 percent. The peer group multiplier is then applied to each PAC's VIP point total (Step 7). SNF A earns a payment adjustment of 6.0 percent, which is equal to \$60,000 (or a net reward of \$10,000 more than its contribution to the pool). SNF B earns a payment adjustment of 4.5 percent, which is equal to \$90,000 (or a net penalty of \$10,000 less than its contribution to the pool). The entire pool of \$150,000 is distributed among the providers in the peer group. ■

**TABLE  
14-7**

**In the illustrative PAC VIP model, payment adjustments varied by IRF characteristics**

IRF characteristics	Number of providers	Average net payment adjustment (after 5% withhold)
All IRFs	991	0.34%
Ownership		
Nonprofit	554	0.64
For profit	330	0.00
Government	107	-0.19
Location		
Urban	855	0.29
Rural	136	0.63
Facility type		
Hospital based	791	0.52
Freestanding	200	-0.40

Note: PAC (post-acute care), VIP (value incentive program), IRF (inpatient rehabilitation facility). The table shows unweighted average net payment adjustments. Although rewards are financed entirely by the pool of withheld payments, average net payment adjustments do not necessarily average to 0 percent because larger providers, which contribute more dollars to the pool, have their net payment adjustments weighted the same as smaller providers, which contribute fewer dollars to the pool in the average.

Source: MedPAC analysis of claims data, 2015–2017.

payments was a small reward (0.3 percent); the largest reward was a 6 percent adjustment to payments (net of the 5 percent withhold), and the largest penalty was 5 percent reduction to payments. While differences across provider groups were small, hospital-based IRFs, nonprofit IRFs, and rural IRFs had higher average net payment adjustments compared with freestanding, for-profit, and urban IRFs (Table 14-7). Regression analysis that predicted performance points based on the share of fully dual-eligible patients, provider type, ownership, and size confirmed these results (size was not a significant factor). It is possible that certain facilities have specialized units or have better information about potential admissions that could help explain these findings.

### HHAs and LTCHs

Because our empirical analysis found that HHAs and LTCHs with higher shares of fully dual-eligible beneficiaries generally had better performance than

providers with lower shares, we modeled a PAC VIP without peer groups. In all other ways, the approach was the same: We compared a provider’s performance with a setting-specific continuous performance-to-points scale and fully distributed a provider-funded pool of dollars.

About equal shares of providers gained and lost under a PAC VIP, and the average adjustment was about zero (Table 14-8). Adjustments to payments ranged from -5 percent to 5 percent (after the 5 percent withhold).

There were differences in performance points across provider characteristics. Nonprofit providers received larger rewards on average than other providers, and hospital-based HHAs received much larger rewards on average than freestanding providers (Table 14-9, p. 540). Regression analysis found that the share of fully dual-eligible beneficiaries was not a significant factor in explaining LTCH performance but was for HHAs: Agencies with higher shares of fully dual-eligible

**TABLE  
14-8**

**Impact of illustrative PAC VIP model without peer groups was similar for HHAs and LTCHs**

	HHAs	LTCHs
Share of providers whose payments would increase	50%	46%
Share of providers whose payments would decrease	50	54
Median net percent change in payments	0.0	0.2
Largest reward (net percent change in payment)	5	4
Largest penalty (net percent change in payment)	-5	-5

Note: PAC (post-acute care), VIP (value incentive program), HHA (home health), LTCH (long-term care hospital). The illustrative VIP used a 5 percent withhold and fully redistributed the pool as incentive payments. Performance was gauged with three outcome measures: hospitalizations within the stay, successful discharge to the community, and Medicare spending per beneficiary. Change in payments is net of the 5 percent withhold. This analysis included 8,618 HHAs and 387 LTCHs.

Source: MedPAC analysis of claims data, 2015–2017.

patients had better performance. HHAs with high shares of beneficiaries admitted from the community had worse performance compared with other HHAs.

**The results for HHAs and LTCHs highlight the complexities of measuring social risk and performance**

Though a conceptual relationship exists between treating more dual-eligible beneficiaries and providers’ poorer performance on the measures, our empirical analysis found that HHAs’ and LTCHs’ dual-eligible patient shares were positively associated with performance. That is, as the share of fully dual-eligible beneficiaries increased, provider performance improved, though the increases were small. Because the empirical finding conflicts both with the conceptual relationship posited for these settings and with the empirical findings in other PAC settings, more work is needed on the definition of social risk and the measurement of performance. Disentangling these relationships is beyond the scope of this report, but we outline some factors below that may complicate the relationship between the share of fully dual-eligible beneficiaries and provider performance.

**Definitions of fully dual eligible vary across states**

As noted above, in its work on social risk factors and Medicare value-based payment programs, the

Department of Health and Human Services’ Assistant Secretary for Planning and Evaluation concluded that dual eligibility for Medicare and Medicaid was a powerful predictor of poor outcomes in Medicare’s VBP programs (Assistant Secretary for Planning and Evaluation 2020b). However, each state has different eligibility requirements for Medicaid. All states are required to provide Medicaid to people who receive Supplemental Security Income (SSI) benefits, but they also have multiple options for providing Medicaid to other populations. For example, 21 states and the District of Columbia cover seniors and disabled individuals who have incomes that are higher than the SSI limit but below the federal poverty level, 31 states and the District of Columbia cover seniors and disabled individuals who have high medical expenses, and 42 states and the District of Columbia cover seniors and disabled individuals who need long-term services and supports and have income below 300 percent of the SSI limit (Musumeci et al. 2019). Thus, some older adults with low incomes and/or functional impairments may qualify for Medicaid in some states but not in others. These eligibility differences contribute to differences in the shares of beneficiaries who are fully dual eligible across states. For example, in 2020, 9 percent of beneficiaries in Maryland were fully dual eligible, compared with 27 percent of beneficiaries in California.

**TABLE  
14-9**

**In the illustrative PAC VIP model, payment adjustments varied by HHA and LTCH characteristics**

Characteristic	HHA		LTCH	
	Number of providers	Average net payment adjustment (after 5% withhold)	Number of providers	Average net payment adjustment (after 5% withhold)
All providers	8,618	0.05%	387	-0.18%
Ownership				
Nonprofit	1,092	1.37	65	0.12
For profit	6,947	-0.24	307	-0.20
Government	398	1.04	N/R	N/R
Location				
Urban	7,085	0.05	N/R	N/R
Rural	1,527	0.05	N/R	N/R
Facility type				
Hospital based	811	1.31	N/A	N/A
Freestanding	7,626	-0.11	N/A	N/A

Note: PAC (post-acute care), VIP (value incentive program), HHA (home health agency), LTCH (long-term care hospital), N/R (not reported), N/A (not applicable). The table shows unweighted average net payment adjustments. Although rewards are financed entirely by the pool of withheld payments, average net payment adjustments do not necessarily average to 0 percent because larger providers, which contribute more dollars to the pool, have their net payment adjustments weighted the same as smaller providers, which contribute fewer dollars to the pool in the average. All LTCHs are considered freestanding. The counts by provider group may not equal the total number due to missing information about the provider. We did not report government-owned, urban, and rural LTCHs because the number of providers in these categories was too small for meaningful comparisons.

Source: MedPAC analysis of claims data, 2015–2017.

**Extent of home- and community-based services varies across states** States also differ in the shares of Medicaid spending on long-term services and supports (LTSS) that are devoted to home- and community-based services (HCBS).<sup>11</sup> HCBS can help beneficiaries remain in their homes—especially relevant for beneficiaries receiving home health care. Some states devote more than 75 percent of their LTSS spending to HCBS, while others spend less than 40 percent (Murray et al. 2021). We found that HHAs in states with higher proportions of their Medicaid LTSS spending devoted to HCBS had, on average, better performance on each of the PAC VIP measures compared with HHAs in other states.

**Small number of providers** Nationwide, there are fewer than 400 LTCHs—far fewer than the 15,000 SNFs and 11,000 HHAs. These figures suggest that there are not

enough LTCHs to meaningfully identify peer groups that could account for differences in the social risk of LTCH patient populations.

**Risk adjustment may not fully capture differences in patient complexity** Accurate risk adjustment is always challenging, but developing an accurate model (for each measure) across four settings is especially so. With different coverage and cost-sharing rules and complicated PAC placement decisions, it is likely that quite a bit of “sorting” occurs before patients are admitted to a PAC provider (assuming most markets have at least two types of PAC providers). Furthermore, providers may selectively admit patients they can effectively treat or expect will be profitable. Risk adjustment may not adequately address the patient selection that occurs across the four settings.

Anticipating a unified PAC PPS, we used a single risk-adjustment model for each performance measure—with a common set of risk adjusters and uniform weights—for all PAC providers. While this approach sets the stage for a unified PAC VIP, it may be less accurate for providers in any given setting compared with using risk-adjustment models that are specific to each setting—that is, the factors and their relative importance would vary by setting. In a single risk-adjustment model, all PAC stays are used to establish the risk adjusters and their relative weights. As a result, the single models (one for each outcome measure) are dominated by HHA and SNF stays, which account for the vast majority of PAC use. As a result, because LTCHs account for only 1 percent of PAC stays, the characteristics of their patients will have little influence on the model. Furthermore, Medicare’s qualifying criteria for LTCHs result in a higher average risk score compared with other settings, which may not be fully captured by the single risk-adjustment model.<sup>12</sup> Therefore, LTCH performance may not be fully adjusted for the complexity of the patients they treat. In another example, some IRFs treat highly complex patients (such as those with traumatic brain injury) who may not be fully accounted for in a single risk-adjustment model.

At the same time, risk adjustment may not capture the characteristics of community-admitted patients, who comprise two-thirds of HHA stays. Although community-admitted beneficiaries are more likely to be dual eligible, they are not, on average, clinically complex. They have fewer chronic conditions but higher rates of Alzheimer’s and dementia and are frailer. Our risk-adjustment model does not include measures of frailty and so may not fully capture the differences between community-admitted and other PAC users.

In finalizing the measures of performance, the Secretary will need to ensure, to the extent possible, that the risk adjustment adequately reflects the differences in the clinical risks of the patients each provider treats. Otherwise, a provider may avoid admitting patients whose risks will put the provider at a disadvantage in performance measurement under the PAC VIP.

**Community risk factors likely play a bigger role in outcomes for HHAs** When beneficiaries are treated with home health care, the social risk factors of the communities they live in are especially relevant. For

example, whether a beneficiary has access to a car or telephone will affect the ease of making follow-up medical appointments or picking up prescribed medications, which could negatively impact an HHA’s hospitalization rate. Thus, communities’ social risk factors can be particularly important in understanding differences in HHA performance, yet these factors are not captured by the dual-eligibility measure (see text box with results of the illustrative PAC VIP model using ADI as a measure of community risk, pp. 542–543).

---

## Other ways to encourage providers to improve performance

---

A PAC quality payment program is not the only tool CMS has to encourage providers to improve their performance. These various approaches are not mutually exclusive; indeed, a combination of approaches might yield the largest and longest-lasting improvements in quality.

Publicly reporting Medicare quality information has two main objectives. The first is to increase the accountability of health care providers by offering patients, payers, and purchasers a more informed basis on which to hold providers accountable (e.g., directly through purchasing and treatment decisions). The second objective is to maintain standards and stimulate improvements in the quality of care through economic competition (reputation and increased market share) and by appeals to health care professionals’ desire to do a good job (Marshall et al. 2003). Researchers have identified and tested best practices on how to display comparative information to best meet the objectives of public reporting. Many such practices are incorporated in the Care Compare website—for example, using only a small number of data points (or the single data point of an overall star rating), with more detailed information available in a second or even third layer for those who want it (Agency for Healthcare Research and Quality 2020b, *Aligning Forces for Quality* 2009). The Commission also believes that public reporting should enable comparisons of individual providers with state and national averages to give consumers meaningful reference points.

To stimulate quality improvement, CMS could also incorporate performance standards in conditions



## Results of using the Area Deprivation Index in the illustrative post-acute care value incentive program model

Some health outcomes are tied to the characteristics of the communities where patients live, so conceptually, area-level measures may better capture the social risk of a provider's patient population (see text box on area-level measures of social risk, pp. 528–529). We explored the empirical relationship between performance and one area-level measure of social risk, the Area Deprivation Index (ADI). Using a published list of ADIs for each nine-digit ZIP code, we matched each beneficiary's home nine-digit ZIP code to its corresponding ADI. For each post-acute care (PAC) provider, we averaged the ADIs of the beneficiaries they treated over a year for a provider-level index. A provider's index captures the social risk of the communities where its patients live, not the community where the provider is located. We awarded performance points to each provider on three performance measures: hospitalization during the stay, successful discharge to the community, and Medicare spending per beneficiary.

We found that the empirical relationship between a provider's average ADI and its performance was not consistent across the four PAC settings. Skilled

nursing facilities (SNFs) and home health agencies (HHAs) with high average ADIs (more social risk) generally had worse performance compared with providers with low average ADIs. The correlations between average ADI and performance points were  $-0.24$  for SNFs and  $-0.26$  for HHAs. Peer groups would counter the disadvantages that SNFs and HHAs face in achieving good performance. In contrast, inpatient rehabilitation facilities (IRFs) and long-term care hospitals (LTCHs) with high average ADIs generally had better performance compared with providers with low average ADIs (correlations were  $0.05$  for IRFs and  $0.21$  for LTCHs). Because ADIs were positively correlated with the performances of IRFs and LTCHs, we modeled these settings without peer groups.

We assigned HHAs and SNFs to peer groups based on the provider's average ADI. As we did with the analyses of the shares of fully dual-eligible beneficiaries, we scaled the number of peer groups (20 groups) for SNFs and HHAs so that each group was sufficiently large to calculate the effects of the peer grouping.

*(continued next page)*

of participation (referred to as “requirements of participation” for SNFs). Providers that repeatedly deliver the poorest quality of care could be removed from the program. In addition, Medicare could target time-limited technical assistance resources to low-performing providers so they could develop the skills and infrastructure needed for successful quality improvement.

Some observers have noted that interventions such as upfront payments would enable providers to invest in activities that could improve their performance on quality measures (Arling et al. 2013, Assistant Secretary for Planning and Evaluation 2020b, Jaffery and Safran 2021). Upfront payments could take two forms: lump-

sum payments or enhanced payment rates. Lump-sum payments would be akin to technical assistance, with the added condition that qualifying providers would need to treat populations at higher social risk. Enhanced payment rates could be made to providers treating a population at higher social risk, similar to the disproportionate share payments made to acute care hospitals and to IRFs. Enhanced rates would not offer providers a block of dollars to use for quality improvement.

There is little evidence about which interventions improve care for high social risk patient populations. Unless the upfront payments were tied to specific

## Results of using the Area Deprivation Index in the illustrative post-acute care value incentive program model (cont.)

In SNFs, a multiplier of 1.4 for SNFs in the highest peer group (the most social risk) would enable the best performers to earn larger rewards compared with SNFs in the peer group with the lowest risk (multiplier of 0.9). Likewise, a multiplier of 1.28 for HHAs in the highest social risk peer group would enable the best performers to earn larger rewards for their outcomes compared with HHAs in the peer group with the lowest social risk (multiplier of 0.75).

Because the results using the ADI as the measure of social risk differed from the results using shares of fully dual-eligible beneficiaries, we examined the relationship between the two measures (share of fully dual-eligible beneficiaries and average ADI). We found that they were inversely related to each other for three of the four PAC settings (IRFs, LTCHs, and HHAs), though the associations were weak (correlations of  $-0.09$ ,  $-0.18$ , and  $-0.12$ , respectively). As their shares of fully dual-eligible

beneficiaries increased, their ADI scores dropped. These correlations indicate that the social risk proxies capture different aspects of the social risks associated with a provider's patient population. In SNFs, the two scores were positively correlated (correlation of 0.20).

These results raise questions about whether the ADI is the best area-based measure to account for social risk. CMS should test various area-level measures for their potential use in accurately accounting for differences in the social risk of patient populations. Further, more research is needed to understand the accuracy of any area-level measure compared with the gold standard of person-reported information. For example, the reliability of an area-level index should be assessed by comparing individual responses from a patient survey with their associated area-level index. ■

quality activities, providers would not necessarily use the payments to improve the quality of care they provide. Tying payments to specific activities would have a greater likelihood of raising the quality of care but would require an administrative structure to select projects and ensure accountability. The impact on access, quality, and costs for either of these forms of upfront payments would need to be tested before broad implementation in the Medicare program.

---

### Implementing a PAC VIP is a complex undertaking

---

Table 14-10 (p. 544) summarizes the key elements of a PAC VIP and the decisions policymakers would need to make when designing such a program. These elements include a small set of performance measures, strategies

to ensure reliable measure results, a system to distribute rewards with minimal cliff effects, a method to account for differences in social risk factors, and a way to fully distribute the incentive payments back to providers.

### Steps to implementing a PAC VIP

Implementing a PAC VIP involves many steps and will be a multiyear endeavor. First, a PAC PPS needs to be implemented so that setting-specific practice patterns begin to converge. Concurrently, CMS needs to begin aligning regulatory requirements. Until these two steps are complete, provider performance initially would need to be compared within settings because practice patterns reflect current regulatory requirements and the payment incentives embedded in the various PPSs. Setting-specific comparisons could be phased out over time, leading up to comparisons across settings.

Design element	Decisions
<p><b>Small set of performance measures.</b> Payments would be adjusted based on provider performance on a small set of outcome, patient experience, and resource use measures.</p>	<ul style="list-style-type: none"> <li>• Should all providers be scored on the same set of measures?</li> <li>• What measures should be used to gauge performance?</li> </ul>
<p><b>Strategies to ensure reliable measure results.</b> Measure results would reflect true differences in performance and not be attributable to random variation.</p>	<ul style="list-style-type: none"> <li>• What reliability standard should be used?</li> <li>• What strategies will ensure reliable measure results for as many providers as possible?</li> </ul>
<p><b>System to distribute rewards with minimal “cliff” effects.</b> A simple scoring approach would award points for every performance, achieved with minimal use of numeric thresholds that create statistical cliffs.</p>	<ul style="list-style-type: none"> <li>• Should a provider meet some minimum performance standard before it earns a reward?</li> </ul>
<p><b>Approach to account for differences in patients’ social risk factors using a peer-grouping mechanism, if necessary.</b> Providers would be stratified into peer groups based on the social risk of their patient populations if higher social risk is tied to poorer outcomes. Providers in peer groups with patient populations at high social risk would receive larger adjustments for attainments in quality compared with other providers.</p>	<ul style="list-style-type: none"> <li>• How should the social risk of a provider’s patient population be defined and measured?</li> <li>• How many peer groups should be used to differentiate providers?</li> </ul>
<p><b>Method to distribute the entire provider-funded pool of dollars.</b> All withheld funds would be distributed back to providers based on their performance.</p>	<ul style="list-style-type: none"> <li>• How large should the rewards and penalties be?</li> </ul>

Note: PAC (post-acute care), VIP (value incentive program).

CMS needs to select a set of performance measures that captures differences across providers. There will be trade-offs between selecting common measures and patient population-specific measures. These measures should eventually align with the quality incentives tied to MA and ACO payment. CMS also needs to choose a measure of social risk that has both a conceptual and an empirical relationship to outcomes.

Finally, CMS needs to design a methodology that scores providers’ performances, ensures reliable measure results, distributes rewards with minimal cliff effects, accounts for differences in the social risks of a provider’s patient population through peer grouping, and fully redistributes provider-financed incentive

payments to providers. Our PAC VIP methodology would be a good starting point for its deliberations.

**Multiple measurement issues confront the implementation of a PAC VIP**

The mixed results of our illustrative model underscore the many challenges in implementing a PAC VIP. These efforts are complicated by the multiple issues involved in defining and measuring performance. First, we need additional measures, most notably accurate measures of functional status and of patient experience. We urge CMS to make progress on developing these measures. Second, while having common measures across the four settings will facilitate comparisons of all providers, the measures may not be the best ones for any



given setting. CMS could consider a mix of common measures and measures tailored to specific patient populations. However, the resulting performance scores may be harder to compare across settings and could run into reliability issues with population-specific measures. Because a payment incentive program is not the only way to encourage providers to improve, CMS could publicly report measures specific to patient populations. Further, the set of performance measures is likely to evolve over time.

Third, developing accurate risk adjustment is elusive within one setting, let alone across four. There is an inherent trade-off between having a uniform risk-adjustment method—more in keeping with a uniform

payment system—and having setting-specific models that are more likely to be accurate for providers in any given setting but could undermine comparisons across settings.

Our investigation also illustrates that more work needs to be done to define measures of social risk. Ideally, a social risk measure should have both conceptual and empirical associations with outcomes. While the share of dual-eligible beneficiaries as a proxy for income is the best currently available measure of a patient's social risk, it does not capture all dimensions of social risk. CMS should test various area-level measures for their potential use in accurately accounting for differences in the social risk of individual patients. ■

## Endnotes

---

- 1 CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality, a U.S. government agency.
- 2 CMS has a similar measure, but its measure does not consider nursing home residents who return to the same facility as discharged to the community (and excludes them from the measure calculation).
- 3 Literature suggests 0.7 is an acceptable standard for reliability (Adams et al. 2010, Kao et al. 2011, Krell et al. 2014, Mehrotra et al. 2010, Scholle et al. 2008). Reliability values range from 0 to 1.0, where 0 indicates that the measure captures no real differences in performance (it captures only noise, or the random variation unrelated to performance) and 1.0 indicates that the measure captures all differences in real performance (all signal).
- 4 In our illustrative PAC VIP modeling, we excluded providers that did not meet the minimum case counts tied to 0.7 reliability for each measure using three years of claims data (i.e., 60 Medicare discharges over the three-year period that met the measure specifications). As a result, we excluded 7 percent of LTCHs, 23 percent of SNFs, 11 percent of IRFs, and 19 percent of HHAs because they did not have reliable results for all three measures (including providers with missing data needed to calculate a measure). In our modeling, these providers with very low Medicare volume would not be rewarded or penalized for their performance.
- 5 A census block group is the smallest geographical unit for which the United States Census Bureau publishes sample data. Typically, census block groups have a population of 600 to 3,000 people. In contrast, census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people.
- 6 In our PAC VIP model, we set each measure's continuous performance-to-points scale using a beta distribution, which helps to smooth the extremes of a distribution by providing estimates of a true percentile independent of associated issues such as ceiling effects.
- 7 The correlations between the share of fully dual-eligible beneficiaries and total performance points are consistent with the correlations between the shares of fully dual-eligible beneficiaries and each raw performance measure (i.e., before the performance was converted to points).
- 8 The Commission previously reported its findings on a VIP for SNFs also using fully dual-eligible beneficiaries as the measure of social risk (Medicare Payment Advisory Commission 2021b).
- 9 The shares of providers that gain or lose under the design are not necessarily equal. With each peer group, the incentive pool would be paid out fully each year, with the funds dispersed to providers that gain balancing out funds dispersed to those that lose. A large provider that gains under the VIP may have to be funded by multiple providers that lose.
- 10 To qualify for payment as an IRF, at least 60 percent of a facility's patients must require and benefit from intensive therapy for the treatment of at least 1 of 13 conditions. Beneficiaries must be able to tolerate "intensive" therapy, often interpreted as three hours a day.
- 11 States also vary in the share of their total Medicaid spending devoted to LTSS. In 2018, spending ranged from 18 percent in Arizona to 52 percent in North Dakota (Murray et al. 2021).
- 12 The qualifying criteria are a preceding acute care hospital stay and either at least 3 days in an intensive care unit during the hospital stay or mechanical ventilation services for at least 96 hours during the LTCH stay. The median risk score for LTCH patients was 3.7 compared with 2.6 for SNF patients and 2.1 for IRF patients.

## References

---

- Adams, J. L., A. Mehrotra, J. W. Thomas, et al. 2010. Physician cost profiling—Reliability and risk of misclassification. *New England Journal of Medicine* 362, no. 11 (March 18): 1014–1021.
- Agarwal, R., J. M. Liao, A. Gupta, et al. 2020. The impact of bundled payment on health care spending, utilization, and quality: A systematic review. *Health Affairs* 39, no. 1 (January): 50–57.
- Agency for Healthcare Research and Quality, Department of Health and Human Services. 2020a. Section 2: Why improve patient experience? <https://www.ahrq.gov/cahps/quality-improvement/improvement-guide/2-why-improve/index.html>.
- Agency for Healthcare Research and Quality, Department of Health and Human Services. 2020b. Talking quality: Reporting to consumers on health care quality. <https://www.ahrq.gov/talkingquality/index.html>.
- Agency for Toxic Substances and Disease Registry. 2021. CDC/ATSDR Social Vulnerability Index. <https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>.
- Aligning Forces for Quality. 2009. *How to display comparative information that people can understand and use: Research conducted on behalf of the Robert Wood Johnson Foundation*. Washington, DC: Aligning Forces for Quality. <http://forces4quality.org/how-display-comparative-information-people-can-understand-and-use.html>.
- Arling, G., V. Cooke, T. Lewis, et al. 2013. Minnesota's provider-initiated approach yields care quality gains at participating nursing homes. *Health Affairs* 32, no. 9 (September): 1631–1638.
- Assistant Secretary for Planning and Evaluation, Department of Health and Human Services. 2020a. *Report to Congress: Social risk factors and performance in Medicare's value-based purchasing program*. Washington, DC: ASPE. [https://aspe.hhs.gov/sites/default/files/migrated\\_legacy\\_files/195191/Second-IMPACT-SES-Report-to-Congress.pdf](https://aspe.hhs.gov/sites/default/files/migrated_legacy_files/195191/Second-IMPACT-SES-Report-to-Congress.pdf).
- Assistant Secretary for Planning and Evaluation, Department of Health and Human Services. 2020b. *Report to Congress: Social risk factors and performance under Medicare's value-based purchasing programs: A report required by the Improving Medicare Post-Acute Care Transformation (IMPACT) Act of 2014*. Washington, DC: ASPE. <https://aspe.hhs.gov/system/files/pdf/253971/ASPESESRTCfull.pdf>.
- Assistant Secretary for Planning and Evaluation, Department of Health and Human Services. 2016. *Report to Congress: Social risk factors and performance under Medicare's value-based purchasing programs*. Washington, DC: ASPE.
- Bennett, K. J., and J. C. Probst. 2016. Thirty-day readmission rates among dual-eligible beneficiaries. *Journal of Rural Health* 32, no. 2 (Spring): 188–195.
- Centers for Disease Control and Prevention. 2021. PLACES: Local data for better health. <https://www.cdc.gov/places/index.html>.
- Gage, B., M. Morley, L. Smith, et al. 2012. *Post-Acute Care Payment Reform Demonstration: Final report, volume 1 of 4*. Prepared under contract to the Centers for Medicare & Medicaid Services. Baltimore, MD: CMS.
- Garrett, B., C. Carter, and D. Wissoker. 2021. Evaluation of the reliability of Medicare spending per beneficiary for post-acute care. *Medical Care* 59, no. 8 (August 1): 721–726.
- Institute of Medicine. 2013. *Interim report of the committee on geographic variation in health care spending and promotion of high-value care: Preliminary committee observations*. Washington, DC: The National Academies Press.
- Jaffery, J., and D. G. Safran. 2021. Addressing social risk factors in value-based payment: Adjusting payment not performance to optimize outcomes and fairness. *Health Affairs* blog. April 19. <https://www.healthaffairs.org/doi/10.1377/hblog20210414.379479/full/>.
- Kao, L. S., A. A. Ghaferi, C. Y. Ko, et al. 2011. Reliability of superficial surgical site infections as a hospital quality measure. *Journal of the American College of Surgeons* 213, no. 2 (August): 231–235.
- Keohane, L. M., D. G. Stevenson, S. Freed, et al. 2018. Trends in Medicare fee-for-service spending growth for dual-eligible beneficiaries, 2007–15. *Health Affairs* 37, no. 8 (August): 1265–1273.
- Krell, R. W., A. Hozain, L. S. Kao, et al. 2014. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surgery* 149, no. 5 (May): 467–474.
- Marrufo, G., A. Bradley, J. Somers, et al. 2021. *CMS Bundled Payments for Care Improvement initiative models 2–4: Year 7 evaluation & monitoring annual report*. Falls Church, VA: Lewin Group. <https://innovation.cms.gov/data-and-reports/2021/bpci-models2-4-yr7evalrpt>.

- Marshall, M. N., P. G. Shekelle, H. T. Davies, et al. 2003. Public reporting on quality in the United States and the United Kingdom. *Health Affairs* 22, no. 3 (May–June): 134–148.
- Medicare Payment Advisory Commission. 2021a. *A data book: Health care spending and the Medicare program*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2021b. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2020. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2019a. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2019b. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2018. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2017a. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2017b. *Report to the Congress: Regional variation in Medicare Part A, Part B, and Part D spending and service use*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2016. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2014. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Mehrotra, A., J. L. Adams, J. W. Thomas, et al. 2010. Cost profiles: Should the focus be on individual physicians or physician groups? *Health Affairs* 29, no. 8 (August): 1532–1538.
- Murray, C., A. Tourtellotte, D. Lipson, et al. 2021. *Medicaid long term services and supports annual expenditures report: Federal fiscal years 2017 and 2018*. Chicago, IL: Mathematica Policy Research.
- Musumeci, M., P. Chidambaram, and M. O'Malley. 2019. *Medicaid financial eligibility for seniors and people with disabilities: Findings from a 50-state survey*. Washington, DC: Kaiser Family Foundation.
- National Academies of Sciences, Engineering, and Medicine. 2016a. *Accounting for social risk factors in Medicare payment: Criteria, factors, and methods*. Washington, DC: The National Academies Press.
- National Academies of Sciences, Engineering, and Medicine. 2016b. *Accounting for social risk factors in Medicare payment: Identifying social risk factors*. Washington, DC: The National Academies Press.
- Navathe, A. S., E. J. Emanuel, A. S. Venkataramani, et al. 2020. Spending and quality after three years of Medicare's voluntary bundled payment for joint replacement surgery. *Health Affairs* 39, no. 1 (January): 58–66.
- Office of Disease Prevention and Health Promotion, Department of Health and Human Services. 2021. Social determinants of health. <https://health.gov/healthypeople/objectives-and-data/social-determinants-health>.
- Pozniak, A., M. Turenne, E. Lammers, et al. 2021. *Evaluation of the Home Health Value-Based Purchasing (HHVBP) Model: Fourth annual report*. Baltimore, MD: Centers for Medicare & Medicaid Services.
- Saliba, D., D. L. Weimer, Y. Shi, et al. 2018. Examination of the new short-stay nursing home quality measures: Rehospitalizations, emergency department visits, and successful returns to the community. *Journal of Health Care Organization, Provision and Financing* 55: 1–11.
- Scholle, S. H., J. Roski, J. L. Adams, et al. 2008. Benchmarking physician performance: Reliability of individual and composite measures. *American Journal of Managed Care* 14, no. 12 (December): 833–838.
- University of Wisconsin School of Medicine and Public Health. 2021. Neighborhood Atlas. <https://www.neighborhoodatlas.medicine.wisc.edu/>.
- Wadhwa, R. K., Y. Wang, J. F. Figueroa, et al. 2020. Mortality and hospitalizations for dually enrolled and nondually enrolled Medicare beneficiaries aged 65 years or older, 2004 to 2017. *JAMA* 323, no. 10 (March 10): 961–969.