# Report to Congress:

# Evaluation of the Independence at Home Demonstration

**U. S. Department of Health and Human Services**
**November 2018**

# Table of Contents

# List of Tables



# List of Figures

# Executive Summary

Section 1866E of the Social Security Act (the Act) (as added by Section 3024 of the Affordable Care Act) mandated a Medicare demonstration titled the "Independence at Home Medical Practice Demonstration Program" (IAH). The demonstration is intended to test a payment incentive and service delivery model that utilizes physician- and nurse-practitioner-directed home-based primary care teams designed to reduce expenditures and improve health outcomes for applicable beneficiaries, who are Medicare beneficiaries with multiple chronic conditions and a substantial burden of functional limitations. This report responds to the law's requirement for an independent evaluation of the demonstration and report to Congress that includes an analysis of the demonstration program on coordination of care, expenditures, applicable beneficiary access to services, and the quality of health care services provided to applicable beneficiaries (Section 1866E(g) of the Act). In this report, we present interim findings, which are based on the three years of the demonstration ending September 2015.[1,2] Results of succeeding years of the demonstration will be forthcoming in a future final report.

The demonstration tests a combination service-delivery and payment-incentive model. IAH relies on multidisciplinary teams led by physicians or nurse practitioners to meet the primary care needs of the target population comprehensively in a home setting.[3] Under this care model, often called "home-based primary care" (HBPC), IAH practices are expected to coordinate patient care, offer 24-hour-per-day accessibility every day of the week, and design and carry out patient-centered care plans. In return, practices can earn incentive payments if actual expenditures for applicable beneficiaries it enrolls are less than estimated spending targets. The incentive amount depends on the size of the savings, and is adjusted for performance on specified quality measures. Medicare also shares in any savings achieved.

This report[4] addresses the following evaluation question: What were the impacts of the IAH demonstration? Specific research questions are:

- Did the demonstration, in which practices were offered payment incentives tied to quality measures, lead to expenditure reductions for Medicare?
- Did the demonstration lead to changes in health care utilization, such as reductions in acute-care stays and emergency department use?
- Did the demonstration lead to changes in health outcomes?
- Were the IAH beneficiaries satisfied with their primary care?

[1] The Medicare Independence at Home Medical Practice Demonstration Improvement Act of 2015 (Pub. L. 114-39) extended the demonstration from three to five years. Most practices ended their first three years in June 2015; others ended their first three years in September 2015. The Bipartisan Budget Act of 2018 (Pub. L. 115-123) extended the demonstration from five to seven years.

[2] Selected qualitative descriptive information pertaining to the fourth year are included in this report as well.

[3] Home settings include private homes and residences such as assisted living facilities and domiciliary care homes.

[4] The contents of this report are based on an independent evaluation of the IAH demonstration.

- Did the IAH practices change their approach to care delivery in response to the demonstration and, if so, how did they change?

The IAH demonstration tests both a payment incentive and service delivery model. However, the evaluation of the demonstration focuses on whether the possibility of earning an incentive payment tied to quality measures resulted in improved outcomes for beneficiaries enrolled in IAH practices, not on the impact of the HBPC delivery model relative to the typically office-based service delivery experienced by most fee-for-service beneficiaries. Because the IAH practices were furnishing HBPC prior to the start of the demonstration and many IAH beneficiaries were receiving HPBC prior to the start of the demonstration, the size of the demonstration is too small to permit a robust analysis of the effects of HBPC as distinct from usual care for similarly ill beneficiaries. Therefore, when this report refers to the impact of the demonstration, it means the impact of the incentive structure of the demonstration on the participating practices.

In this report, the approach to estimating changes in expenditures differs from that used to determine the demonstration's payment incentives, which were released by CMS in 2015 and 2016 for years 1 and 2 of the demonstration, respectively. The evaluation addresses the question of expenditures using the IAH practices' collective past performance compared to their performance at baseline (the year before the demonstration). The evaluation uses a comparison group to account for changes over time that would have affected the practices in the absence of the demonstration. In contrast to the evaluation's approach, incentive payment results for years 1 and 2 of the demonstration generally were based on comparing the Medicare expenditures for IAH beneficiaries with expenditures of a contemporaneous comparison group. This report's estimated reductions in expenditures do not take into account incentive payments made to IAH practices.

## Background

The statute set forth requirements for participation of health care providers and beneficiaries, and it established basic structural features of the demonstration, such as rules for determining incentive payments and the size of the demonstration. In particular, in selecting practices to participate in the demonstration, the Secretary was required to limit the number of practices so that the number of beneficiaries in the demonstration did not exceed 10,000. Among other requirements, practices were to be experienced in delivering HBPC with a team approach. Beneficiaries qualified only if they were Medicare fee-for service beneficiaries, had two or more chronic illnesses and functional dependencies, had a nonelective hospital admission within the past 12 months, and had used rehabilitation services within the past 12 months.

To address the enrollment limit, CMS approved 18 practices[5] to participate in IAH at the start of the demonstration in 2012. Although all practices specialized in HBPC, they varied in their geographic location (14 states and the District of Columbia), size, structure, and organizational

[5] We use the term "practice" to refer to any practitioner group that participated in IAH, including any consortia of practitioner groups that participated as a single entity for purposes of the demonstration.

affiliation. Broadly speaking, practices could be categorized into three major subgroups: units of the Visiting Physicians Association, academic or medical center practices, and independent practices. The extent to which the participating practices are nationally representative is unknown and, given the limited number of practices, the samples are relatively small from a statistical point of view.

Two important implications follow from these characteristics of the demonstration. Because the practices volunteered to participate in the demonstration and we have not compared the characteristics of the participating practices to HBPC practices more generally, these results apply only to the participants in the demonstration and not to HBPC practices more generally. Second, because the demonstration involves a limited number of practices, and because these practices have small practice sizes, a relatively large change in any given outcome is necessary to engender confidence that a real change occurred. For example, the evaluation analysis can reliably detect an impact on monthly beneficiary expenditures of approximately 7 percent or more.

## Results: Expenditures

We found indications that the demonstration may have reduced total Medicare expenditures and inpatient hospital expenditures. Moreover, expenditure estimates indicated consistent, though small, yearly reductions in total Medicare Part A and B expenses per beneficiary per month (PBPM) (Table ES-1). This measure declined by $123, $31, and $177 in Years 1, 2, and 3, respectively. The average reduction over the entire three years was $111 per beneficiary per month, or 2.5%. In the aggregate, before taking into account incentive payments made to IAH practices, the demonstration could have lowered Medicare expenditures by $25 million for the 221,379 beneficiary-months of care analyzed for this report. However, total savings could have been considerably higher or lower, as the changes were not statistically significant. This means evidence of savings is inconclusive.

The possibility that real change occurred is supported by qualitative information gathered annually from the practices. Many reported that they instituted new mechanisms designed to achieve the IAH quality targets and to strengthen the effectiveness of their primary care services. However, in view of the inconclusive results, we believe analysis of the additional years of the demonstration is necessary, as further experience may allow us to make a final judgment about its impact.

**Table ES1. IAH demonstration annual and cumulative beneficiary-months of care, savings per beneficiary-month, and total savings: Years 1 to 3**

*Aggregate estimates of expenditures reductions suggest the demonstration may have lowered expenditures in its first three years. However, no estimates of a reduction in total Medicare expenditures were statistically significant in any year or on average during the entire three years.*

| | Year 1 | Year 2 | Year 3 | Cumulative total, Years 1 to 3 |
|---|---|---|---|---|
| Beneficiary-months of care | 79,396 | 69,768 | 72,215 | 221,379 |
| Savings per beneficiary per month | -$123 | -$31 | -$177 | -$111[a] |
| Total savings | -$9,741,494 | -$2,193,523 | -$12,758,376 | -$24,693,394 |
| 90% confidence interval | -$22,412,928; $2,929,941 | -$18,161,992; $13,774,946 | -$31,413,985; $5,897,234 | -$69,740,518; $20,353,731 |
| 80% confidence interval | -$19,616,739; $133,752 | -$14,638,251; $10,251,205 | -$27,297,276; $1,780,525 | -$59,800,028; $10,413,241 |

Note: No results were statistically significant. Total savings estimates varied each year due to changes in savings per beneficiary-month and in total months of care. Savings estimates do not take into account incentive payments to IAH practices.
[a]Savings of $111 per beneficiary per month were calculated as an average value based on all beneficiary months, regardless of the year.

We note that the demonstration uses a different methodology for measuring spending for the purposes of incentive payments for practices. For year 1, incentive payments of $11.7 million were paid to 9 practices, and for Year 2 seven practices out of 15 that completed the first two years earned incentive payments in the amount of $5.3 million. Due to differing purposes and methodologies, the incentive payment results are not comparable to the evaluation's results for demonstration savings. The annual and cumulative savings amounts in this report are the estimated reductions associated with the impact of the demonstration—that is, how the sites changed over time. This comparison strategy—which measures practices against their performance in the year before the demonstration—differs from the strategy CMS used for calculating incentive payments. CMS determined that the appropriate baseline for determining incentive payments is not past performance; rather, the appropriate baseline is the contemporaneous expenditures of a benchmark population. Because the two objectives (evaluation vs. incentive payments) and their respective analytic strategies are different, and because of numerous other technical differences between the two analyses, the incentive payments that CMS reported during the course of the demonstration are independent of the savings we report here as outcomes of the demonstration. (For CMS' public release of incentive payment results, see https://innovation.cms.gov/initiatives/independence-at-home.)

## Results: Utilization

We also examined multiple acute care-related measures to assess whether the demonstration led to reductions in acute care utilization. Among them, we found the following measures tended to decrease each year, with each achieving a statistically significant reduction in Year 3:

- emergency department (ED) visits leading to hospitalization showed statistically significant declines in both Years 1 (-0.07 per person annually, or -4.8 percent) and 3 (-0.12, or -8.4 percent);
- the proportion of beneficiaries with at least one unplanned hospital readmission in the year decreased by 1.71 percentage points (-8.7 percent) in Year 3; and
- the number of preventable hospital admissions registered a decrease of 0.03 per person per year (-7.6 percent) in Year 3.

Utilization of skilled nursing facilities, hospices, inpatient rehabilitation, and home health agencies remained unchanged as a result of the demonstration.

Although impact estimates for the demonstration as a whole are generally modest, one subgroup of practices, independent practices, did make notable progress in key outcomes. As a group, these practices had strong, statistically significant reductions in expenditures, hospital admissions, ED visits, and the probability that a beneficiary would have an unplanned readmission. For example, for the independent practices as a group, annual savings in Medicare expenditures PBPM ranged between 9.1 percent and 13.7 percent across the demonstration years 1 through 3.

## Results: Health outcomes

Two important measures of health outcomes came from Medicare administrative data: mortality and the rate of long-term care (LTC) placement, which we would expect to occur less often for beneficiaries receiving HBPC from IAH practices. Our estimates suggest the demonstration did not affect mortality. An unexpected finding was that IAH beneficiaries' rate of entry into LTC did not change as much as the comparison group's rate changed. Although both groups showed a decline in the rate of LTC entry, the smaller decrease in IAH beneficiaries' entering LTCs is a potentially unfavorable finding. Further analysis is necessary to confirm this result.

## Results: Beneficiary satisfaction with primary care

Overall satisfaction with the HBPC services, as reported by the IAH beneficiaries and their caregivers surveyed, was high and similar to satisfaction levels among Medicare beneficiaries in general. More than nine in ten beneficiaries and their caregivers reported that they were satisfied or very satisfied with the overall quality of care they had received from the IAH practice in the past six months. When asked specifically how much he or she likes receiving in-home care compared to primary care in an office or clinic, more than eight in ten beneficiaries said "a lot" or "somewhat" more. However, results on perceptions of physician care quality

tended to suggest IAH beneficiaries' views on physician quality issues were somewhat less favorable than views of a similarly ill comparison group. For example, when asked whether the practitioner often seems to be in a hurry, a larger minority of IAH beneficiaries said yes, compared to a comparison group.

## Results: Practice changes

The IAH practices' activities in response to the demonstration appeared to build over time, as they modified care delivery processes, invested in quality improvement, and strengthened relationships with hospitals and other care partners, such as home health agencies. Their accumulation of experience with practice change over the course of the demonstration could be related to the improving trend between the second and third years in several acute care measures: ED visits leading to hospital admission, readmissions, and preventable admissions. As previously noted, those decreases in utilization were statistically significant and largest in Year 3.

## Results: Operation of incentive payments for HBPC practices

As specified in the IAH statute, subject to performance on quality measures, IAH practices were eligible to receive incentive payments, which were to be based on performance against a target spending level for each practice. The target was to reflect the estimated amount that would have been spent under Parts A and B for items and services furnished to IAH beneficiaries in the absence of the demonstration. The incentive payment is a portion of the difference between the target and actual spending, after adjustments and allocation of Medicare's share. One lesson learned from this demonstration is that, because of the unique characteristics of IAH beneficiaries and the frequently small size of practices in IAH, estimating a target reflecting the amount that would have been spent absent the demonstration presented technical difficulties in the HBPC context, even though several different methodologies have been used. Therefore, implementing incentive payments is challenging.

## Conclusion

Results of the first three years of the demonstration are promising but inconclusive, in large part due to the small size of the demonstration. We see indications that the demonstration may have had some small, favorable effects in each year on important measures of Medicare utilization and expenditures. The changes that we observed suggest that the practices may have the ability to increase the efficiency with which they manage beneficiaries' care. As for health impacts, we find no serious indications of worsening outcomes. Nor do we find that IAH beneficiaries are any less satisfied with their primary care than their peers who do not receive primary care in the home, although IAH beneficiary responses to several specific survey questions on physician care quality were somewhat less favorable. If the practices continued to make performance improvements in the demonstration's later years, we may see stronger results from adding additional years of demonstration data to our analysis. A final report from the evaluation will present results for succeeding years of the demonstration.

# Background

Section 1866E of the Social Security Act (as added by Section 3024 of the Affordable Care Act) mandated a Medicare demonstration titled the "Independence at Home Medical Practice Demonstration Program" to test a payment incentive and service delivery model that utilizes physician and nurse practitioner directed home-based primary care teams designed to reduce expenditures and improve health outcomes for applicable beneficiaries. Primary care practices serving these patients could earn payment incentives beyond the standard reimbursement for home visits under the Medicare fee schedule. This report responds to the law's requirement for an independent evaluation of the demonstration and report to Congress (Section 1866E(g)). In this report, we present interim findings, which are based on the three years ending September 2015.[6,7] Results of the entire demonstration will be issued in a future final report.

In establishing the IAH demonstration, the legislation set forth certain requirements, including the following:

- Applicable beneficiaries enrolled in IAH practices were to be high-need Medicare fee-for-service beneficiaries who had two or more chronic illnesses, two or more functional dependencies, a nonelective hospital admission within the past 12 months, and acute or subacute rehabilitation services within the past 12 months.
- The physicians, nurse practitioners, or physician assistants of practices participating in IAH were to care for the beneficiaries as part of a team experienced in delivering home-based primary care (HBPC) to high-cost chronically ill beneficiaries, with care delivery featuring 24-hour accessibility, individualized care plans, and the use of electronic health information systems, remote monitoring, and mobile diagnostic technology.
- A participating practice was eligible to receive an incentive payment if actual expenditures for a year for the applicable beneficiaries it enrolls are less than estimated spending targets for those beneficiaries that year. The incentive payment is subject to performance on quality measures. The law required the Secretary to establish quality performance standards.
- The demonstration would involve a limited number of practices, with the number selected so that the number of applicable beneficiaries that may participate in the demonstration does not exceed 10,000.

We provide key details of the demonstration requirements below.

[6] The Medicare Independence at Home Medical Practice Demonstration Improvement Act of 2015 (Pub. L. 114-39) extended the demonstration from three to five years. Most practices ended their first three years in June 2015; others ended their first three years in September 2015. The fourth and fifth years began for all practices in October 2015. The Bipartisan Budget Act of 2018 (Pub. L. 115-123) extended the demonstration from five to seven years.
[7] Selected qualitative descriptive information pertaining to the fourth year is included in this report as well.

### Beneficiary and practice requirements

Enrolled beneficiaries had to meet program-related, utilization-history, and health-status criteria, including needing the assistance of another person in at least two activities of daily living (ADL); at least two chronic illnesses typically associated with high costs, such as chronic obstructive pulmonary disease or stroke; a recent hospital stay, and receiving recent acute or subacute rehabilitation services (see Figure 1). The beneficiary criteria collectively signify poor health status and are suggestive of difficulties accessing primary care in doctors' offices--characteristics which could lead to unnecessary use of acute care in the emergency department and the hospital. CMS provided practices with utilization history and program-related information to assist them in verifying beneficiary eligibility for enrollment. The practices themselves determined whether beneficiaries met the health status requirements.

**Figure 1: Patient requirements upon IAH enrollment**

- Covered by Part A and Part B
- At least two high-cost chronic conditions
- Requires assistance of another person (supervision, cueing, or hands-on help) for 2 or more activities of daily living such as bathing or dressing
- Had a nonelective hospital admission in past 12 months
- Used acute or subacute rehabilitation services (skilled nursing/inpatient rehabilitation/home health) in past 12 months
- Not in long-term care or hospice
- Not in a PACE or Medicare Advantage plan

Participating physician- or nurse practitioner (NP)-led practices were required to be experienced serving high-cost, chronically ill patients at home, to work in teams that may include social workers, pharmacists, and others, to be available 24 hours a day 7 days a week, and to follow patient-centered care plans. Practices were required to use electronic health information systems, remote monitoring, and mobile diagnostic technology. Practices were required to furnish services to at least 200 IAH-eligible beneficiaries per year and to report on quality measures specified by the Secretary.

Three of the original 18 practices[8] left the demonstration within the first three years. Participants that dropped out were unable to meet demonstration enrollment size requirements, the data reporting responsibilities, or both. A fourth practice completed the first three years but did not participate in the two-year extension. This report's quantitative analysis of administrative data comes from the remaining 14 practices. Qualitative information in this report in general comes from the same 14 practices. Beneficiary survey samples were drawn from beneficiaries who entered the demonstration in 2012-2014.

[8] We use the term "practice" to refer to any practitioner group that participated in IAH, including any consortia of practitioner groups that participated as a single entity for purposes of the demonstration.

### Incentive payments

In return for generating savings and achieving standards on quality measures, the practices could share in savings each year through an incentive payment. The incentive payment design in the legislation required the Secretary to determine an estimated annual spending target representing an estimate of what would have been spent for items and services furnished to IAH beneficiaries under Medicare Parts A and B in the absence of the demonstration. If a practice exceeded the target, it would not receive an incentive payment. If actual expenditures (including the incentive payment) of the applicable beneficiaries enrolled by a practice were at least five percent below the target, then the practice was eligible for an incentive payment, subject to the practice's performance on quality measures. CMS determined that the appropriate spending target should be estimated from the expenditures of a contemporaneous population with a similar clinical and demographic profile, rather than relative to the past performance of the IAH practices.

The law specified that eligibility for the incentive payment was subject to performance on quality measures, and it authorized the Secretary to decide how quality measures would affect the financial incentive. CMS established the following policy: A practice was eligible to share in a maximum of 80% of any savings beyond the first five percent, according to a schedule that depended on the number of quality measures achieved. If a practice achieved all six quality measures tied to payment, then it earned the entire available maximum. If a practice met five, four, or three quality measures, it earned, respectively, 83%, 67%, or 50% of the available maximum.

The six quality measures tied to payment were selected by the Secretary to reflect processes and outcomes that promote effective primary care. For example, the process-related measures included making contact with patients around the time of any hospitalizations or ED visits, and outcomes measures included risk-adjusted hospital readmission rates (see Figure 2). For the three process measures, CMS established a required minimum rate of achievement for each one; for example, at least 80% of IAH beneficiaries had to have their care preferences documented annually. For the three outcome measures, utilization rates among IAH-enrolled beneficiaries had to be as low as or lower than utilization rates among IAH-eligible beneficiaries not receiving home-based primary care (HBPC) in the

**Figure 2: Quality measures tied to incentive payments**

***Process:***

**1**. Follow-up contact within 48 hours of hospital admissions, discharges, and ED visits (required for at least 50% of events)

**2**. Medication reconciliation in the home within 48 hours of hospital discharges and ED visits (required for at least 50% of events)

**3.** Patient preferences documented annually (required for at least 80% of enrolled patients)

***Outcome:***

**4.** Hospital admissions within 30 days for ambulatory care sensitive conditions (diabetes, heart failure, chronic obstructive pulmonary disease)

**5.** All-cause hospital readmissions within 30 days

**6.** ED visits for ambulatory care sensitive conditions (diabetes, heart failure, chronic obstructive pulmonary disease)

same geographic area. The Secretary required additional quality measures that were not tied to payment; these included documenting goals for the patient and family caregiver; conducting screenings and assessments (depression, home safety, caregiver stress, fall risk, cognitive deficits); managing symptoms (pain, shortness of breath, cognitive deficits, fatigue, sleep disturbances); and managing medications.

For more information about the implementation of the demonstration, see Appendix A.

### Incentive payment methodology

In the demonstration solicitation, CMS described the incentive payment methodology it intended to use, and the practices selected for participation signed agreements accepting that methodology and other terms and conditions. A series of developments that unfolded later revealed several challenges associated with designing and implementing a methodology to determine the incentive payment.

When the practices signed their participation agreements, the methodology they agreed to was derived from a risk-adjusted target based on average monthly expenses in each IAH beneficiary's county. This approach was referred to as the actuarial methodology.

However, during the second year of the demonstration, independent analyses provided to CMS by the American Academy of Home Care Medicine suggested that such a methodology may understate spending for the frail and medically complex type of beneficiary who qualified for IAH; and therefore the calculations would tend to result in smaller savings for the practices than if the spending for this type of beneficiary population was fully accounted for in the methodology.

As a result, CMS reconsidered its technical approach to determining the spending targets, and offered practices a new methodology that used comparison groups identified in administrative data. The treatment group consisted of IAH beneficiaries enrolled by the practices, as long as their eligibility could be confirmed in administrative data. All but one practice agreed to switch to the new comparison-group-based methodology. Incentive payments totaling $11.7 million for the first year of the demonstration were paid to 9 practices (including the practice that remained with the original methodology). The other 8 practices that completed the first year did not achieve savings or had insufficient savings to earn an incentive payment.

Due to CMS-identified issues with the comparability of the treatment and comparison samples that became clear during the analysis of Year 2 results, further methodology modifications preceded the release of payments for Year 2. The methodology modifications involved revising the comparison group matching procedures and conducting analysis of two samples, both enrollees and an expanded sample that included all beneficiaries who were treated by the practices (and who also met IAH eligibility criteria in administrative data). Practices that had switched to the comparison-group methodology for Year 1 had the option to agree to use the modified version of that methodology in Year 2. Seven practices out of 15 that completed the

first two years earned incentive payments in the amount of $5.3 million. (See Appendix B for details about the methodology modifications.)

During Year 5 of the demonstration, prior to the payment of Year 3 incentives, representatives of the practices requested that CMS consider returning to an actuarial methodology but with certain technical changes to it. For example, the practices recommended stratifying the IAH population into beneficiaries with end stage renal disease (ESRD) and non-ESRD beneficiaries. CMS, taking into account this input from the practices, has been considering a revised methodology for the demonstration, and results from Years 3 through 5 of the demonstration have not yet been released. This revised methodology would be the fourth methodology.

Over time, experience with the data and the methodologies revealed a set of important lessons associated with applying the spending target requirement in the HBPC context. First, HBPC practices are often small and lack sufficient numbers of eligible beneficiaries to assure savings estimates with high statistical confidence. The statute requires that spending targets include a risk corridor that takes into account normal variation in expenditures, with the size of the corridor being related to the number of applicable beneficiaries furnished services by each practice. CMS implemented this requirement by testing each practice's savings percentage for statistical significance at the 90 percent and 95 percent confidence levels. With small numbers of patients, savings estimates generally need to be relatively large to be statistically significant. For example, in some instances even savings percentages between five and ten percent for IAH practices were not statistically significant. Further, with small numbers of patients, minor changes in the sample can lead to relatively large changes in results for a given practice. When results are unstable from year to year, practices seeking feedback information from savings results would not receive a clear signal of how well they are performing.

Second, the high mortality characteristic of the IAH population can add another source of instability in the estimates, and also lead to uncertain financial results due to random factors. Unusually high costs often attend the weeks and months leading up to death. Death is inherently difficult to model statistically, and therefore mortality events that randomly concentrate in either the treatment or the comparison group (if one is used) can inordinately affect financial outcomes.

Third, in the absence of a randomized design, experience with a comparison group methodology that relied on administrative data to construct a comparison group closely matched to treatment beneficiaries highlighted additional issues when analysis focuses on a population like the IAH beneficiaries. One challenge was finding comparison groups that are sufficiently comparable to IAH beneficiaries in their health characteristics. Some unusual health conditions that collectively drive high costs are likely poorly represented in administrative data samples, leading to differences between the treatment and control groups that are not able to be controlled but that may be important to the results. For example, many IAH beneficiaries are homebound, but methods of reliably identifying homebound patients in administrative data are not available for the purpose of constructing a well-matched comparison group. Another challenge stemmed from the voluntary decision of beneficiaries to be seen by IAH practices,

which specialize in frail and limited-mobility patients. Beneficiary preferences regarding their desired approach to care—which are unmeasurable with administrative data—likely play a role when a beneficiary chooses HBPC. The typical office-based practice may not be as attractive to patients or families who especially favor a model oriented to care in the home whenever possible. If the same preferences place relatively little emphasis on readily accessing specialist and institutional services, expenditures could be lower regardless of the model of care chosen by the beneficiary.

We also learned how enrollment bias can arise in this type of demonstration, despite efforts to minimize it. Furthermore, ways of mitigating the bias can have limitations. Before making the changes to the comparison–group methodology for the Year 2 results, we detected some likely causes of noncomparability of the enrolled and comparison samples attributable to enrollment procedures at the practices. Because we used comparison samples drawn from administrative data, enrollment procedures at the practices needed to be timely and complete, and avoid introducing clinical judgment in enrollment decisions. However, we found that practices' enrollment procedures and clinical decisions led to both under- and over-enrollment relative to the information available in administrative data for determining which beneficiaries were eligible for IAH. Close examination of the reasons for discrepancies, as well as other data analysis, led to the conclusion that enrollment procedures at the practices couldn't be replicated in administrative data when selecting comparison group members. One difficulty was that enrollment procedures must be timely to avoid omitting some patients from the treatment group (e.g., patients who die or become ineligible shortly after joining the practice's panel of patients). The counterparts of such patients unavoidably are selected when extracting a comparison sample from administrative data. Another example was that the measure of ADL limitations often differed between the administrative data[9] and the practices' medical records. One reason is that ADL status may shift in individuals over time, and it is impossible to equalize the timing of measurement between the two sources of ADL information. Another likely reason is that standardization of ADL measurement may be difficult to achieve across the two sources.

Our solution to mitigate enrollment bias in the results was to make an adjustment to the savings estimate we derived from analysis of the enrollee sample.[10] The adjustment was based on the analysis of samples selected under the exact same selection rules applied to administrative data for both treatment and comparison subjects, with some allowances for beneficiaries who continued as patients of the IAH practices beyond their first year in the demonstration. In other words, to derive the adjustment factor, in general we did not use the IAH practices' actual enrollment of beneficiaries, but rather used the administrative data to identify eligible beneficiaries treated by the practices for inclusion in the sample. When relying

[9] ADL information in administrative data came from assessments required during rehabilitation stays (including home health, as well as skilled nursing stays and inpatient rehabilitation hospital stays).

[10] A significance test of the enrolled-sample-based estimate was used to determine whether a practice would be eligible for an incentive payment. If results indicated that the practice was eligible, the adjusted savings estimate was also tested for statistical significance in order to implement the risk corridor requirement.

on a comparison group extracted from administrative data, the administrative data provide the only consistently collected common source of information for identifying comparable samples of treatment and comparison subjects. However, this approach means that the final treatment samples used for determining the adjustment to the savings amount differed in some instances from the practices' actual determination of which patients were IAH-eligible. This was one concern that practice representatives cited when they requested that CMS consider implementing a fourth methodology, one based on the actuarial approach. The practices also were concerned that using a comparison-group method entailed delays in paying incentives.

For further details about the incentive payment methodology in the demonstration, see Appendix B.

It should be noted that, due to differing purposes and methodologies, the incentive payment results are not comparable to the evaluation's results for demonstration savings. The annual and cumulative savings amounts in this report are the estimated savings associated with the impact of the demonstration incentives—that is, how the sites changed over time. This comparison strategy—which measures practices against their performance in the year before the demonstration—differs from the strategy CMS used for calculating incentive payments. CMS determined that the appropriate baseline for determining incentive payments is not past performance; rather, the appropriate baseline is the contemporaneous expenditures of a benchmark population. Because the two objectives (evaluation vs. incentive payments) and their respective analytic strategies are different, and because of numerous other technical differences between the two analyses, the incentive payment results that CMS reported during the course of the demonstration are independent of the savings we report here as outcomes of the demonstration. (For CMS' public release of incentive payment results, see https://innovation.cms.gov/initiatives/independence-at-home.)

## Who Are the IAH Beneficiaries?

The IAH beneficiaries are among the very sickest in the Medicare fee-for-service (FFS) population. Their mortality is approximately 19 percent annually, compared to about 4.5 percent in the Medicare population overall[11]. Average monthly Part A and B expenditures per IAH beneficiary exceeded $4,000 ($4,191 to $4,374 depending on the year of the demonstration). In comparison, average Medicare payments for the entire year in 2013 were $9,231 for all FFS beneficiaries nationally.[12] The treatment group used to generate these statistics and others in our evaluation analyses consists of beneficiaries who received home visits from the IAH practices and met the demonstration eligibility criteria each year, according

[11] Krumholz HM et al., JAMA. 2015 Jul 28;314(4):355-65.

[12] See https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMSProgramStatistics/2013/Downloads/MDCR_UTIL/CPS_MDCR_UTLZN_AB_1.pdf. Statistic based on FFS beneficiaries enrolled in Part A and/or B. The vast majority of such beneficiaries are enrolled in both Parts.

to administrative data.[13] (See Appendix C for information about the practices' IAH beneficiaries not in our treatment group sample.) Beneficiaries in this group accounted for a majority of the practices' qualifying Medicare beneficiaries.

### Demographic characteristics and health conditions

Even when compared to non-participating FFS beneficiaries who met the same IAH eligibility requirements for the demonstration, IAH beneficiaries more often had characteristics typically associated with serious health problems requiring high expenditures (see Figure 1 for requirements). Prior to carefully matching IAH beneficiaries to a comparison group on dozens of characteristics, we identified all FFS beneficiaries in the local area who did not use HBPC but would have met the IAH eligibility criteria according to Medicare administrative data. Comparison of the IAH beneficiaries with all those eligible FFS beneficiaries in the same area shows that IAH beneficiaries were a distinct subgroup (Table 1). IAH beneficiaries were older, and more likely to be black, dually eligible for Medicare and Medicaid, and originally entitled to Medicare due to disability (Table 1). IAH beneficiaries had more chronic conditions, and a higher burden of illness as measured by hierarchical condition category (HCC) scores.[14] IAH beneficiaries had notably higher rates of dementias, paralysis, pressure ulcers, and depression. In interviews, the IAH practices commonly characterized their patients as frail elderly, who often were homebound.

[13] We determined IAH eligibility using CMS routine administrative data (enrollment files, claims, and patient assessments submitted by skilled nursing facilities, home health agencies, and inpatient rehabilitation hospitals). Whereas demonstration rules allowed practices to determine eligibility based on their own health assessments and allowed patients to continue in the demonstration from year to year, regardless of whether their health status and utilization history continued to support eligibility after their time of entry, the evaluation's data come only from beneficiaries who met the qualifying criteria for purposes of each year independently. We determined health status and other qualifying criteria by interrogating administrative data in a consistent manner each year, including the two baseline years ending the day before the demonstration began. This approach enabled us to deploy an evaluation design capable of assessing the impact of the demonstration without the risk of confounding by a change in inclusion criteria for the study. Our treatment samples exclude some enrolled beneficiaries because we could not confirm their eligibility in administrative data. See Appendix C for further information.

[14] The CMS-HCC scoring system was designed for risk adjustment in the Medicare Advantage program. We computed the score using the individual's claims history over the 12 months before the date of IAH eligibility.

## Table 1. Comparison of IAH beneficiaries and beneficiaries in the same geographic areas who were IAH-eligible but did not use HBPC: Year 3

| Beneficiary characteristics | IAH beneficiaries N=7,553 | Local-area beneficiaries who met IAH qualifications but did not use HBPC* N=214,603 |
|---|---|---|
| **Total** | 100% | 100% |
| Dual eligible | 38% | 27% |
| Female | 67% | 62% |
| Race | | |
| White | 72% | 76% |
| Black | 24% | 18% |
| Other | 5% | 6% |
| Original reason for Medicare entitlement | | |
| Age | 68% | 74% |
| Disability | 31% | 24% |
| ESRD or ESRD and disability | 1% | 2% |
| Age | | |
| <65 | 16% | 14% |
| 65-79 | 32% | 41% |
| >79 | 53% | 45% |
| No. of CCW chronic conditions | | |
| <6 | 12% | 18% |
| 6–9 | 48% | 50% |
| >9 | 40% | 32% |
| Depression (CCW flag) | 55% | 42% |
| HCC score | 3.622 | 3.336 |
| HCC group | | |
| HCC 8, metastatic cancer | 1% | 4% |
| HCC 9-10, lung, lymphoma and other cancers | 3% | 6% |
| HCC 11-12, colorectal, bladder, breast, prostate, and other cancers | 8% | 11% |
| HCC 18, diabetes with chronic complications | 29% | 27% |
| HCC 21, protein-calorie malnutrition | 16% | 13% |
| HCC 27, end-stage liver disease | 1% | 2% |
| HCC 28-29, cirrhosis of liver and chronic hepatitis | 2% | 3% |
| HCC 46, severe hematological disorders | 1% | 2% |
| HCC 48, coagulation defects and other specified hematological disorders | 13% | 16% |
| HCC 51, dementia with complications | 15% | 7% |
| HCC 52, dementia without complications | 34% | 20% |
| HCC 54-55, drug/alcohol psychosis and drug/alcohol dependence | 7% | 6% |
| HCC 57-58, schizophrenia, major depressive, bipolar and paranoid disorders | 20% | 14% |
| HCC 70-71, quadriplegia, paraplegia | 5% | 2% |
| HCC 72, spinal cord disorders/injuries | 2% | 3% |
| HCC 85, congestive heart failure | 48% | 45% |

| | | |
|---|---|---|
| HCC 96, specified heart arrhythmias | 35% | 38% |
| HCC 103-104, hemiplegia/hemiparesis, monoplegia, other paralytic syndromes | 12% | 8% |
| HCC 106, atherosclerosis of the extremities with ulceration or gangrene | 4% | 4% |
| HCC 107-108, vascular disease with or without complications | 46% | 42% |
| HCC 111, chronic obstructive pulmonary disease | 35% | 33% |
| HCC 134, dialysis status | 4% | 5% |
| HCC 136-138, chronic kidney disease, stage 3-5 | 8% | 8% |
| HCC 139-140, chronic kidney disease stage 1-2, unspecified renal failure | 7% | 5% |
| HCC 157-159, pressure ulcer of skin with necrosis or skin loss | 12% | 6% |

Source: Medicare administrative data, 2014-2015
*Beneficiaries who met the IAH qualifying criteria and lived in the same geographic areas where IAH beneficiaries resided but did not use HBPC. These data come from the large pool of beneficiaries identified using administrative data before we used matching techniques to select the final comparison group used for analyses of impacts of the demonstration. For more details, see Appendix C.
Note: Categories for race, original reason for entitlement, age, and number of chronic conditions may add to more to 100% because of rounding.
CCW=Chronic Condition Warehouse; HCC=hierarchical condition category. The CMS-HCC model uses HCC categories, consisting of groups of diagnoses, in determining risk-adjusted Medicare Advantage plan payments.

## Extent of disability

Results from a survey of IAH beneficiaries illustrate the extent of disability among IAH beneficiaries (Table 2). The beneficiary eligibility criteria required that beneficiaries must need human assistance for at least two functional dependencies—activities of daily living— and close to half (47.7 percent) of enrollees reported needing human assistance with four or more ADLs (data not shown). A majority needed help with bathing or showering (77.0 percent) and dressing (64.3 percent); a substantial minority (33.5 percent) needed help with eating. Nearly two in three received one or more types of assistance completing the survey, such as reading questions or writing answers (data not shown). Roughly one in five lived alone in a private residence (19.8 percent), but the majority lived in assisted living (26.2 percent) or with one or more family members (44.6 percent). Another indication of dependency is that two-thirds of caregivers said they were normally in attendance when the IAH primary care clinician came to visit.

**Table 2. IAH beneficiaries' functional status and living situation: Beneficiaries enrolled in IAH, 2013-2015**

| **Limitation in activities of daily living for which beneficiary needs human assistance** | |
|---|---|
| Bathing or showering | 77.0% |
| Dressing | 64.3 |
| Eating | 33.5 |
| Getting in or out of bed or chairs | 52.9 |
| Walking | 54.1 |
| Using the toilet | 46.4 |
| **Beneficiary's current living situation** | |
| Lives alone in a private residence | 19.8% |
| Lives with family member | 44.6 |
| Assisted living | 26.2 |
| Other setting with non-family members | 11.0 |
| **Caregiver present at IAH visits*** | |
| All or most visits | 66.3% |

Source: Surveys of beneficiaries enrolled by IAH practices between June 2012 and June 2014. Surveys conducted 2013-2015.
* Based on a companion survey of caregivers of the IAH beneficiary sample. N = 3,870 for beneficiaries; N = 2,519 for caregivers.

# Who Are the IAH Practices and How Did They Deliver Care?

To understand the IAH practices and how they delivered HBPC, we collected information from practice providers and staff annually through site visits during the first three years. We also analyzed practices' Medicare claims. To gain insight into how IAH beneficiaries experience their primary care at home, we conducted beneficiary surveys and claims analyses. Both provider and beneficiary information contribute to understanding the practices and their care delivery approaches.

This section of the report provides descriptive information on IAH practices, including structural and operational characteristics, and brief sketches of three practices selected for illustrative purposes. We also discuss data on non-hospital visit patterns that highlights the strong dominance of primary care visits associated with the IAH practices. The section concludes with information about beneficiary experiences based on survey data.

### Practices' structural characteristics

To address the legislation's cap of 10,000 beneficiaries, CMS selected 18 practices (of which three were consortia of multiple separate practices) into the demonstration in two waves in 2012. All practices specialized in HBPC and treated some beneficiaries outside of the demonstration who did not meet the IAH requirements. By intention, the participants chosen were diverse geographically, hailing from 14 states and the District of Columbia, and were varied in their size, structure, and organizational setting (Table 3). For descriptive purposes, we classified the IAH practices into three categories:

1) independent community practices;
2) academic or medical-system-affiliated practices; and
3) units of a multistate corporation, the Visiting Physicians Association (VPA).

The academic medical center practices were small (about 10 providers or less) and nonprofit, whereas the independent practices were usually larger (13 to 75 providers) and mostly for-profit (Table 4). VPA practices were also relatively large (10 to 23 providers) and all for-profit. Practices had varying complements of staff supporting the care team, but care coordination was a common job function in all categories. Staff dedicated to scheduling visits were common among independent and VPA practices. VPAs used medical assistants to pair with visiting medical doctors. Academic medical center practices almost invariably had social workers on staff, who typically help manage care transitions by coordinating with hospital staff, in addition to linking patients to community resources and addressing psychosocial issues.

## Table 3. Practices' structural characteristics as of 2017

| Site | Medicare beneficiary census, Yr. 1* | Affiliation | Ownership | Full-time providers making house calls | Part-time providers making house calls | Other staff involved in care team |
|---|---|---|---|---|---|---|
| **Independent practices (n=4)** | | | | | | |
| Austin, TX | 2,294 | Kindred Health Care | For-profit | 4 physicians, 9 NPs, 4 PAs | 2 physicians | 5 LPNs, 2 MAs serving as patient service coordinators, 2 intake coordinators, 1 office manager, 1 medical record personnel |
| Brooklyn, NY | 1,661 | None | For-profit | 10 physicians, 15 PAs, 9 NPs[a] | None[a] | Quality assurance nurse, patient liaison[a] |
| Durham, NC | 2,501 | None | For-profit | 33 physicians, 35 PAs, 7 NPs | None | 6 podiatrists 2 psychologists, 1 social worker, 130 additional office support staff, 40 of which are MAs serving in clinical service, management, and scheduling capacities. |
| Portland, OR | 734 | None | Nonprofit | 4 physicians, 3 NPs, 1 PAs | 1 physician, 1 PA, 3 NPs | 17 RNs, 4 LPNs, 7 social workers, care coordinators, care coordinator supervisor, DME specialist |
| **Academic medical centers (n=7)** | | | | | | |
| Boston, MA | 787 | Boston Medical Center | Nonprofit | None | 6 physicians | 5 nurses, 1 office manager, 3 ambulatory service representatives, 1 project coordinator |
| Cleveland, OH | 587 | Cleveland Clinic | Nonprofit | 7 physicians, 3 NPs | 1 PA | 3 RNs, 4 MAs, 1 nurse manager, 1 social worker, 3 schedulers, 1 pharmacist |
| North Shore, NY | 659 | Northwell Health | Nonprofit | 4 physicians, 2 NPs | 2 physicians | 6 nurses, 6 medical coordinators, 5 social workers, 1 clinical data analyst, 1 DME coordinator |
| Philadelphia, PA[b] | 524 | University of Pennsylvania | Nonprofit | 1 NP | 3 physicians, 1 NP | 1 social worker |

| Site | Medicare beneficiary census, Yr. 1* | Affiliation | Ownership | Full-time providers making house calls | Part-time providers making house calls | Other staff involved in care team |
|---|---|---|---|---|---|---|
| Richmond, VA[b] | 754 | Virginia Commonwealth University | Nonprofit | 2 physicians, 6 NPs | 2 physicians, 1 NP | 2 RNs, 1 consulting pharmacist, 3 social workers, 1 office manager, 3 patient access representatives |
| Washington, DC[b] | 630 | MedStar Health | Nonprofit | 6 physicians, 5 NPs | 1 NP | 1 RN, 1 LPN, 5 MA, 1 social worker, 1 outcomes analyst |
| Wilmington, DE | 960 | Christiana Care Health Systems | Nonprofit | 1 physicians, 3 NPs | 4 physicians, 1 PA, 1 NP | 1 phlebotomist, 4 RNs, 4 MAs, 3 social workers, 1 office manager |
| **VPAs (n=5)** | | | | | | |
| Dallas, TX | 3,643 | US Medical Management | For-profit | 17 clinicians[c] | None | 18 MAs, 2 clinical educators on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[d] |
| Flint, MI | 4,754 | US Medical Management | For-profit | 23 clinicians[c] | None | 24 MAs, 5 clinical educators on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[d] |
| Jacksonville, FL | 1,696 | US Medical Management | For-profit | 14 clinicians[c] | 2 clinicians | 10 MAs, 1 clinical educator on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[d] |
| Lansing, MI | 1,935 | US Medical Management | For-profit | 10 clinicians[c] | None | 11 MAs, 2 clinical educators on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[d] |
| Milwaukee, WI | 1,814 | US Medical Management | For-profit | 12 clinicians[c] | None | 11 MAs, 1 clinical educator on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[d] |

Source: Information from practice representatives and staff originally collected in 2013 and updated with information from interviews held in 2015 and, for all practices except Louisville, in 2017.

*Information on number of beneficiaries seen comes from analysis of all paid claims for all Medicare beneficiaries in Year 1.

[a] The Brooklyn, NY, site did not provide updated information on the number of full- and part-time providers making house calls, or other staff involved in the care team.

[b] As members of a consortium formed for purposes of the demonstration, these three sites (Philadelphia, Richmond, and Washington, DC) are considered one practice.

[c] VPAs reported on the total number of clinicians and did not provide information on the number of physicians, NPs, and PAs.

[d] Additional care team staff are located at the corporate office in Troy, MI, and provide support to local sites: 1 social worker, 1 DME intake, 1 care manager.

DME = durable medical equipment; MA = medical assistant; NP = nurse practitioner; PA = physician assistant; LPN = licensed practical nurse; RN = registered nurse.

## Practices' operating characteristics

Additional operational information collected from IAH practices is summarized in Appendix D. There we discuss a number of distinguishing features of each category of practices. Summary information is provided in tables. Appendix Table D1 provides the approximate patient panel size of a typical IAH clinician and several visit characteristics, such as after-hours visit policies. Panel sizes are notably smaller than those for an office-based physician working in a team.[15] Many practices provide visits that are not covered by Medicare, such as visits by social workers or other paraprofessionals. Appendix Table D2 covers a variety of information, such as how the practice receives notification of inpatient admissions and whether routine patient monitoring includes proactive outreach to beneficiaries.

## Visits to IAH beneficiaries by location and practitioner type

HBPC takes place in private homes or group dwellings such as board and care homes and assisted living facilities (ALF)[16]. National data indicate that a slight majority of home visits take place in group dwellings such as ALFs and board and care homes. However, among IAH beneficiaries, private homes predominate, with the exception of two independent practices (Durham and Portland) and one VPA (Milwaukee), sites where assisted living was the most common venue for patient visits by far (Table 4; Year 2 shown for illustrative purposes). In fact, these three practices—two independent practices and one VPA practice—delivered 70 percent or more of their home visits to IAH patients in ALFs. In contrast, most of the academic medical center practices rarely or never saw patients in ALFs. Several academic medical center practices made visits to their hospitalized IAH patients but this was not true among any of the other types of practices.

Most practices shared patient visiting responsibilities between physicians and nurse practitioners (NPs), but in the academic medical centers, physicians and NPs had a tendency to share visiting more evenly than did the other groups. VPAs follow a physician-oriented model, with the vast majority of visits delivered by physicians. At the other extreme, at the Portland site NPs delivered about 7 of every 8 visits. Physician assistants (PAs) delivered about one-quarter of visits in two independent sites (Austin and Durham).

## Illustrative provider sketches

Data tables alone don't portray the wide variation among IAH practices in their approach to carrying out HBPC. We illustrate the variation with three sketches, one from each IAH practice category, following Table 4 (Figures 3-5).[17]

---

[15] Module 20. Facilitating Panel Management. Content last reviewed May 2013. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/professionals/prevention-chronic-care/improve/system/pfhandbook/mod20.html

[16] In this report, we use the term "assisted living facilities" to denote all types of domiciliary care residences.

[17] Practices used for illustration are not intended to be representative of an entire category, and their selection does not imply any particular status in the demonstration. Most of the information was collected during a site visit in the year shown.

## Table 4. IAH practice visits by provider type and location, Year 2

| | Percentage of visits by provider type[a] | | | | Percentage of visits by location[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Practice** | **Physician** | **NP** | **PA** | **Other** | **Home** | **ALF** | **Inpatient Hospital** | **Office/ Outpatient Clinic** | **Other** |
| **Independent practices (n=4)** | | | | | | | | | |
| Austin, TX | 21.3 | 47.7 | 25.3 | 5.8 | 84.4 | 15.6 | 0 | 0 | 0 |
| Brooklyn, NY | 94.1 | 5.9 | 0 | 0 | 100.0 | 0 | 0 | 0 | 0 |
| Durham, NC | 64.0 | 0 | 28.0 | 8.0 | 10.7 | 84.4 | 0 | 5.0 | 0 |
| Portland, OR | 5.8 | 87.1 | 7.0 | 0 | 16.6 | 83.4 | 0 | 0 | 0 |
| **Academic medical centers (n=7)** | | | | | | | | | |
| Boston, MA | 100.0 | 0 | 0 | 0 | 77.7 | 4.1 | 14.3 | 3.3 | 0.7 |
| Cleveland, OH | 66.4 | 30.6 | 3.0 | 0 | 84.1 | 13.7 | 0 | 2.3 | 0 |
| North Shore, NY | 80.7 | 19.3 | 0 | 0 | 100.0 | 0 | 0 | 0 | 0 |
| Philadelphia, PA* | 38.3 | 61.7 | 0 | 0 | 92.8 | 0 | 1.5 | 0.4 | 5.3 |
| Richmond, VA* | 28.7 | 71.3 | 0 | 0 | 93.6 | 0 | 0.3 | 1.9 | 4.3 |
| Washington, DC* | 31.4 | 68.6 | 0 | 0 | 87.9 | 0.5 | 10.8 | 0.7 | 0.1 |
| Wilmington, DE | 40.1 | 43.6 | 16.3 | 0 | 93.3 | 1.6 | 2.4 | 1.9 | 0.7 |
| **VPA (n=5)** | | | | | | | | | |
| Dallas, TX | 69.5 | 30.5 | 0 | 0 | 88.3 | 11.7 | 0 | 0 | 0 |
| Flint, MI | 98.0 | 1.8 | 0 | 0.2 | 62.5 | 37.3 | 0 | 0.1 | 0.1 |
| Jacksonville, FL | 92.2 | 7.4 | 0 | 0.4 | 69.0 | 30.9 | 0 | 0 | 0 |
| Lansing, MI | 91.1 | 8.9 | 0 | 0 | 62.3 | 37.7 | 0 | 0 | 0 |
| Milwaukee, WI | 87.8 | 12.2 | 0 | 0 | 28.5 | 71.5 | 0 | 0 | 0 |

Source: Medicare claims for 2013-2014, for IAH enrollees in each IAH practice. Note: ALF=assisted living facility.
* Philadelphia, Richmond, and Washington, DC, participated as members of a single consortium.
[a] Percentage of the total number of evaluation and management visits billed in CMS professional claims files.
[b] All figures are percentages of the total number of evaluation and management visits provided by IAH practices.

## Figure 3: Provider Profile, Washington, D.C. (academic medical center) (2015)

Housed in the medical center's geriatrics department, the practice serves patients from several zip codes in the surrounding area. Physicians and NPs are grouped, with each group dedicated to its own list of patients. In weekly meetings of an interdisciplinary team (e.g., social worker, home health nurse), discussion focuses on "unstable" patients; any staff person can identify a patient for discussion. Most IAH patients are dually eligible for Medicare and Medicaid. Social workers are partly supported by a contract to provide case management under a Medicaid waiver for home and community-based services. Another source of support for the practice is subsidy from the medical center.

Busy office coordinators develop personal relationships with patients, as they answer phone calls; arrange for physician order fulfillment, referrals, and transportation; prepare charts for clinicians' use in upcoming visits; and confirm appointments. An office-based nurse triages the many incoming clinically focused calls. The practice hired a registered nurse for weekend coverage to ensure patient contacts and home visits occur as necessary. The practice strives to follow patients across all settings—including into the medical center's hospital, if necessary—supported by a mobile electronic health system (EHR) for timely communication among staff, which they say makes it possible to coordinate care during acute episodes, including for treatment and discharge planning.

## Figure 4: Provider Profile, Visiting Physicians Association practices (2016)

VPA, with more than 40 offices across 12 states, is an affiliate of U.S. Medical Management (USMM), whose integrated home delivery medical model includes physician house calls, home health care, and hospice as well as radiology, diagnostic, laboratory, and pharmacy/medical supplies. VPA corporate infrastructure includes an education center, call center, and various administrative support departments (e.g., information technology, data analytics, and finance). Home-based primary care is delivered by mobile teams consisting of a physician and a medical assistant; on average visits occur at least every four weeks. The medical assistant routes the visits and drives the physician to see the patient. In the home, the medical assistant takes patient vitals, blood samples and gives injections, and carries a tablet that stores reminders and other information to identify systematically patients' needs and risk factors. Teams are backed by office-based care coordinators who follow up on physician orders and serve as the routine point of contact for patients. Schedulers and a clinical care educator round out the office staff.

VPA relies on its central office for some aspects of care coordination. Upon learning of a hospital or ER admission, typically via a state health information exchange, the central office nurse care manager contacts discharge planners and others to coordinate care, while communicating with the patient's local clinician and care coordinator to plan for the return home and follow-up visits. The central office also provides on-call physician coverage after hours, supported by a company-wide EHR.

.

## Figure 5: Provider Profile, Brooklyn, NY (independent practice) (2016)

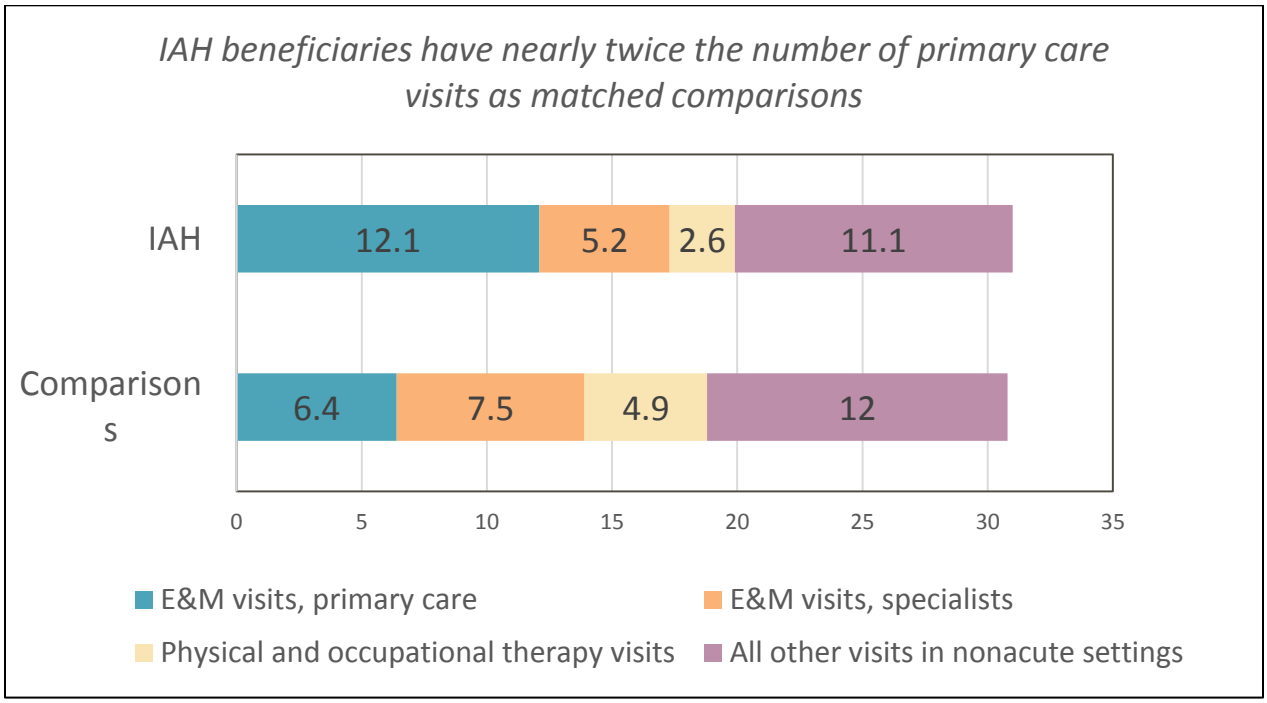
The practice is overseen by its medical director, who is the owner, and an associate medical director. They are assisted by a vice president of strategic planning, a senior operations director, and a director of communication and logistics. The practice employs 37 drivers to ferry visiting clinicians to patient appointments around the five boroughs of New York City, often in congested vehicle traffic. Mapping software helps drivers optimize their routing each day. Two clinicians are assigned to each patient, allowing for backup of the one designated as lead. One section of the busy back office houses call center operators who take calls from patients; in many cases operators came to the practice with valuable customer service experience in outside industries. Another section houses patient liaisons, who call each patient weekly and help coordinate patient care.

In March 2016, the practice began using a service that sends real-time notification of patient ED registrations, ED discharges, and acute hospital admissions. For example, upon receiving notice of an ED registration, the practice protocol calls for the associate medical director to contact the ED clinician to discuss the case and plan for discharge or hospital admission. If discharged, the patient is called by a practice patient liaison to schedule a next-day appointment. Because only about one-third of area hospitals participate in the notification service, the practice is not always made aware of patient events by the system (or by other means such as direct data feeds sent by specific hospitals).

### Primary care visits and other non-hospital visits

A noteworthy feature of IAH care delivery is the dominance of primary care visits in the visit mix. Figure 6 shows the number of non-hospital visits per beneficiary per year (PBPY) for evaluation and management (E&M) services or physical and occupational therapy visits. Non-hospital visits reflect liberal use of primary care compared to the mix for the evaluation's "usual care" matched comparison group. (As detailed in Appendix C, the matched comparison group members resided in the same geographic areas as the IAH beneficiaries, and had demographic and health characteristics comparable to IAH beneficiaries', but did not receive their primary care at home.) Although the total of non-hospital visits for IAH beneficiaries approximately equals the total for their comparisons, IAH beneficiaries have an annual average of 12 visits from primary care providers, nearly twice as many as matched comparisons. In the IAH group, smaller numbers of non-hospital visits are from specialists and therapists, compared to the matched comparison group.

**Figure 6: Yearly average number of visits outside of the hospital, according to type of visit: IAH beneficiaries and matched comparisons, Year 3**




Source: Medicare claims 2010-2015. Data come from regression analysis of two pre-demonstration years and three demonstration years, using a total of 37,250 treatment-group observations and 158,828 comparison-group observations. In Year 3, the numbers of treatment and comparison beneficiaries were 7,564 and 31,259, respectively.
Note: Settings for visits in Figure 6 are limited to home, office, hospital outpatient clinic, rural health clinic, federally qualified health centers; Part B therapy delivered by skilled nursing facilities are also included

The large number of primary care visits among IAH beneficiaries likely reflects, at least in part, scheduling of regular visits in the demonstration for the purpose of monitoring patients, but we also found similarly large numbers of primary care visits among the practices' patients who did not qualify for IAH. Factors such as patient preferences or a different approach to primary care practice on the part of IAH providers may also be contributing to the IAH pattern. Another part

of the explanation may be found in the dominance of office-based practice among specialists and therapists. IAH patients often experience physical, cognitive, and other barriers in accessing care outside of the home, and the result can be less access to specialists and therapists.

## Beneficiary reports of physician care quality, access, and patient satisfaction

During the first three years of the demonstration, we surveyed beneficiaries about physician care quality, access to care, and patient satisfaction. We also surveyed the same beneficiaries' caregivers. Beneficiary responses about physician care quality were compared with answers from respondents to the Medicare Current Beneficiary Survey (MCBS) who met the IAH beneficiary eligibility criteria.[18] We found a relatively small number of MCBS respondents for this comparison (N=337), and had to make some adjustments to the IAH-qualifying criteria when selecting MCBS respondents; therefore, the results from comparisons should be considered tentative.[19]

Viewpoints revealed in the seven questions on the topic of physician care quality were somewhat less favorable among IAH beneficiaries than among the MCBS respondents, who were asked about their usual source of care. Specifically, IAH respondents more often agreed or strongly agreed with the view that the primary care team often seems to be in a hurry (18.8 percent vs. 13.0 percent); that they often have health problems that should be discussed but are not (21.5 percent vs. 13.2 percent); and that the primary care team often acts as though it were doing the beneficiary a favor by talking to him or her (14.9 percent vs. 9.7 percent) (all differences statistically significant at 0.05 or below). Three additional questions—about confidence in the primary care team, whether the team answers all of the patient's questions, and whether the team has a good understanding of the patient's medical history—all showed high ratings for IAH practices (exceeding 90 percent) but not quite as high as ratings among the

[18] For comparison, we used the 2013 Medicare Current Beneficiary Survey files to select IAH-eligible beneficiaries who responded to the MCBS and said they did not receive home visits. The questions related to physician quality of care were not administered to MCBS beneficiaries residing in domiciliary facilities, so for those questions we only present responses from IAH beneficiaries in a similar living situation. Questions related to access to care and patient satisfaction were not analyzed in comparison to the MCBS respondents we selected from the MCBS survey files.

[19] We interpret the differences conservatively as an exploratory analysis for several reasons. We made some adjustments to the eligibility criteria used for the MCBS sample in an attempt to make the sample consistent with the circumstances of the IAH respondents. For example, respondents selected from the MCBS could have had a hospitalization or rehabilitation stay within two years before the survey, because the IAH respondents received the survey as long as two years after experiencing their IAH-qualifying utilization events. Although we replicated the IAH eligibility criteria as best we could for the MCBS benchmark group, and used regression analysis to adjust for differences between the groups, differences in unobserved characteristics may remain. Observable differences between the groups in age, race/ethnicity, living arrangement, prevalence of dementia, and functional status in particular suggest differences may exist in unobserved variables. In addition, IAH beneficiaries are drawn from a small number of states, whereas the MCBS benchmark group is designed to be representative of the national sample of beneficiaries who meet IAH eligibility criteria.

MCBS respondents (differences were about 5 percentage points and statistically significant). On the question of whether "the primary care team is very careful to check everything when examining you," a small difference favored IAH practices and was not statistically significant.

Responses from IAH beneficiaries suggest access to primary care was good. Access-related questions in our survey were specifically about the IAH team and had no counterparts in the MCBS. The vast majority of IAH respondents (87. 4 percent) reported no trouble obtaining the in-home care they needed from the IAH practice. Among the one in eight respondents who said they experienced trouble getting needed in-home care, the most common problem concerned availability of the primary care team—either the wait was too long or the team too busy (38.6 percent)—followed by trouble contacting the office to make appointments (33.0 percent).[20] When asked about the frequency with which the beneficiary would like the IAH clinicians to visit, approximately three-quarters said the visit frequency was about right, and nearly one in five said they would like visits more often. When it came to helping beneficiaries arrange for medical care outside the home from non-IAH providers, more than half of beneficiaries reported the IAH practice gave them a significant amount of help (36.1 percent) or some help (19.0 percent). However, 13.4 percent said their IAH practice provided no help with making those arrangements. Perspective on these statistics, via comparison with samples not using home visits, is unavailable. Other evidence on care coordination suggests that some practices took steps to improve care-partner relationships and involvement, with potential salutary effects on care coordination.

Overall satisfaction with the primary care team seems similar to that which we find nationally[21] for Medicare beneficiaries. Overall satisfaction questions in the IAH survey asked about satisfaction with the primary care team, whereas MCBS asks about satisfaction with the beneficiary's health care from doctors and hospitals generally. About 93 percent of beneficiaries and caregivers reported that they were very satisfied or satisfied with the overall quality of care they had received from the IAH practice in the past six months. In national MCBS data for 2013, 52.1 percent of beneficiaries were "very satisfied"[22]; similarly, 50.2 percent of IAH beneficiaries were "very satisfied" with care overall from the IAH practice. When asked specifically how much he or she likes receiving in-home care compared to primary care in an office or clinic, 72.5 percent of beneficiaries said "a lot more," 10.5 percent said "somewhat more," and 12.4 percent said "about the same." Caregivers' opinions about receiving in-home care mirrored those of beneficiaries. These indicators suggest patient satisfaction is strong, notwithstanding IAH respondents' relatively unfavorable views on some aspects of physician care discussed earlier.

[20] Beneficiaries could select more than one response to this question.

[21] MCBS data cited here come from CMS' published national tables rather than the analysis of care quality discussed earlier in this section; the care quality analysis was based on specific IAH survey questions using wording in the MCBS.

[22] Table 5.2, accessed in March 2017 at https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables-Items/2013CNP.html?DLPage=1&DLEntries=10&DLSort=0&DLSortDir=descending

# What Was the Demonstration's Impact and How Did the IAH Practices Change during the Intervention Period?

This section presents the impacts of the demonstration in two parts: the impact on beneficiaries and the practice changes reported to us during the intervention period. First, we provide an impact analysis in terms of quantitative outcomes for beneficiaries, inclusive of Medicare expenditures, key utilization indicators, and health-related outcomes.[23] Second, we report the qualitative information we collected from the practices, which was focused on the organizational and operational changes they undertook during the demonstration's first three years. Their reports of how they changed and what they did specifically in response to the demonstration may be indicative of the impact of the demonstration on the sites' approach to care. The qualitative findings are followed by a summary of the individual practices' achievement of the demonstration's quality measures, and a brief discussion of our preliminary assessment of best practices. While not an impact analysis, the best practices assessment is an exploration of the association between how practices operated and their financial and quality performance under the demonstration.

### Quantitative impacts: expenditures, utilization, and health status indicators

The quantitative analysis of the demonstration's impact takes into account the pre-existing difference between IAH beneficiaries on average and matched comparisons on average at baseline, i.e., in the year before the demonstration began. In other words, the impact of the demonstration for a given demonstration year is the difference in the given year minus the pre-existing baseline difference (see Appendix C for further explanation). This is known as a "difference-in-differences" analysis, which in effect measures change at the IAH practices remaining after netting out change that would have occurred absent the demonstration. In other words, we are comparing an extrapolation of the difference observed in the baseline year between the treatment group and the comparison group to the actual differences observed during the treatment period, and departures from the extrapolated difference indicate impacts due to the demonstration (see Figure C1 for a depiction of the strategy). For substantiating the assumption that net change is indicative of impacts, we also check to see that differences between the two groups within a multiyear baseline period remain statistically the same, i.e., trends are parallel. For most quantitative measures in the evaluation, we examined the two years leading up to the demonstration, and we found evidence of parallelism.

In Tables 5 through 9, we present the amount of net change attributable to the demonstration as an annual value for each year and as an average value for the entire three-year intervention period. For example, the change attributable to the demonstration, or impact, for the annual number of hospitalizations was a reduction of 0.08 hospitalizations PBPY in Year 3 (see Table 6). The change for the entire intervention period on average was a reduction of 0.05 hospitalizations PBPY.

[23] Sample sizes for the quantitative data except the survey are in Appendix C, Table C1.

For representing the entire intervention-period impact in percentage terms, we divided the estimated impact for the entire intervention period by the average value for the treatment group in the baseline year.[24] The baseline value used for the denominator is also provided in the tables. To continue with the previous example, the baseline number of hospitalizations PBPY was 1.78 (Table 6). Therefore, for the entire intervention period, the approximate percentage change in the number of hospitalizations PBPY was -0.05 ÷ 1.78, or -2.8 percent. In this case, and in the tables that follow, values are shown after rounding.

With the exception of the survey data, quantitative impacts in this section are based on analyzing the five successive annual samples given in Appendix C, Table C1. Sample members had to meet the same IAH qualifying criteria each year, and therefore the samples exclude beneficiaries who did not meet the IAH criteria in years after their first year.[25] Results were robust to undue influence from outliers in the data[26] and to the potential influence of Medicare shared savings programs in which some comparison beneficiaries took part.[27]

The expenditure analysis in this report does not include Medicaid expenditures for dually eligible beneficiaries; nearly four in ten IAH beneficiaries in our samples are dually eligible for Medicare and Medicaid. The Medicaid expenditure data will be analyzed later in the evaluation if all needed data become available. As noted earlier in this report, estimates of expense reductions are independent of incentive payment results released by CMS for Years 1 and 2 of the demonstration.

**Part A and B expenditures analysis.** All estimates of the change in total expenditures due to the demonstration indicate reductions, which range from $31 (less than one percent) in Year 2 to $177 (nearly four percent) in Year 3, for an average reduction of approximately 2.5 percent over the course of the demonstration. However, no estimates were statistically significant, which means our confidence that the demonstration resulted in real savings is low. The all-practice analysis showed $123, $31, and $177 saved per beneficiary per month (PBPM) in total Part A and Part B expenditures in the first, second, and third years, respectively (Table 5). For the three intervention years overall, the average reduction was $111 PBPM, or approximately 2.5 percent, relative to the treatment group's mean PBPM expenditures in the year before the demonstration.

[24] Each yearly impact estimate can be estimated as a percent change by using the year's impact estimate in the numerator and the average value in the year before the demonstration for the treatment group in the denominator.

[25] The demonstration rules allowed a beneficiary to continue to participate in each successive year without continuing to meet the IAH qualifying criteria that were responsible for her admission initially (see Figure 1 for requirements). Unlike the participant rules in the demonstration, the evaluation sample members had to meet the requirements each year. We adhered to this sample selection rule to ensure that the analysis would not erroneously attribute impacts to the incentive structure that were actually due to changes in sample case mix. For more details on sample selection, see Appendix C.

[26] The following key outcomes were tested for sensitivity to outliers: total expenses, hospital admissions, ED visits leading to hospital admission.

[27] We tested the total expenses outcome for robustness to participation on other shared savings programs.

The 10,000-beneficiary cap in the law operated to restrict the number of practices that could be selected for the demonstration, which in turn limited the effective sample size available for statistical analysis of all variables in the impact analysis. Given the design of the demonstration and its resulting small number of practices, a change nearly three times as large as the three-year average we measured ($111) would have been statistically significant. Notably, however, the largest estimate was for the third year. Based on qualitative findings discussed later in this report, the sites' performance in that year could reflect an accumulation of experience with practice change, in response to the demonstration. If that is the case, then we might expect to see at least as large an impact after we analyze data from remaining years of the demonstration, possibly with a concomitant improvement in statistical confidence in the estimates.

Demonstration impact estimates for the typically most costly component of expenditures, inpatient care, were not statistically significant. However, as with total expenditures, inpatient care estimates consistently indicated savings (-$59, -$42, and -$81 in Years 1 through 3, respectively) and suggested notable improvement by Year 3. The non-significant changes for professional services, another relatively large component of expenditures, also consistently suggested reductions and appeared to trend towards improvement (-$18, -$20, -$26).
Most other categories did not show an impact of the demonstration. A clear exception was a reduction in expenditures for durable medical equipment (DME). The savings on DME, $10, $22 and $35, in Years 1, 2, and 3, respectively, amounted to double-digit savings rates of about 15 percent during the demonstration's first three years overall. Savings in the DME category were the only changes that were statistically significant.

### Table 5. IAH impact on expenditures by service category: Years 1 to 3

| | Treatment group mean, PBPM | Demonstration Impact Estimates† | | | Entire Intervention Period Estimate†† | Approximate Percentage Impact††† |
|---|---|---|---|---|---|---|
| **Expenditures PBPM** | **Baseline Year** | **Year 1** | **Year 2** | **Year 3** | **Years 1- 3** | **Years 1-3** |
| Total Part A and Part B $ | $4,397 | -$123 | -$31 | -$177 | -$111 | -2.5% |
| — Inpatient hospital | $1,741 | -$58 | -$41 | -$78 | -$59 | -3.4% |
| — Skilled Nursing Facilities | $605 | -$15 | $12 | $0 | -$2 | -0.3% |
| — Hospice | $153 | 0 | $10 | $4 | $4 | 2.9% |
| — Home Health | $781 | -$8 | $30 | -$32 | -$4 | -0.5% |
| — Professional Claims Expenditures | $715 | -$18 | -$20 | -$26 | -$21 | -3.0% |
| — Outpatient Claims Expenditures[a] | $253 | -$13* | 0 | -$10 | -$8 | -3.2% |
| — Durable Medical Equipment | $150 | -$10** | -$22*** | -$35*** | -$22*** | -14.6% |

Source: Medicare claims 2010-2015 (baseline and demonstration periods). Expenditures were risk-adjusted and weighted to reflect partial year observations. We did not adjust for Medicare price changes as the regression method we used accounts for them.
†Demonstration impact estimate is the amount of change due to the demonstration after accounting for the pre-existing difference between the treatment and comparison groups. See text and Appendix C for further explanation of the methods.
††Entire intervention period estimate is the average change in the measure (e.g., PBPY) for the 3 years combined into a single period.
†††The approximate percentage impact is the entire intervention period impact divided by the average value in the treatment group in the baseline year, which is the year before the demonstration.
[a]Expenditures from institutions such as hospital outpatient departments and rehabilitation agencies.
*, **, *** indicate statistical significance at .10, .05, and .01 levels, respectively. PBPM=per beneficiary per month.

**Acute hospital and ED utilization measures.** As with most of the expenditures impacts, the demonstration's effect on overall hospital admissions suggested a modest, consistent reduction that was not statistically significant, but some subcategories of hospital use showed more reliable signs of improved performance. Total hospital admissions per person, which averaged nearly two per beneficiary per year (PBPY) in the baseline year, saw reductions ranging from 0.03 to 0.08, with none statistically significant (Table 6). Again, the largest estimate was for the third year of the demonstration. The percentage change for the entire intervention period was only -2.9 percent. However, two important subcategories of hospital admissions improved somewhat more, particularly in Year 3. By the third year, both the number of preventable[28] admissions PBPY and the percentage of beneficiaries experiencing an unplanned readmission declined by a statistically significant amount; preventable admissions PBPY declined by 0.03 admissions per person, and the percentage of beneficiaries experiencing at least one unplanned readmission fell by 1.71 percentage points. While the sequence of yearly estimates suggests improvement over time in these subcategories of hospital admissions, neither set of estimates was consistently large enough to produce a statistically significant effect for the three years on average; in both cases, the percentage change was under 5 percent.

Emergency department (ED) use improved the most, when considering the ED visits that result in admission. Total ED visits PBPY—including both those that led to admission and those that did not--decreased by an average of 0.09 visits during the intervention period, given reductions of 0.12 and 0.15 in the first and third years of the demonstration (both statistically significant). However, when total ED visits were broken down according to the disposition of the patient after the visit, the reduction was attributable to ED visits that resulted in hospitalization. This category of ED use saw statistically significant reductions of 0.07 and 0.12 in the first and third years, and a statistically significant intervention-period decrease of approximately 0.08 visits per person, or 5.8%. ED visits that did *not* lead to admission were not affected by the demonstration; nor were preventable ED visits affected. Reasons for the different results for ED visits leading to inpatient admission versus those that did not are unclear. The result may be partly attributable to practices' occasional ability to intervene while an ED visit was in progress and to work with the ED to discharge the patient to home rather than to the hospital inpatient

[28] Preventable utilization events such as hospitalizations and ED visits were identified using an Agency for Healthcare Research and Quality measure, Prevention Quality Indicators (PQI). PQIs comprise a variety of potentially ambulatory-care-sensitive conditions, such as diabetes and asthma.

service. Moreover, the result is consistent with practices' reports that they focused more resources on patients they assessed as having high health risks; by Year 2 use of risk stratification to allocate clinical resources was widespread among the IAH practices (discussed later in this report). Those same risk indicators may be associated with a high likelihood of being hospitalized when the patient presents to the ED. The extra resources directed to such IAH patients could have paid off by forestalling some trips to the ED.

### Table 6. IAH impact on hospital admissions and ED visits: Years 1 to 3

| | Treatment Group Mean | Demonstration Impact Estimates† | | | Entire Intervention Period Estimate†† | Approximate Percentage Impact††† |
|---|---|---|---|---|---|---|
| **Outcome measure** | **Baseline Year** | **Year 1** | **Year 2** | **Year 3** | **Years 1-3** | **Years 1-3** |
| Number of hospital admissions PBPY | 1.78 | -0.05 | -0.03 | -0.08 | -0.05 | -2.9% |
| Number of preventable hospital admissions PBPY | 0.46 | -0.01 | -0.01 | -0.03* | -0.02 | -3.8% |
| Had at least one unplanned readmission in the year (%) | 19.55% | 0.02% | -0.96% | -1.71%** | -0.84% | -4.3% |
| Total number of ED visits PBPY | 2.90 | -0.12* | -0.01 | -0.15* | -0.09* | -3.2% |
| — Number of ED visits resulting in hospital admission PBPY | 1.44 | -0.07** | -0.06 | -0.12** | -0.08** | -5.8% |
| — Number of ED visits without hospital admission PBPY | 1.46 | -0.03 | 0.07 | -0.02 | 0.00 | 0.3% |
| Number of preventable ED visits without hospital admission PBPY | 0.19 | 0.00 | 0.02 | 0.00 | 0.01 | 4.0% |

Source: Medicare claims, 2010-2015 (baseline and demonstration periods). Utilization measures were risk-adjusted and weighted to reflect partial year observations.
†Demonstration impact estimate is the amount of change due to the demonstration after accounting for the pre-existing difference between the treatment and comparison groups. See text and Appendix C for further explanation of the methods.
††Entire intervention period estimate is the average change in the measure (e.g., PBPY) for the 3 years combined into a single period.
†††The approximate percentage impact is the entire intervention period impact divided by the average value in the treatment group in the baseline year, which is the year before the demonstration.
*, **, *** indicate statistical significance at .10, .05, and .01 levels, respectively.
PBPY=per beneficiary per year; ED=emergency department.

**Key expense and utilization outcomes for independent practices.** Key acute-care utilization outcomes for one subgroup of practices, independent practices, were markedly stronger than results overall. This was revealed when we analyzed five key outcomes separately for each type of practice: independent practices, academic medical center practices, and VPAs. As shown in Table 7, the outcomes were: (1) total Medicare expenditures PBPM, (2) number of hospital admissions PBPY, (3) number of ED visits resulting in hospital admission PBPY, (4) number of ED

visits without hospital admission PBPY, and (5) whether the beneficiary had at least one unplanned readmission. For brevity, Table 7 reports results for independent practices only; estimates for the other two groups tended to be small and usually were not statistically significant.

Independent practices saved a statistically significant $415 PBPM in Year 1, and savings grew in the next two years, reaching $626 PBPM in Year 3. The Year 3 savings percentage was 13.7 percent. We found similarly large, or larger, percentage reductions for three other utilization measures; for example, the rate of any unplanned admissions declined by more than 15 percent in Year 1, and more than one-third in Year 3. ED visits that did not lead to hospital admission changed the least among these measures, but the Year 3 result, a decline of 10.0 percent, was still notable.

These results indicate that the independent practices were the main driver behind the overall results, but the reasons for the exceptional performance of this group are not clear. Our observation of the trends in annual mean outcomes suggested that, unlike the other IAH practice categories, outcome measures in the independent practices' comparisons did not generally trend downward during the intervention period (data not shown). Rather, comparison group measures for the independent practices had a tendency to be flat. Given that the independent practices' outcome measures did decrease, the net reductions, or impacts, were relatively large. The trends in impacts suggest that the independent practices as a group built upon their improvements in each successive year, with the exception of the ED visits that did not lead to hospital admission.

### Table 7. IAH impact on expenditures and selected acute care measures, independent practices subgroup: Years 1 to 3

| | Demonstration Impact Estimate† | | | Range of Approximate Percentage Impacts†† |
|---|---|---|---|---|
| **Outcome measure** | **Year 1** | **Year 2** | **Year 3** | **Range: Years 1 to 3** |
| Total Part A and Part B $ PBPM | -$415** | -$438** | -$626** | -9.1% to -13.7% |
| Number of hospital admissions PBPY | -0.26*** | -0.25*** | -0.32*** | -14.9% to -19.2% |
| Had at least one unplanned readmission in the year (%) | -2.76%** | -4.49%*** | -6.01%*** | -15.4% to -33.6% |
| Number of ED visits resulting in hospital admission PBPY | -0.22*** | -0.23*** | -0.35*** | -16.2% to -26.2% |
| Number of ED visits without hospital admission PBPY | -0.15*** | -0.02 | -0.14* | -1.2% to -10.0% |

Source: Medicare claims, 2010-2015 (baseline and demonstration periods). Expense and utilization measures were risk-adjusted and weighted to reflect partial year observations. We did not adjust for Medicare price changes as the regression method we used accounts for them. Results are based on 63,028 unweighted observations, including baseline years.

†Demonstration impact estimate is the amount of change due to the demonstration after accounting for the pre-existing difference between the treatment and comparison groups. See text and Appendix C for further explanation of the methods.
††The approximate percentage impact is the annual impact divided by the average value in the treatment group in the baseline year, which is the year before the demonstration.
*, **, *** indicate statistical significance at .10, .05, and .01 levels, respectively.
PBPY=per beneficiary per year; ED=emergency department.

**Other organizations providing care.** Our examination of institutional care providers, comprising skilled nursing facilities (SNF), inpatient rehabilitation hospitals (IRF), hospices, and home health agencies (HHA), showed the demonstration had no statistically significant impact on the percentage of beneficiaries using these services (Table 8). During the entire intervention period, SNF and IRF use registered increases, and hospice use decreased, with annual changes generally small and not consistently either positive or negative. Home health agency utilization, which is very common among IAH beneficiaries—about 90% use it during a year—showed somewhat mixed results. With an average reduction of 0.5 percentage points PBPY, the HHA use rate did not change appreciably, but total visits from HHAs rose by about five per person in the second year. Days of stay increased (data not shown), and total HHA expenses per beneficiary per month also increased in that year (Table 5), though neither estimate was statistically significant. The possible uptick in intensity of HHA services in Year 2 could be a random fluctuation but it is generally consistent with Year 2 results in various other services suggesting that utilization and cost performance was weaker in Year 2 than in other years. For example, Year 2 savings estimates for hospital inpatient, SNF, hospice and outpatient services were the smallest of any year, notwithstanding that all of those estimates were not statistically significant.

## Table 8. IAH impact on institutional rehabilitation, home health, and hospice use: Years 1 to 3

| | Treatment Group Mean | Demonstration Impact Estimate† | | | Entire Intervention Period Estimate†† | Approximate Percentage Impact††† |
|---|---|---|---|---|---|---|
| **Outcome measure** | **Baseline Year** | **Year 1** | **Year 2** | **Year 3** | **Years 1-3** | **Years 1-3** |
| Used SNF in the year (%) | 41.01% | -0.52% | 1.29% | 0.26% | 0.31% | 0.8% |
| Used IRF in the year (%) | 4.82% | 0.35% | 0.44% | 0.28% | 0.35% | 7.3% |
| Used hospice in the year (%) | 17.86% | -0.56% | 0.02% | -0.47% | -0.34% | -1.9% |
| Used HHA in the year (%) | 91.26% | -0.44% | -0.25% | -0.83% | -0.5% | -0.6% |
| Number of HHA visits PBPY | 62.32 | -1.64 | 5.07** | -1.21 | 0.60 | 1.0% |

Source: Medicare claims, 2010-2015 (baseline and demonstration periods). Utilization measures were risk-adjusted and weighted to reflect partial year observations.
†Demonstration impact estimate is the amount of change due to the demonstration after accounting for the pre-existing difference between the treatment and comparison groups. See text and Appendix C for further explanation of the methods.

††Entire intervention period estimate is the average change in the measure (e.g., PBPY) for the 3 years combined into a single period.
†††The approximate percentage impact is the entire intervention period impact divided by the average value in the treatment group in the baseline year, which is the year before the demonstration.
*, **, *** indicate statistical significance at .10, .05, and .01 levels, respectively.
PBPY=per beneficiary per year

**Professional visits.** To the extent that IAH practices were expected to make more home visits to prevent unnecessary hospital use and to meet IAH patient contact targets, the demonstration was expected to boost the use of primary care, relative to what would have happened in the absence of IAH. (Later in this report, we describe operational changes reported by IAH practices, changes directed at both augmenting patient monitoring and meeting IAH quality targets.) At the same time, it was expected that, if at-home primary care became more effective in IAH practices, specialist visits might see a decrease relative to what might have happened without the IAH demonstration. We found limited indications that both types of changes may have occurred, as changes were small, generally not statistically significant, and not always consistent across the three years (Table 9). Following a trivial decrease in the first year, primary care visits outside of hospitals increased by about one-half of a visit per person (or approximately five percent) in the second and third years, although neither estimate was statistically significant.

In contrast to the estimates for primary care visits, estimates for specialist visits outside of the hospital setting registered decreases, and a reduction was measured in each demonstration year, although only one year's impact estimate was large enough to reach statistical significance. This category of visits may have declined by about one-quarter of a visit in the first year, and about half of a visit in the second and third years, for an average of 0.42 fewer visits per person annually during the entire three-year intervention period, or -7.3 percent. Although the second- and third-year values were similar (-0.52 and -0.48, respectively), only the second year's estimate was statistically significant. An impact on this category of visits might have resulted from acute-care-prevention strategies used by the IAH practices, given that substantial specialist care tends to occur during hospitalizations and may lead to follow-up visits. Further, primary care practitioners might have provided services that could substitute for specialist services, or perhaps they reduced referrals to specialists, or both.

In terms of all types of professional visits in all settings, the demonstration appears to have led to a decline of more than two visits PBPY in Years 1 to 3, but only in the first year was this change statistically significant. Total professional visits include all visits billable under Medicare Part B and include home visits, office, clinic, and rehabilitation therapist visits, and visits occurring during a hospitalization or other institutional stay. In the year before the demonstration, the average annual number of professional visits per IAH beneficiary was 63 PBPY (Table 9). The -3.7 percent impact for total professional visits is in line with the impact shown in Table 5 for professional and outpatient claims expenses.

## Table 9. IAH impact on primary care and specialist visits outside of hospitals and total visits: Years 1 to 3

| | Treatment Group Mean | Demonstration Impact Estimate† | | | Entire Intervention Period Estimate†† | Approximate Percentage Impact††† |
|---|---|---|---|---|---|---|
| **Outcome measure** | **Baseline Year** | **Year 1** | **Year 2** | **Year 3** | **Years 1-3** | **Years 1-3** |
| Number of E&M visits outside of hospitals by primary care[a] providers PBPY | 11.24 | -0.11 | 0.44 | 0.58 | 0.28 | 2.5% |
| Number of E&M visits outside of hospitals by specialists PBPY | 5.66 | -0.27 | -0.52* | -0.48 | -0.42 | -7.3% |
| Total number of visits in all settings by all clinical providers PBPY | 63.06 | -2.60* | -2.16 | -2.24 | -2.35 | -3.7% |

Source: Medicare claims, 2010-2015 (baseline and demonstration periods). Utilization measures were risk-adjusted and weighted to reflect partial year observations.

†The demonstration impact estimate is the amount of change due to the demonstration after accounting for the pre-existing difference between the treatment and comparison groups. See text and Appendix C for further explanation of the methods.

†† Entire intervention period estimate is the average change in the measure (e.g., PBPY) for the 3 years combined into a single period..

†††The approximate percentage impact is the entire intervention period impact divided by the average value in the treatment group in the baseline year, which is the year before the demonstration.

[a]Primary care providers include primary care physicians, nurse practitioners, and physician assistants.

[b]Therapist visits in this table are Part B visits to beneficiaries in offices, clinics, federally qualified health centers, etc., but do not include visits by therapists when a beneficiary uses the Medicare home health benefit.

*, **, *** indicate statistical significance at .10, .05, and .01 levels, respectively

PBPY=per beneficiary per year; E&M = evaluation and management.

**Health status indicators.** Available indicators for health status impacts were mixed but individual measures are subject to limitations or further review. The demonstration appears to have had no adverse impact on mortality, a possible unfavorable impact on the likelihood of beneficiaries' entering long-term care (LTC), and no impact or a possibly favorable one on self-reported general health status.

Twelve-month-mortality impact estimates for each year were very small, lacked consistent signs, and were not statistically significant (Table 10). During the demonstration, mortality differences between the treatment and comparison groups changed little, another indication that mortality was likely not affected. Mortality rates among IAH beneficiaries were below those of their comparisons during the pre-demonstration and demonstration years (data not shown). However, mortality rose slightly between the two years before the demonstration for treatment beneficiaries, while it fell slightly during the same period for comparisons, causing some uncertainty about how to extrapolate the difference in baseline mortality rates in the absence of the demonstration. The implication is that the mortality analysis should be interpreted with caution and continue with more years of data before drawing a final conclusion.

To examine LTC entry, we analyzed the probability that a beneficiary would enter a nursing home each year, for both IAH beneficiaries and their comparisons. The rate of entry was lower for IAH beneficiaries in all the years, but the pattern of changes each year produced a smaller difference between the groups, relative to the group difference in the year before the demonstration (data not shown). Therefore, the difference-in-difference analysis resulted in a net *increase* in the LTC entry rate, attributable to the demonstration (Table 10). The impact--a statistically significant two-percentage-point increase in the rate of LTC entry on average, or nearly 16 percent--was unexpected (Table 10). Aside from the unexpected outcome, the trajectory of rates featured an unusually large drop in rates in Year 2 for both groups (data not shown). Both the unusual pattern as well as the impact findings are undergoing further investigation and additional validation.

### Table 10. IAH impact on one-year mortality and rate of entry into LTC: Years 1 to 3

| | Treatment Group Mean† | Demonstration Impact Estimate† | | | Entire Intervention Period Estimate†† | Approximate Percentage Impact††† |
|---|---|---|---|---|---|---|
| **Outcome measure** | **Baseline Year** | **Year 1** | **Year 2** | **Year 3** | **Years 1-3** | **Years 1-3** |
| One-year mortality | 18.13% | -0.13% | 0.28% | 0.82% | 0.31% | 1.7% |
| Beneficiary entered long-term care[a] (%) | 12.31% | 1.06%* | 2.61%*** | 2.33%*** | 1.94% | 15.8% |

Source: Medicare claims 2010-2015 (baseline and demonstration periods) and Medicare enrollment data. Measures were risk-adjusted and the long-term-care entry measure was weighted to reflect partial year observations.
†The demonstration impact estimate is the amount of change due to the demonstration after accounting for the pre-existing difference between the treatment and comparison groups. See text and Appendix C for further explanation of the methods.
††. Entire intervention period estimate is the average change in the measure (e.g., PBPY) for the 3 years combined into a single period.†††The approximate percentage impact is the entire intervention period impact divided by the average value in the treatment group in the baseline year, which is the year before the demonstration.[a]LTC entry was established using a claims-based algorithm that found physician bills for required periodic assessment visits during Part B nursing home stays, from the first appearance of such a bill until the end of the year.
*, **, *** indicate statistical significance at .10, .05, and .01 levels, respectively.
LTC=long-term care

We also examined health status with survey data. We compared IAH beneficiaries' self-assessed health status with responses from beneficiaries in the Medicare Current Beneficiary Survey (MCBS). Self-assessed health status results are not necessarily an outcome of the demonstration because we were unable to account for pre-existing differences at baseline. Also, we were unable to apply the IAH eligibility criteria precisely in selecting the MCBS sample. Therefore, results are considered exploratory. IAH beneficiaries' self-reported health status was comparable to that reported by national respondents to the MCBS in 2013 who approximately met the IAH qualifying criteria (Table 11). Results also suggested that the subgroup of IAH beneficiaries who lived in private homes (as opposed to assisted living) were nearly 25% more likely than the MCBS respondents to say that their health improved compared to the previous year.

## Table 11. Self-reported health status and health status change: IAH enrollees compared to MCBS respondents

| | Percentage of IAH beneficiaries reporting yes (A) | Percentage of MCBS respondents reporting yes (B) | Difference (A-B) | Percentage difference (A-B)/B |
|---|---|---|---|---|
| **Health compared with other people the same age: excellent, very good, good or fair (versus poor)** | | | | |
| Beneficiaries living in assisted living facilities | 76.3 | 79.5 | -3.2 | -4.0% |
| Beneficiaries living in the community | 69.5 | 67.7 | 1.8 | 2.7% |
| **Health compared with one year ago: much better, somewhat better, or about the same (versus somewhat or much worse)** | | | | |
| Beneficiaries living in assisted living facilities | 56.0 | 58.8 | -2.8 | -4.8% |
| Beneficiaries living in the community | 63.2 | 51.0 | 12.1*** | 23.7% |

Source: 2013-2015 survey of beneficiaries enrolled in the demonstration and 2013 MCBS.
Note: We approximated the IAH eligibility criteria when selecting MCBS respondent records for this comparison. The sample consisted of 3,293 MCBS respondents and 341 IAH enrollees.
MCBS=Medicare Current Beneficiary Survey

### Qualitative data: practices' changes potentially attributable to the demonstration

Our periodic site visits and interviews allowed us to observe evolution in the IAH practices' quality-related activities and organizational strategies.[29] We cannot determine whether all the reported changes were attributable to the demonstration, because we lack comparative information from other practices outside of the demonstration; the evaluation was not designed to collect external practice information. Therefore, findings of the qualitative data collection should not be taken as definitive information about practices' responses to the demonstration.

Over the course of three years, many practices told us that they enhanced their traditional structures and strategies for achieving multiple aims: managing transitions across settings, coordinating care in an effort to prevent unnecessary hospital and emergency department use, and improving quality of care. The approaches taken were as varied as the practices themselves. However, relationships and communication with patients, caregivers, and other organizations that served IAH beneficiaries were targets of improvement for many practices. Care processes were also a focus of improvement; for example, practices designed new care protocols and changed home visit scheduling. This section summarizes IAH practices' self-

[29] Following an initial round of site visits to the IAH practices (February 2013 to May 2013), we conducted a second round of visits during Year 2 (February 2014 to July 2014), and a third round in Year 3 (April 2015 to October 2015).

reported activities directed at meeting the IAH process and outcome quality measures tied to incentive payments (Figure 2). We also describe additional organizational strategies they deployed to improve care quality.

**Strategies to meet the IAH quality measures.** The demonstration's incentive payment methodology incorporated process quality measures that required timely follow-up in the home in the transition from acute care, as well as contact with the patient within 24 hours of admission to, or discharge from, a hospital stay or emergency department (Figure 2). Because of the focus on these process quality measures, practices reported that a good deal of effort was expended on achieving timely notification of the patient's recent admission or pending discharge. Communication with hospitals to learn when IAH patients would be discharged was critical to meeting the IAH targets for medication reconciliation at home within 48 hours of discharge. But reliable notification of such events was not always assured. Although many practices were aided by automated notices from some area hospitals, or received automatic notification through interconnectivity with a related institution, these supports didn't always solve the problem in individual cases. Practices were often dependent on the cooperation of external parties to relay when and where the patient was being treated. A common response on the part of the practices was to attempt to strengthen relationships with other providers or institutions likely to have knowledge of patients' whereabouts, such as social service agencies and assisted living facilities, with the objective of facilitating more-reliable notification. Some practices added staff for tracking admissions and discharges.

Another challenge, as reported by IAH practices, was conducting the follow-up home visit for medication reconciliation within the period required to meet the demonstration process measure (48 hours). Most of the academic medical center practices and one independent practice added nursing staff or contracted with nursing services to ensure a visit occurred within the specified time frame, especially on weekends. One practice changed scheduling procedures to allow for last-minute changes that enabled a clinician to work a post-discharge visit into the day's appointments.

Regular telephone outreach to patients, newly instituted by some practices, was also a strategy used to help practices meet the process and outcome quality measures. On an ongoing basis, practices tried to educate patients and caregivers to contact the practice when contemplating a trip to the ED. Practices used their encounters with IAH beneficiaries to educate patients about available home services that could substitute for use of the ED. To facilitate patient-initiated communication, in Year 2, some practices concentrated on improving the ability of patients to contact the practice after hours—for example, by contracting with a visiting nurse association to triage after-hours calls.

Collaborative relationships with care partners may have improved due to the demonstration, as suggested by information from a number of practices' care partners.[30] These sources noted IAH

[30] Care partner interviews took place by telephone in 2016 and early 2017. We asked each IAH practice to provide contact information for a small number of care partners, such as home health agencies and suppliers. We conducted a total of 48 interviews with care partners.

providers' efforts to provide education about the goals of the demonstration and how care partners fit in with those goals, particularly the intent to reduce unnecessary acute care. Goal-sharing and education resulted in domiciliary facility staff communicating more closely with a few IAH practices on a routine basis. In some cases social workers or nurses involved in managing acute-care transitions directly intervened with the hospital to explain that HBPC could substitute for discharge to a post-acute institution.

**Other care improvement responses.** Improving primary care quality was a multi-faceted effort. Practices introduced new care protocols, such as reminders for preventive services, or in other ways sought to standardize patient care processes or implement specific quality improvement initiatives. Some instituted periodic or additional home visits, in part to identify emerging problems before they developed into acute issues. Practices sought to better capitalize on the clinical expertise of their leaders or colleagues through staff meetings and supervisory activities in which solutions for managing patients more effectively were discussed. A few practices involved clinicians in reviewing their patients' acute-care episodes to see what might have been done differently to avoid use of the hospital or ED. Practices also reported using hospital admission notifications to immediately implement readmission prevention strategies for the patient involved.

During the demonstration, IAH practices' use of data to improve the effectiveness of the organization also grew, but some of their experiences showed there were limits to the ability to effectively use data. In several practices, IAH project managers or data analysts regularly created reports for tracking post-hospitalization follow-up and other IAH requirements. A few practices hired or deployed staff to work with data, while others instituted methods such as routine chart audits to assess compliance with treatment standards. Although a few practices obtained customized reports through their EHR, some practices could not use their EHR usefully to create summary reports, or resorted to copying data into other software programs or databases for analysis. A few practices worked on data-sharing across organizations to enhance their access to information about their panel of patients.

By Year 2, most practices established some form of patient risk assessment or risk stratification using either data sources or clinician judgment. Practices variously conducted assessments to assign risk level, used risk classification tools, or used internal or external data resources to understand patient characteristics and conditions associated with increased risk of admission. Such information was used to allocate more care coordinator contacts and more frequent home visits to patients perceived to be at higher risk of hospitalization.

## Progress on IAH demonstration quality measures during the first three years

The demonstration's quality measures consisted of one set tied to incentive payments and another set intended to monitor performance for general administrative purposes. Two IAH process quality measures tied to incentive payments were highly dependent on quick identification of patients admitted to or discharged from the hospital or the ED (Figure 2). Despite the practices' efforts to improve timeliness of notifications, achievement rates on the

measures suggest the challenges in receiving timely notifications. In Year 3, a minority (five) of the 14 practices met both the measure for contacts within 48 hours for at least 50 percent of hospital admissions, hospital discharges, and ED visits, and the measure for medication reconciliation in the home within 48 hours for at least 50 percent of hospital and ED discharges. Furthermore, the achievement rates for process measures show little evidence of change over the years, possibly suggesting that operational and other changes made by the sites had limited effect. However, baseline data for the process measures are not available to enable a proper test of the demonstration's impact on them. As with the process measures, the outcome quality measures such as all-cause readmissions were stable and lack a baseline.[31] (See Appendix B, Table B2, for tabular summary of quality measures achieved.)

IAH quality measures that were not tied to payments required systematic recordkeeping and reporting on the part of the practices. However, practices' reporting on these measures, such as symptom assessments, screenings, and other care-process activities, appeared highly variable, even across years within a practice. Given some of the problems that sites encountered in data management for quality improvement and incentive-related reporting, we were reluctant to interpret missing data as an absence of the reportable activity (for more information about the reporting system, see Appendix A). This circumstance limited what we could learn about additional quality-related care processes for possible future application.

### Preliminary analysis of factors associated with successful practices

Although our sample is small—only 16 practices[32]—we drew some preliminary conclusions about characteristics that might portend successful overall performance for a HBPC practice under the demonstration. We used information from site visits and other qualitative data about sites' structural and operating characteristics to associate performance with best practices. To classify the sites according to how successfully they performed, we considered both annual expenditures and IAH quality performance measures.[33] The approach to examining annual expenditures was to compare average expenditures of IAH beneficiaries and a contemporaneous comparison group (after adjusting for differences in demographic and health characteristics).[34] This approach differs from the impact analysis presented earlier in this report

[31] The quality measures tied to incentive payments were calculated by the demonstration implementation contractor. The measures were based on IAH beneficiaries enrolled by the practices. See Appendix C for information about the sample differences between the enrollee lists and the evaluation sample. Some measures reported in the quantitative findings section may seem conceptually related to the quality outcome measures; for example, the IAH quality measures for ACSC hospital admissions and ACSC ED visits are conceptually similar to the evaluation's outcome analysis of preventable admissions and preventable ED visits. However, the two approaches differ in many methodological details, including the goal of the analysis and the basic strategy for identifying an effect. Therefore, results from the two sources are not comparable.

[32] For purposes of this analysis, the three academic practices in the consortium (Philadelphia, PA; Richmond, VA; and Washington, DC) were treated as separate participants.

[33] Quality measures for Years 1 and 2 only were used, solely due to availability at the time of analysis.

[34] In the expenditure analysis, we compared all of a practice's beneficiaries who met IAH eligibility criteria in administrative data with a sample of matched comparisons who met the IAH criteria but did not use home visits. We matched a contemporaneous set of comparisons each year, and we matched separately the group

in that it does not rule out pre-existing unobservable differences, such as beneficiary care preferences, as a reason for expenditure differences. Furthermore, it should be noted that, because only statistically significant expenditure differences (at the 10 percent level) were used to determine whether a practice was financially successful, smaller practices would generally need to show larger differences relative to the expenditures of the contemporaneous comparison group.

Based on the expenditure and quality performance measures, we then assigned the sites to one of three groups: high performers, moderate performers, and low performers (Table 12). The highest-performing group consisted of two practices; they achieved statistically significant, lower expenditures than their comparison group in all three years and met all 6 quality measures tied to incentive payments in the first two years. The moderate performers consisted of three sites that achieved statistically significant, lower expenditures in two or three of the years and met four quality measure standards for both of the first two years. The low performers consisted of the remaining 11 practices; they achieved lower expenditures in only one year or not at all, and their quality performance varied.

### Table 12. Classification of IAH practices for preliminary analysis of best practices

| Group 1: HIGH PERFORMERS (N=2) | Group 2: MODERATE PERFORMERS (N=3) | Group 3: LOWER PERFORMERS (N=11) |
|---|---|---|
| Achieved savings (3 years) and met all 6 quality measure goals (2 years) | Achieved savings (2 to 3 years) and met 4 quality measure goals (1 to 2 years) | No savings or savings in one year only |

We looked for commonalities among members of the three groups, in terms of selected operating and structural information found in our qualitative data. The characteristics we examined had to have a conceptual relationship to cost and quality. We concluded that the following operational features seem to be associated with success:

- frequent in-home visits,
- reliable weekend coverage by clinicians for post-hospital-discharge follow-up,
- ongoing coordination with community resources, and
- regular quality tracking and improvement efforts.

In addition, practices that conducted a greater proportion of visits in assisted living facilities as opposed to private homes had lower costs and met quality measure goals. Anecdotal information from several interviews with assisted living facilities housing IAH beneficiaries suggested that assisted living facility staff proactively notified IAH practices when their patients experienced a change in condition, fall, or medication issue. Timely communication was important for effective medical management of IAH beneficiaries. Some assisted living facility

---

of IAH continuing beneficiaries who did not meet all IAH criteria. We statistically adjusted for demographic and other differences between the groups using regression analysis. Due to differences between samples and methodologies agreed to by practices for determining their incentive payments, these results do not necessarily reflect incentive payment results.

staff reported that they gained new clarity about their potential role in care coordination early in the demonstration, after the IAH practice explained the demonstration's goals.

## Summary, Conclusions, and Limitations

For this report on the first three years of the IAH demonstration, we examined impacts from multiple sources, both quantitative (administrative data and surveys) and qualitative (site visits and interviews with IAH practices and their care partners). In this concluding section, we summarize the key findings, discuss their implications, and note limitations of the study.

### Expenditures and utilization impacts

We saw indications in both qualitative and quantitative data suggesting that the demonstration might have led to a modest decrease in the key metric of total Part A and Part B Medicare expenditures, particularly in the first and third years. By the third year, the estimate of change in total annual expenditures was greatest, -$177 PBPM (or a 4 percent reduction), but the reductions in expenditures were too small to be statistically significant and, therefore, could not be conclusively attributed to the demonstration. The sole exception was DME expenditures PBPM, which fell by an average of nearly 15 percent. Table 13 summarizes the estimated savings in the aggregate. For the three-year period, our point estimate for the reduction in total expenditures is $24.7 million, but the confidence intervals are very wide and indicate the possibility that the demonstration did not produce savings.

In keeping with the inconclusive expenditures results, in each year we measured reductions in several key utilization measures, but usually the changes were not statistically significant. These key utilization measures that consistently indicated reductions were total hospital admissions, preventable hospital admissions, ED visits leading to hospitalization (all shown in Table 6), specialist visits outside of hospitals, and total clinician visits (both shown in Table 9). In many of these measures, the Year 3 change was largest but not large enough to reach statistical significance, given the small sample size. Among the measures with consistent reductions, only preventable hospital admissions and the annual number of ED visits leading to hospitalization declined in Year 3 by statistically significant amounts—0.03 admissions (-7.6 percent) and 0.12 visits (-8.4 percent), respectively. With one exception, all other Year 3 reductions in utilization measures (Tables 6, 8, and 9) were not statistically significant. The exception was the proportion of beneficiaries with at least one unplanned readmission, which declined by 1.71 percentage points in the third year (Table 6). Finally, the only outcome measure demonstrating a statistically significant reduction for the entire three-year intervention period was the annual number of ED visits leading to hospitalization, which averaged -0.08 ED visits PBPY, or -5.8 percent (Table 6).

Aggregate estimates of total hospital admissions and readmissions suggest that 974 hospital admissions might have been avoided, and 197 cases with at least one readmission within a year were possibly avoided, during the entire three-year intervention period. As with expenditures, statistical precision for these estimates was poor and allowed for the possibility that the demonstration did not lead to reductions in admissions and readmissions (see Table 13 confidence intervals). The most statistically reliable measure of events avoided was for ED

visits leading to hospital admission. We estimate a total of 1,560 such ED visits were avoided during the entire three-year period; we can be 90 percent confident that ED visits were avoided, given the confidence interval (-2,668, -451).

No other major category of utilization—including skilled nursing, hospice, and inpatient rehabilitation stays, and ED visits not leading to admission—demonstrated statistically significant change. Whereas annual numbers of visits with clinicians overall may have decreased by 3.7 percent—due in part to a reduction in specialist visits outside of hospitals of 7.3 percent—primary care visits outside of hospitals apparently rose by 2.5 percent (Table 9). Although the increase in primary care visits was not statistically significant in any year or on average, greater use of primary care is consistent with some of the acute-care preventive strategies pursued by IAH providers under the demonstration. For example, opinion among substantial numbers of IAH practice staff was that the IAH quality measure relating to home visits for medication reconciliation is extremely useful in avoiding future episodes of acute care, in view of IAH patients' extensive health care needs. Clinical staff pointed to difficulties that can result when patients or caregivers misunderstand how to modify existing care regimens upon returning home from a hospital stay or ED visit.

The diversity of practices that participated in the demonstration permitted us to test whether impacts differed across three different categories of IAH practices. This analysis was an attempt to understand whether some types of practices were more successful than others at lowering expenditures and hospital utilization. The answer is yes. Results for independent practices provided strong evidence that as a group they achieved substantial savings and reductions in acute care, as measured by total hospital admissions, readmissions, and ED visits leading to hospitalization. The exceptional performance of this category of practices suggests that the demonstration can be effective with some practices. However, we lack a clear understanding of why the independent practices were more effective.

## Table 13. Aggregate impact of IAH demonstration on selected outcomes: Years 1-3 and cumulative results for the three years combined

***The point estimates show savings (indicated by minus signs) but the wide confidence intervals indicate there is uncertainty in whether the demonstration actually led to savings***

| Outcome | Aggregate impact in Year 1 | 80 percent CI | 90 percent CI | Aggregate impact in Year 2 | 80 percent CI | 90 percent CI | Aggregate impact in Year 3 | 80 percent CI | 90 percent CI | Cumulative aggregate impact across the three years | 80 percent CI | 90 percent CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total Medicare expenditures | -$9,741,494 | -$19,616,739; $133,752 | -$22,412,928; $2,929,941 | -$2,193,523 | -$14,638,251; $10,251,205 | -$18,161,992; $13,774,946 | -$12,758,376 | -$27,297,276; $1,780,525 | -$31,413,985; $5,897,234 | -$24,693,394 | -$59,800,028; $10,413,241 | -$69,740,518; $20,353,731 |
| Number of hospital admissions[a] | -328 | -633; -23 | -719; 63 | -177 | -559; 204 | -667; 313 | -468 | -952; 15 | -1,088; 152 | -974 | -2,035; 88 | -2,335; 388 |
| Number of outpatient ED visits[b] | -219 | -653; 214 | -776; 337 | 411 | 8; 815 | -106; 929 | -109 | -635; 418 | -784; 566 | 84 | -1,002; 1,169 | -1,309; 1,476 |
| Number of ED visits resulting in hospital admissions | -457** | -716; -199 | -789; -126 | -375 | -702; -48 | -794; 45 | -728** | -1,095; -361 | -1,199; -257 | -1,560** | -2,424; -696 | -2,668; -451 |
| Number of beneficiaries having a qualifying hospital discharge and an unplanned readmission within 30 days of discharge | 2 | -78; 81 | -100; 103 | -69 | -171; 33 | -200; 62 | -129** | -213; -45 | -236; -22 | -197 | -436; 42 | -503; 110 |

Source: Medicare claims and enrollment data for 2011–2015.

Note: This table shows the aggregate impact estimates for key outcomes for IAH-eligible treatment beneficiaries, over all IAH practices during years 1, 2, and 3 of the demonstration. The estimates for expenditures do not take into account incentive payments made to IAH practices. The calculations are based on the beneficiary-level estimates reported in Tables 5 and 6 and on the number of IAH beneficiaries and eligible beneficiary months in each year. Specifically, the aggregate results for total expenditures, number of acute-care admissions, and ED visits are calculated by multiplying the beneficiary-level impact estimate by the number of IAH beneficiary months (for expenditures) or by the number of IAH beneficiary months in each year divided by 12 (for admissions and ED visits that are measured yearly). The aggregate results for unplanned readmission are calculated by multiplying the beneficiary-level impact estimate by the number of IAH beneficiaries in each year. The total numbers of IAH beneficiaries in the annual analysis sample were 8,216 in Year 1, 7,266 in Year 2, and 7,564 in Year 3. The numbers of eligible beneficiary months for the same numbers of IAH beneficiaries were 79,396 in Year 1, 69,768 in Year 2, and 72,215 in Year 3, or a total of 221,379 months of care.

[a]The number of hospital admissions includes observation stays.

[b]The number of ED visits not resulting in hospital admission includes those resulting in observation stay.

*/**/***Significantly different from zero at the 0.10/0.05/0.01 level, two-tailed test.

CI = confidence interval; DD = difference-in-differences; ED = emergency department; LL = lower limit; UL = upper limit

### Health status indicators

We found no impact on mortality from the demonstration, a result that suggests the demonstration did not adversely affect beneficiary health in general. An unexpected finding was that LTC entry rates rose as a result of IAH, because the reduction in the rate was smaller for IAH than for comparisons. Further analysis is necessary to confirm this result.

### IAH practices' response to the demonstration

The fact that various change estimates for expenditures (e.g., total, inpatient hospital, professional services, DME in Table 5) and for utilization (e.g., hospital admissions, preventable admissions, readmissions and ED visits in Table 6) indicated the strongest impact occurred in Year 3 could mean that impacts of the demonstration were taking time to develop during the first three years. We learned of activity among the practices consistent with this pattern, based on qualitative data collected from the sites and their care partners. In particular, practices worked to strengthen their relationships and refine communication methods with information sources, particularly hospital personnel, domiciliary care staff, and other professionals involved in the care of the patient, not to mention patients and their caregivers. Relationship-building was perceived to be a means to improve care coordination, promote the benefits of HBPC, and meet IAH process quality measures by improving real-time notification of patient acute-care episodes. Adding staff members, such as social workers, or redeploying staff to help in coordinating care with outside organizations further indicated that communications could have improved under the demonstration. Practices' educating hospitals and care partners about their model of care may also have increased opportunities to smooth the process of post-discharge planning and thereby potentially avoid readmissions (an outcome measure that improved by 1.71 percentage points in Year 3 [Table 6]). The large number of area hospitals for some practices made the task of strengthening relationships with partner institutions particularly challenging, implying there is room for more progress in achieving consistent timely notice of patient admissions.

Practices redoubled efforts or developed new strategies to avoid acute-care use. Among the responses were increased use of regular visit schedules and more reliance on periodic telephone outreach to patients or caregivers. Practices also engaged in more education of patients about available home services and, most important, about the need to contact the office before going to the ED. To realize the benefits of these activities, some practices had to make themselves more accessible to patients by, for example, hiring weekend staff. The change in ED visits resulting in hospitalization—the strongest and most persistent acute-care change we found (Table 6)—is one possible indication of the effectiveness of these and other actions taken by the sites. Although the evaluation cannot trace a direct relationship from the activities of the sites to the expenditures and utilization changes we measured, the evaluation found consistency between qualitative information and key quantitative estimates, notwithstanding that the latter were modest.

### IAH beneficiaries' views of their care

As shown in the section describing IAH practices' structure and operations, patient perspectives on the quality of their primary care during the demonstration were a mix of highly favorable and somewhat unfavorable views when compared to the views of non-IAH beneficiaries. While patient and caregiver satisfaction with HBPC appears to be high, some specific measures of care quality reported by IAH beneficiaries were unfavorable when compared to views among IAH-qualified MCBS beneficiaries asked to assess their usual source of care. However, conclusions from the survey data are subject to clear limitations. Because beneficiaries chose the type of primary care to use, differences between the groups are not necessarily attributable to the demonstration or HBPC delivery model, as they could reflect self-selection factors. And because we surveyed beneficiaries at a single point in time, differences are not necessarily attributable to the demonstration.

### Preliminary information on best practices

Given the available information to date, our analysis of best practices suggests the following four characteristics may be linked to successful performance of a practice in the demonstration: frequent in-home visits, reliable weekend coverage arrangements for post-discharge follow-up, ongoing coordination with community resources, and regular quality tracking and improvement efforts. We also found that high performance may be associated with a large share of visits in assisted living, which is possibly a reflection that communication about multiple patients is easier when a smaller number of sources are involved. Our preliminary analysis of best practices is subject to limitations, the main one being that results of the analysis of financial performance are liable to reflect unmeasured differences between the treatment and comparison groups. Misclassification of practices into the three performance groups (high, moderate, low) could have resulted.

### Lessons learned in implementing incentive payments in HBPC

We encountered challenges in the methodology for calculating incentive payments under the demonstration (see Appendix B for more details about the IAH incentive payments methodology). We learned that for practices of the size that participated, and with the demonstration's target population being severely chronically ill and functionally limited, calculating statistically reliable savings estimates is a formidable task technically. These group practices generally do not have large numbers of clinicians, and the delivery model, including the amount of time spent in travel, inherently constrains the size of a clinician's feasible patient panel. Moreover, patient panels include patients who don't meet the IAH eligibility requirements, a situation that whittled down the sample size available for analysis. With small samples, results can be volatile from year to year and not precisely measured in any year, which means the actual amount of savings can be subject to fairly large uncertainty. Under the statute's risk corridor requirement, CMS avoided making incentive payments for uncertain estimates (CMS used a 90 percent confidence level). The implication is that small practices are less likely to earn a payment unless the savings estimate is large.

Data augmentation strategies may help address some of these challenges in calculating incentive payments, but could require longer periods to accumulate data. Strategies may

include a longer performance period than a single year for calculations or more frequent use of consortia by participating practices; with longer performance periods, more data accumulate for a practice, and with consortia, the combining of data from multiple practices enlarges the volume of data. Additional research into the practices' full population of patients, instead of just those meeting the IAH eligibility criteria, could provide additional information on the impact of HBPC and has the potential to mitigate small-sample-size problems. Short of a sample size large enough to estimate savings with more certainty, alternative financial incentives, such as care management fees linked to quality measures, may avoid some of the challenges. Specifically, care management fees can be awarded for achieving process benchmarks whose measurement does not involve statistical uncertainty.

Even assuming adequate amounts of data, given the types of data available, assuring valid methods for establishing a spending target representing amounts that would have been spent in the absence of the demonstration is likely to be a persistent challenge for populations like IAH beneficiaries. Except for the one practice that did not switch from the actuarial method described in the solicitation, we used matched comparison groups extracted from CMS administrative data to identify the spending target for Years 1 and 2 incentive payments. The IAH beneficiary qualifications are markers for very costly, ill patients; finding a comparable group for benchmarking in administrative data is possible but the process doesn't guarantee that all important characteristics (such as being homebound) are accounted for. A challenge that estimating savings has in common with most observational designs is the risk that treated sample members may self-select the treatment while the characteristics that determine the decision are unobservable and therefore cannot be statistically controlled. For example, administrative data contain no information about attitudes that lead to choosing HBPC—a variable that can theoretically drive results but could be confounded with the HBPC model effect we intended to measure in the analysis employing data from a comparison group to generate the spending target. Because unobservable group differences could be responsible for, or at least contribute to, estimates of group expenditure differences, further research into additional confounding variables is worth pursuing.

### Limitations of the evaluation study

This report addresses one basic question: What were the outcomes of the demonstration? The design of the demonstration did not allow us to address another important question: What are the outcomes of HBPC relative to usual care? The main reason was that the demonstration was not limited to new users of HPBC; it is not possible to obtain a clean assessment of HPBC's effects when study subjects have prior experience with it. Given the demonstration design, we could reasonably address whether the IAH demonstration had an impact over and above any effects (which remain unknown) attributable to the HBPC delivery model itself.

The demonstration's small size—starting with only 18 practices, to adhere to the 10,000-beneficiary cap—led to a major limitation of the analysis: samples too small to measure modest changes with statistical confidence. Moreover, the constrained number of participating practices prevented us from projecting results to HBPC practices generally.

Additionally, this report's study of the demonstration was observational and, as such, involved inherent caveats. First, our research design depended on the assumption that treatment and

comparison groups would have trended in parallel had there been no demonstration. The assumption is fundamentally untestable, although we found support for it in data showing that pre-demonstration trends of the treatment and comparison groups were parallel for most outcome measures.

Second, we relied on comparison samples that we matched to the IAH beneficiaries for the analysis, but truly rich data for establishing comparable treated and untreated groups are not available. We applied the same procedures when selecting matched samples of treatment and comparison groups in each of the five years and in measuring risk-related characteristics that we held constant in the analysis. However, unlike a randomized trial, the study is not capable of ruling out effects from unmeasurable characteristics that may differ between the groups– effects that could make comparisons falsely attribute an impact to the demonstration. An example would be a change between the baseline and intervention periods in unmeasurable variables that lead to patients choosing HBPC and that also affect outcomes such as expenses and utilization. A specific concern of this kind is that the longevity of some beneficiaries' disability is not measurable in administrative data. Although HBPC in IAH was typically targeted to beneficiaries with long-term functional impairment, beneficiaries with short-term as opposed to long-term disablement could make up differentially changing shares of the treatment and comparison groups over time. But since we could not measure this type of risk factor, we were unable to control for it.

Another caveat is that the comparison group did not come from HBPC practices outside of the demonstration; it is doubtful that suitable comparator practices could have been found. Therefore, the design also depended on the assumption that any IAH practice changes underlying the outcomes were not part of a broader trend specific to HBPC regardless of the demonstration.

We also acknowledge two issues with generalizability of the evaluation to the target population of beneficiaries. First, the results of the quantitative impact analysis do not necessarily pertain to all beneficiaries whom the demonstration intended to include as participants. In the interests of a research design capable of identifying an impact from the IAH demonstration, IAH beneficiaries who continued in the demonstration each year without meeting the IAH requirements again are not represented in the samples. Frequently, carryovers did not re-experience both an acute and a rehabilitation stay. The restriction of the study population to beneficiaries who met the IAH criteria each and every year could mean that our results are not applicable to chronically ill and frail patients of IAH practices who avoided recent acute or rehabilitation stays. Similarly, each year we excluded potential new entrants whose claims histories did not confirm that they met all IAH requirements, even if the practices enrolled them in the belief that they met the requirements based on information at the site[35]; and we included other beneficiaries whom the practices did not confirm as eligible, because administrative data supported their eligibility. The difference in eligibility determination mostly related to the required ADL limitations, which can differ in an individual across different sources and measurement occasions. The problem in measuring ADL limitations points to a second generalizability issue. For example, if adding nonenrolled patients who met IAH eligibility

[35] Both excluded groups are discussed further in Appendix C.

requirements in administrative data made the sample represent eligible patients and others who no longer had at least two ADL limitations, then the results may be generalizable to a broader population than one having at least two limitations.

Additional years of data will be incorporated into forthcoming analyses. With the added data, we intend to update impact estimates in a future report. Analysis for more years of the demonstration is expected to further refine conclusions. We will also add Medicaid expenditure data to the analysis if sufficient data become available.

In summary, results of the IAH demonstration are promising but inconclusive after the end of three years. We found generally small estimates of cost and utilization avoided and, though measures were often consistent with one another, these changes may or may not have been real. At the same time, performance in the third year signaled that results might turn more favorable as we add more years of data. Our experience implementing the demonstration also revealed that a well-functioning incentive payment methodology for practices of the kind that participated faces technical challenges.

# Appendix A: IAH Demonstration Implementation

In this appendix, we provide an overview of how CMS selected practices to participate in the IAH demonstration, and explain reasons for the reduction in the number of practice participants after the demonstration's launch. We also describe the processes and resources used to oversee implementation of the IAH demonstration. Additionally, we recount key developments that affected the demonstration.

**Applicant selection process.** In response to the authorizing legislation, CMS issued a solicitation for practices that qualified for IAH under the legislation. Only practices that were already providing home-based primary care (HBPC) and met certain criteria (e.g., HBPC services must be provided to at least 200 demonstration-eligible beneficiaries per year) were qualified to apply (see main report). The solicitation allowed practices to apply as a single entity, consisting of one solo practice that operates independently; a consortium, consisting of two or more small practices that form together to create one legal entity; or as part of a national pool, consisting of single entities that are combined to calculate the expenditure target, but remain independent for the calculations of the quality measures and incentive payments.

The application period for the IAH demonstration was held in 2011-2012. Because the statute directed the Secretary to limit the number of selected practices so that the number of applicable beneficiaries does not exceed 10,000 beneficiaries, CMS selected 18 applications to participate in the IAH Demonstration: 15 single entities and three consortia.

To evaluate applications, CMS followed four steps:

1) Identified and selected only applications that were complete.
2) Verified eligibility of all applicant organizations with respect to the following legislatively specified criteria:
    - Practices that demonstrated experience in serving high-cost chronically ill patients at home;
    - Possessed the capacity to provide in-home care for at least 200 eligible patients annually;
    - Were organized to provide care in a team-based environment that includes staff such as social workers, pharmacists, and others;
    - Had around-the-clock availability;
    - Used patient-centered care plans;
    - Indicated in their proposals that they would meet CMS data submission requirements for enrollment, quality monitoring, and quality measurement purposes.

    As part of this verification process, CMS examined Medicare claims data and conducted internet searches for information on the applicant organizations.
3) Categorized applicants by the number of potentially eligible beneficiaries, geographic location, and type of organizational affiliation (e.g., whether an applicant operated independently or was a part of a health system).

4) Selected the final set of applicants to represent a good mix of geographic variation, practice size, years of experience, and practice affiliation. (For a broad snapshot of the practices selected, see Table 3 in Report to Congress.)

Before officially accepting the 18 selected applicants into the demonstration, CMS worked with its Center for Program Integrity to conduct a program integrity screening on each. Following this step, the selected practices began their first annual period of performance in 2012.

**Reduction in the size of the demonstration.** Practices could choose to end their participation in the IAH demonstration early. In Year 1, one practice left the demonstration because it could not continue to meet the data reporting requirements. In Year 2, two consortia left IAH. One consortium withdrew from Year 2 because of staffing difficulties as well as ongoing internal business issues, such as purchasing a new electronic health record system. The other consortium left because most of its individual members ended their relationship with the main HBPC practice. The main practice tried to continue on with the demonstration as a single entity but found it was unable to comply with demonstration requirements around data entry and the 200-beneficiary minimum, so it later withdrew. Lastly, CMS did not extend participation to one practice for the fourth and fifth demonstration years due to issues with the practice's recent program integrity history. Thus, four of the original 18 practices had ceased participation in the IAH demonstration by the beginning of Year 4.

**CMS Reporting System and data sharing.** In order for CMS to implement the demonstration operations of enrolling patients and administering incentive payments, IAH practices were required to submit data on their patients' IAH enrollment or disenrollment status, documentation of patients' IAH eligibility, and information for use in calculating certain quality measures. To collect the needed information, CMS' implementation contractor created the IAH Reporting System (RS). The RS stored information about beneficiary eligibility criteria determined from CMS administrative data (e.g., date of qualifying event); eligibility information reported by practices (e.g., disenrollment, ADL limitations); and quality measure information reported by practices for the outcomes and processes tied to the incentive payment and for the six performance measures not tied to the incentive payment. The RS also functioned as a repository for demonstration-wide documents such as the legislation text and IAH practice handbook. Overall, the RS was widely used with little issue by the IAH practices.

One major improvement made to the RS during the demonstration was the ability of practices to create reports. The two reports the practices could extract at any time from the RS enabled them to see their individual beneficiary participant data as well as information they had inputted to the RS for quality measure calculations and enrollment tracking purposes. CMS' implementation contractor was also able to extract reports for the purposes of tracking enrollment counts and the progress of demonstration-wide quality measure reporting.

In early 2013, CMS began to provide quarterly spreadsheet workbooks to the practices to share data. Initially the workbooks provided were designed to help the practices identify beneficiaries who met the qualifying criteria for IAH using administrative data. In 2016, the workbooks were enhanced to add information relevant to the practices' financial performance. The data included average expenditures per beneficiary per month for services such as home

health, DME, or hospice, as well as selected utilization averages (e.g., average number of observation stays and average length of stay for a hospitalization). Another enhancement completed later in the year was to show beneficiary-specific summary information about expenditures and utilization such as ED visits and non-IAH practice physician expenditures.

**Communication and technical support.** For the IAH practices, most communication was via the IAH Help Desk, which was operated by CMS' implementation contractor. The Help Desk had both a telephone number and email. The most common reasons a practice contacted the Help Desk were for technical assistance around RS data entry or data security, and to perform administrative functions, such as scheduling calls or site visits with CMS or the implementation contractor. On some occasions, a practice would reach out directly to the CMS project officer. This was typically for questions regarding timelines, information for the practice's stakeholders, legislative activity, or specific scenarios the practice believed only CMS could answer. Information flow was coordinated with the implementation contractor by including the Help Desk in the CMS response.

In addition to maintaining the Help Desk, CMS' implementation contractor was responsible for conducting "touch-base" telephone calls and multiple rounds of site visits to each practice. The touch-base calls were held either with individual practices or with all IAH practices. Individual practice calls tended to focus on practice-specific needs and problems, such as claims processing complications that arose when a practice obtained an additional Tax Identification Number. All-practice calls focused on implementation topics, such as training on the RS. The site visits were conducted to learn about each practice's delivery of care, through such activities as attending a home visit and practice team meetings. Another purpose of the site visits was to perform medical record audits. The medical record audits were designed to confirm the information entered into the RS for a small sample of cases and to resolve any discrepancies between the practice-reported information in the RS and the Medicare claims system. Discrepancies discovered in this manner were usually due to user input errors and were easily fixed once identified.

**Independent Learning Collaborative (LC) and its role.** In 2012, the American Academy of Home Care Medicine obtained grant funding to convene an independent learning collaborative for practices participating in the IAH demonstration. One of the LC's purposes was to facilitate the participating IAH practices' discussions about the demonstration, and the LC also proposed modifications to the demonstration. Proposals included suggestions for alternatives to the incentive payment methodology, changes to the quality measure requirements, and research opportunities related to the delivery of care outside of the home. The LC sought to generally serve as advocates on behalf of the IAH practices. CMS accepted some of the LC's recommendations. For example, for purposes of meeting the quality measure requiring contact with each patient within 48 hours of admission to or discharge from a hospital stay or ED visit, CMS adopted the LC's recommendation to alter its definition of contact to include telephone contact.

**Shared Savings and Quality Measures Evolution.** During the demonstration, CMS modified the incentive payment calculation twice and quality measure thresholds for two measures once.

Both modifications were made to account for issues discovered during the course of the demonstration.

For a description of changes to the incentive payment calculation, see Appendix B. In addition to the incentive payment adjustments, CMS reduced the threshold for two of the quality measures used to determine the proportion of the calculated savings that was disbursed to the practices as an incentive payment. In the first year of the demonstration, the IAH practices informed CMS how difficult it was to find out when a beneficiary visited an emergency department or was admitted to a hospital. From this feedback, CMS believed it was in the best interest of the demonstration to reduce the threshold so that the level used would reflect high-quality but obtainable levels of performance. CMS did not want to penalize practices that were doing a good job of tracking their beneficiaries in the face of impediments related to interoperability with healthcare systems. Therefore, in 2014, CMS reduced the threshold from 80 percent of events to 50 percent of events for the following two quality measures:

1) the 48-hour follow-up contact for admission to the hospital and discharge from the hospital and/or emergency department, and
2) the home visit for medication reconciliation within 48 hours after hospital or emergency department discharge.

The change in the threshold was made before CMS finalized Year 1 incentive payment disbursements.

# Appendix B: Incentive-payment and Quality-measurement Methodologies

The statute stipulated that incentive payments must be based on savings in total Part A and Part B expenses and must be subject to performance on quality measures determined by the Secretary. To meet this requirement, CMS designed a methodology in which the estimated savings were adjusted according to quality performance. Below we explain the methodologies underlying both components of the incentive payment—savings measurement and quality measurement—and how they were used together to determine the incentive payment. We also report the outcomes of both components of the incentive payment for the practices.

The savings estimation methodology underwent changes during the demonstration. Specifically, the approach to determining the target expenditure for comparison with the practices' actual expenditure was revised as reasons for making changes were raised by the practices or became known after additional data came in. In brief:

- Both CMS' solicitation and the demonstration participation agreement described an actuarial methodology based on county-level average expenditures adjusted for two summary measures of beneficiary health. For determining incentive payments, this target was to be compared with actual average expenditures of the demonstration enrollees.
- During the demonstration's second year, at the practices' request and before any incentive payments were made, CMS agreed to reconsider the target methodology in the participation agreement. Practices' concern was that the target would be set too low.
- CMS delayed making incentive payments and in the following year offered practices a choice to stay with the participation agreement methodology or move to a different one which was based on identifying a comparison group in Medicare administrative data. All but one practice selected the new approach, and incentive payments were issued.
- When generating the second year's incentive payments, CMS concluded that comparison group selection in administrative data was not capable of accounting for enrollment issues that could arise at the practices, resulting in likely imbalances in beneficiary characteristics between the two groups. As a result, CMS offered a change that both revised the comparison group selection procedure and--provided that the analysis of enrollees showed actual expenses were below the target--made an adjustment to the target and expense calculations. The adjustment was based on analysis of a sample that included qualifying beneficiaries beyond those enrolled by the practices.
- For the second year's incentive payments, one practice continued with the original actuarial methodology, and most other practices elected to move to the newly revised comparison group methodology for Year 2.

The incentive payment amounts in Table B2 (below) reflect the methodology chosen by each practice in each year. So that practices could make an informed choice, CMS explained the reasons for and details of each methodology before they made their selection, including the anticipated incentive size of each alternative. Practices then chose their preferred methodology. This resulted in practices choosing the methodology that gave them the highest payment. Therefore, incentive payments made under the demonstration appear to reflect the most advantageous approach as perceived by the practices.

Later in this appendix we explain in more detail the reasons why changes were made.

## Overview of incentive-payment and quality-measurement methodologies

**Incentive payment methodology.** The statute required CMS to establish a per-capita spending target representing the estimated amount that would have been spent on total Part A and Part B services for IAH beneficiaries in the absence of the demonstration. For a practice to be eligible for an incentive payment, actual expenditures had to be at least five percent below the target (after accounting for the incentive payment amount). Medicare retained the first five percent saved. The remaining savings were then adjusted for quality performance. Specifically, a practice was eligible to share in a maximum of 80% of any savings beyond the first five percent, according to a schedule that depended on the number of quality measures achieved. Practices had to achieve a minimum of three of the six quality measures to qualify for any incentive payment at all. (See Background section for further details.)

The law also stipulated that spending results should be evaluated and incentives determined for each year. The mechanics of the savings calculation were performed on the average monthly per-beneficiary expenditures for the practice's applicable beneficiaries during a 12-month performance period. The spending target was also expressed as average expenses per beneficiary per month (PBPM). The percentage saved was derived from comparing a practice's actual average PBPM with the target average PBPM. The estimate of gross spending in the absence of the demonstration was computed as the product of total beneficiary months in the demonstration for enrollees and the average monthly target.

**Quality measurement methodology.** CMS sought to use a broad range of quality measures for purposes beyond adjusting the incentive payments. Other purposes included a need to identify assessment, planning, implementation, and monitoring/evaluation priorities and activities for the demonstration. Several process measures, mainly assessments of various kinds, were not tied to payment; these were briefly discussed in the Report but the focus in this appendix is on the measures tied to payment.

The demonstration's six quality measures linked to the incentive payment were a mix of process and outcome measures (Figure 2). Together, the process and quality measures were intended to focus on activities that drive cost reduction while promoting overall good quality. The legislation indicated that an objective of the IAH model was to reduce unnecessary acute care; consequently, CMS selected outcome measures that dealt directly with acute care: ambulatory-care-sensitive (ACSC) hospitalizations, ACSC ED visits, and hospital readmission. The three process measures tied to payment were contact with beneficiaries within 48 hours

upon admission to the hospital and discharge from the hospital and/or ED; medication reconciliation in the home within 48 hours of hospital discharges and ED visits; and whether patient preferences were discussed with the patient and documented at least annually. These measures addressed considerations of timeliness for measurement, making early feedback possible, and were also considered drivers of spending.

### Quality measure results

If a practice achieved all six quality measures tied to payment, it earned the entire available maximum of savings. CMS established a minimum achievement rate, or threshold, for each process measure to be able to credit a practice with that measure; for example, at least 80 percent of IAH patients had to have their care preferences documented annually to meet the patient preferences measure. The process measures required practices to report that a contact, visit, or discussion and documentation occurred. The implementation contractor conducted on-site audits to examine consistency of the reports with information in the medical record. For outcome measures, risk-adjusted utilization among IAH enrollees was compared to expected utilization among IAH-eligible beneficiaries not receiving HBPC in the same geographic area. If the practice's number of cases was equal to or below the expected number of cases, the practice was credited with achievement.

Table B1 lists the quality measures tied to payment and shows achievement results for each practice by year. The first three measures, from left, are the process quality measures:

1) follow-up contact within 48 hours of hospital admissions, discharges, and ED visits, which had to be performed in at least 50 percent of events;
2) medication reconciliation in the home within 48 hours of hospital discharges and ED visits, which had to be performed in at least 50 percent of events; and
3) patient preferences discussed with the patient annually and documented in the medical record (required for at least 80% of enrolled patients).

The next three quality measures are the risk-adjusted utilization outcome measures:

4) number of hospital readmissions within 30 days;
5) number of ambulatory-care-sensitive[36] hospital admissions; and
6) number of ambulatory-care-sensitive emergency department visits.

[36]Hospital admissions and ED use for diabetes, CHF, and COPD were considered ambulatory-care-sensitive admissions for purposes of the fifth and sixth quality measures tied to payment.

## Table B1. IAH demonstration quality measures tied to payment: Measures achieved, by practice, Years 1 to 3

| | 48-hour follow-up | | | Medication reconciliation | | | Patient preferences recorded | | | Risk-adjusted readmissions | | | Risk-adjusted ACSC admissions | | | Risk-adjusted ACSC ED visits | | | Number of measures achieved | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 | Year 1 | Year 2 | Year 3 |
| Austin, TX | | | | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 3 | 3 | 3 |
| Boston, MA | | ✓ | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 5 | 4 |
| Brooklyn, NY | | | | | | | ✓ | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 4 | 2 | 3 |
| Cleveland, OH | | | | | | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 4 | 3 |
| Dallas, TX | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 4 | 4 |
| Durham, NC | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 6 | 6 | 6 |
| Flint, MI | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 4 | 4 |
| Jacksonville, FL | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 4 | 4 |
| Lansing, MI | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 4 | 4 |
| Mid-Atlantic Consortium | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 6 | 6 | 6 |
| Milwaukee, WI | | | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 4 | 4 | 4 |
| North Shore, NY | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 5 | 6 | 6 |
| Portland, OR | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 6 | 6 | 6 |
| Wilmington, DE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | 6 | 4 | 6 |
| Number of practices that met quality measure | 5 | 6 | 5 | 4 | 5 | 5 | 13 | 12 | 11 | 14 | 14 | 14 | 14 | 13 | 14 | 14 | 12 | 14 | AVG 4.6 | AVG 4.4 | AVG 4.5 |

Source: IAH demonstration reporting system data obtained from demonstration implementation contractor for all IAH practices that participated in demonstration Year 3, excluding Louisville.

Note: See text for measure description. ACSC=ambulatory-care-sensitive condition, ED=emergency department

To determine benchmarks for the utilization measures, statistical models generated risk-adjusted estimates of the numbers expected in the geographic area among the IAH-eligible population not using HBPC.

The amount of the final incentive payment depended upon quality performance. To qualify for incentive payments, a practice must have met or exceeded performance requirements on at least three of the six quality measures tied to payment. Practices that met fewer than three of the quality measures were not eligible for an incentive payment. Practices that met three, four, five, or six of the quality measures were eligible to receive 50%, 67.7%, 83.3%, or 100% of the savings that qualified for sharing, respectively.

Tallies of measures achieved, as shown in Table B1, indicate that the individual practices performed with consistency across the years. In the third year the average number of measures achieved was not the highest, but Year 3 was the first time that 5 practices (Wilmington, Durham, Mid-Atlantic, Portland, North Shore) achieved all six measures, making all of them eligible to receive the entire maximum share of savings. The entire maximum share of savings was 80 percent of the amount remaining after CMS retained the first five percentage points. An example will be provided below in the presentation of the incentive payment results.

### Implementing an IAH incentive payment methodology

CMS made technical changes to the incentive payment methodology two times during the early years of the demonstration, resulting in three methodologies in use for Year 2 incentive payments. Below we briefly describe the three methodologies for determining savings for incentive-payment purposes, and follow the description with an explanation of reasons for the changes. The three payment methodologies were 1) the original "actuarial" approach; 2) a comparison-group-based approach that used a matched comparison group and regression adjustment; and 3) a revision of the second approach.

To determine incentive payments, CMS initially developed, and the IAH practices had agreed to, the methodology laid out in the solicitation. The methodology used a risk-adjusted target based on average monthly expenses in each IAH beneficiary's county. This approach is referred to as the actuarial methodology. Analysis of additional information provided by practice representatives led to CMS' offering practices the option to remain with the original actuarial methodology or to agree to switch to a new methodology based on matched comparison groups. This was the second demonstration methodology.

During preparation of the second year's incentive payment analysis, CMS determined that the comparison-group-based savings analysis was affected by sample selection issues that distorted results. As we will explain below, the issues arose because the comparison-group sample was identified by algorithms applied to administrative data, and they did not align with processes occurring in the practices to identify enrollees for the demonstration. Therefore, revisions were made to the second methodology that resulted in the third methodology.

**Actuarial methodology.** For determining incentive payments CMS originally established a methodology using a risk-adjusted target based on average monthly expenses in each IAH beneficiary's county. This approach was called the "actuarial" method.

Savings were determined by comparing actual spending for IAH enrolled beneficiaries relative to a spending target. The formula for determining the target expense for a beneficiary in the actuarial method was:

> Target Expense for each IAH enrollee=
> Average FFS Cost in County of Residence * Trend * (Risk Adjustment Score[37] + Frailty Factor)

FFS cost in county of residence (per beneficiary per month), trend, and frailty adjustment factors are established each year by CMS. The trend represents the expected average increase in the per-beneficiary-per-month (PBPM) Medicare Part A and B costs since the last year on

[37] CMS-HCC and CMS ESRD risk models, version 21, 2013 software, were used. Version 21 is used by CMS for setting rates in the PACE program.

record. Applying a trend factor is necessary because the IAH benchmark is intended to measure costs contemporaneously with the demonstration performance year, but FFS county costs are only available with a lag. The risk adjustment score was the CMS hierarchical condition category (CMS-HCC) score, calculated individually for each IAH enrollee by the demonstration implementation contractor. CMS developed the CMS-HCC model to set Medicare Advantage and PACE program payment rates. The model predicts expected Medicare expenditures based on diagnostic information and demographics in the prior year. The score is derived from the predicted expenditures. The frailty factor was a quantity added to the risk score to reflect the beneficiary's impairments in activities of daily living (ADLs) that may increase the costs of care. The practices reported the impairments of each of their enrolled beneficiaries. Given data for each element of the formula, the average cost for a practice was computed over all of its enrolled beneficiaries (weighted by months of enrollment) and formed the practice's expenditure target. The target was compared with the actual average cost incurred by the same beneficiaries.

During the second year of the demonstration, before CMS calculated incentive payments for Year 1, analyses undertaken independently by the American Academy of Home Care Medicine suggested the actuarial method might cause the spending target to be set too low. The information raised concerns that underestimation would cause CMS to mismeasure savings, i.e., the difference between actual and target cost level might be erroneously small or negative. Specifically, there were concerns that the actuarial method components—average FFS cost and risk adjustment score—were derived from data that did not accurately represent the intended IAH-eligible population. Therefore, the methodology might not correctly project the expenses of IAH beneficiaries, i.e., beneficiaries whose poor health status and other characteristics made them a sicker group with costs tending to the high end of the expense distribution.

The actuarial method for Year 1 projected an average loss of $73 PBPM (about 1.0%). Using this actuarial method, of the 17 practices that completed Year 1, seven showed losses, and five others had small savings results that were not statistically significant. The remaining five had estimates of savings that ranged from approximately $1.1 million to about $2.5 million, before CMS' share of savings was retained and quality adjustments were applied.

**Comparison-group-based methodology, Year 1.** In response to these concerns regarding the actuarial method, CMS delayed disbursing incentive payments until analysis of the evaluation samples could be completed. In the evaluation, the samples were not deployed to compare average expenses of HBPC and non-HBPC beneficiaries, either for practices individually or demonstration-wide. However, the design of the samples lent itself to making expense comparisons, with limitations (discussed further below). In contrast to the $73 PBPM loss estimated by the actuarial approach, the comparison-group-based method for Year 1 projected an average savings of $341 (7.4%) PBPM; five of the 17 practices that completed Year 1 showed losses, and four practices had savings estimates that were not statistically significant. The remaining 8 practices had estimates of savings ranging from approximately $1.9 million to about $7.5 million, before CMS' share of savings was retained and before quality adjustments.

The evaluation team created a matched comparison group to study the demonstration's impact. The method of matching was propensity scores. Matching was intended to identify patients who were like the IAH population but did not use HBPC. The comparison group consisted of non-participants residing in the same geographic areas where IAH beneficiaries lived, and comparison group members were very similar in observable characteristics to IAH beneficiaries. Because of these similarities, it was thought that the evaluation sample would better represent the projected expenses in the absence of the demonstration, thus addressing the concerns about the actuarial method.[38] To ensure consistency in the start of observation for all beneficiaries, each beneficiary had an assigned IAH eligibility date, which was the month after the month in which the beneficiary met all the IAH requirements. Many subjects met the eligibility criteria before the performance year began, and thus had an IAH eligibility date predating the performance year; in such cases, the start of observation was the first month of

[38] In contrast to the purpose for which it was used in the incentive-payments application, the purpose of the comparison group in the evaluation analysis was to identify the expenditure trend in the absence of the demonstration.

the performance year. If a beneficiary met the requirements after the performance year began, observation began with the month after eligibility.

In creating a comparison group to implement the evaluation's analytic strategy, we identified a set of individuals well-matched to the IAH beneficiaries, but deploying it to estimate savings for purposes of generating incentive payments was not without challenges. Two main challenges arose out of our reliance on secondary data sources, i.e., CMS administrative data, to identify the comparison group. First, we could not use the entire treated population. Demonstration procedures allowed participating practices to enroll beneficiaries upon presenting confirmation to CMS that a patient met IAH health status criteria (at least two ADL limitations and at least two chronic conditions) based on their assessment of the patient. We did not have any physician-reported functional status and chronic condition information for comparison beneficiaries. However, for both treatment and comparison individuals, we did have administrative data on these characteristics from claims and rehabilitation-provider assessments. We considered it necessary for scientific reasons to verify IAH criteria using the same procedures for both treatment and comparison group members. Therefore, for identifying the evaluation sample, we did not use information provided by the practices to confirm that patients met any of the IAH criteria. Instead, sample selection for both treatment and comparison group members was based on administrative information. The result was that approximately 29 percent of the IAH enrolled population in 17 practices that completed Year 1 were excluded from the analysis in that year because, according to administrative data, they did not meet eligibility criteria.[39] In subsequent years, we continued to omit from the analysis enrolled beneficiaries whose eligibility we could not confirm in administrative data. For more information on reasons why we could not confirm IAH eligibility, see Figure B1.

Second, CMS administrative data may contain only some of the relevant information we need to assure equivalence of the treatment and comparison groups. Specifically, administrative data do not allow measurement, for example, of beneficiary motivations and preferences, residence in assisted living, and extent of family support. These variables are plausible considerations in enrollment decisions and they also may be related to beneficiaries' expenses. As a result of our inability to measure these potentially relevant variables, equivalence of the two groups is not assured, and the analysis could end up crediting the program with savings that are partially due to group differences. A key assumption of propensity score matching is that achieving a good match on observable characteristics also reduces unobservable differences between the treated and comparison beneficiaries, if the latter are correlated with matching variables. However, this assumption is not testable with available data. Using regression analysis, the comparison-group-based method attempted to adjust for observable differences that remained after matching. Nonetheless, non-equivalence of the groups is still a possibility in the incentive-payments methodology we developed from the evaluation's matched sample.

Upon reviewing the two sets of results, CMS decided that the comparison group approach for assessing actual versus target expenditures resulted in a more accurate assessment of the performance of IAH providers than the actuarial method. Consequently, CMS decided to allow practices the option to agree to switch to the comparison-group-based method as the foundation for determining incentive payments in Year 1. Under either the actuarial or the comparison-group method, CMS applied each practice's enrolled-beneficiary months of care (officially recorded by the demonstration implementation contractor) to the estimate of average monthly savings generated by the analysis. This procedure yielded each practice's gross savings (or loss).

For contractual reasons, CMS gave the practices the option to agree to switch to the new methodology or to remain with the actuarial method, the latter being the method the practices had agreed to as part of the demonstration terms and conditions. Seventeen IAH practices that completed Year 1 received their incentive-payment results for that performance period under both methods. Only one practice chose to remain with the actuarial approach; this practice's savings rate doubled under the actuarial approach relative to the comparison-group method.

[39] In contrast to the purpose for which it was used in the incentive-payments application, the purpose of the comparison group in the evaluation analysis was to identify the expenditure trend in the absence of the demonstration.

## Figure B1: IAH enrollees not confirmed in administrative data

The official enrollment in IAH Year 1 was 8,445. Reasons why beneficiaries who were enrolled in the demonstration were not found eligible in administrative data in Year 1 are shown below.

| Reason | Number of beneficiaries |
|---|---|
| No ADL information in administrative data in the given year | 1,157 |
| No ADLs or one ADL needing human assistance in the given demonstration year | 575 |
| Fewer than two chronic conditions | 17 |
| No qualifying hospitalization | 38 |
| No qualifying rehabilitation services stay | 6 |
| In hospice | 66 |
| Failed to meet FFS criteria* after becoming eligible for the demonstration | 40 |
| Identified in administrative data as a long-term nursing home resident for the entire demonstration year after they became eligible for the demonstration | 187 |
| Eligible for IAH but had no Part B claims for home visits from the IAH site | 180 |
| Other | 1 |
| **Total** | **2,267** |

Note: Our procedure for attributing beneficiaries to a practice required the beneficiary to reside in the same state as the IAH practice; an additional 138 beneficiaries who enrolled in the demonstration could not be included in the evaluation sample for this reason. *FFS criteria were that beneficiary had to be enrolled in both Part A and Part B.

**Comparison-group-based methodology, Year 2.** For the analysis of Year 2, CMS anticipated that it would continue to follow the Year 1 enrollees and their matched comparisons, and add to the analysis sample any newly eligible enrolled beneficiaries, along with new matches for them.[40] However, upon analyzing the Year 2 samples, several issues became apparent, which led to revisions of the comparison-group-based approach. The methodological concerns leading to revisions included multiple signs that the comparison group in Year 2 was not sufficiently well matched. The eventual solution involved revising the methodology in the following ways:

- we changed the design from one in which we followed IAH beneficiaries and their initial matched comparisons over multiple years, to a design in which we rematched continuing IAH beneficiaries to new comparisons each year;
- we used new variables in matching;
- we used a separate matching procedure for beneficiaries who continued their IAH care from the previous performance year without meeting the IAH beneficiary requirements again;
- we expanded the sample to include beneficiaries who were eligible for IAH and served by IAH practices but who were not necessarily enrolled by the practices; and
- if a practice's savings estimate was statistically significant at the 90 percent confidence level in the analysis of the enrolled sample, we used results from the expanded sample to adjust the results from the analysis of the enrolled sample.[41]

The revised methodology for Year 2 and the process of developing it is described below.

*Signs of non-comparable groups in Year 2*

Upon examining outcomes of the sample selection process in Year 2, we found that the treatment and matched comparison samples had very similar rates of health conditions and other variables used in matching. However, we also noted several unexpected results. When put together, the unexpected results indicated that the comparison beneficiaries were not adequately matched to the treatment population.

First, a large mortality difference suggested that the health status of the treatment beneficiaries and their matched comparisons was not truly comparable. The adjusted mortality

[40] We continued to restrict the sample to beneficiaries whose eligibility we could confirm in administrative data.

[41] The significance test of the enrolled-sample-based estimate was used to determine whether a practice would be eligible for an incentive payment. If the results indicated that a practice was eligible, the adjusted savings estimate was also tested for statistical significance in order to implement the risk corridor requirement.

of newly eligible treatment beneficiaries was much lower than that for their matched comparisons. Although it is possible that care from the IAH practices affects mortality, the size of the difference was not plausible. Furthermore, large differences in mortality could complicate the assessment of savings, because expenditures in the last months of life typically rise, and can raise the average monthly expense during the performance year. As a result, the spending target might be set too high, simply due to group differences in the timing of death. For example, even with comparable overall health status, one group with earlier mortality on average could have magnified average expenses in the first of two years, and the second group could have magnified expenses in the second year, if mortality in the second group occurs later.

Second, we saw indications that differential attrition between the treatment and comparison populations in Year 1 caused non-comparable groups of survivors by Year 2. In following surviving IAH patients and their surviving matches into Year 2, it appeared that the two populations no longer had the same profile of health characteristics. The surviving members of the comparison group from Year 1 appeared to be somewhat healthier than the surviving members of the treatment group from Year 1; for example, they needed assistance with fewer activities of daily living and they had fewer chronic conditions. With the survivor subgroups' reduced comparability, comparisons on cost of care would be problematic.

Third, we saw a need to account more explicitly for the trajectory of expenditures that is typical of beneficiaries after an acute stay. In Year 1, IAH practices had a large group of patients eligible for IAH in June, when the demonstration began. After the year started, only a relatively small number of additional patients became eligible. In contrast, although many comparison beneficiaries were eligible in June (based on hospitalizations during the previous 12 months), an equally large number became eligible during the year, i.e., a relatively high percentage of comparison group beneficiaries experienced their qualifying hospitalization or post-acute care use during Year 1, as opposed to the preceding 12 months. As a result, matched comparison beneficiaries were more likely than IAH beneficiaries to have their performance-year data include the costs of their IAH-qualifying hospitalization, as well as related high costs typical of the aftermath of an acute stay. The comparison-group based method used regression to control for this, based on the number of months observed during the performance year.[42] However, the trajectory of costs related to acute care could also be an issue in data from beneficiaries who continue in the demonstration from the previous year, because many re-experience hospitalization. These trajectory-related considerations for both new and many continuing beneficiaries created a need for a different approach to dealing with the problem.

*Revisions to improve comparability*

We addressed the problems observed in the preliminary Year 2 samples by making several changes. First, we revisited the health-status variables used in matching. It was assumed that revising the list of matching variables could change the composition of the matched comparison group. If the matching variables were associated with the risk for mortality specifically and expenditures more generally, their inclusion would achieve better overall comparability of the treatment beneficiaries and their comparisons. We focused on finding variables associated with high mortality, in view of the mortality differences we observed and the evidence of differential attrition, which was mostly due to differences in mortality. We also included a matching variable to model the length of time between the beneficiary's last hospitalization prior to becoming eligible for IAH and the IAH eligibility date, since there is a trajectory of expenditures following hospitalization, with expected costs declining over time. Specifically, the revised approach added a requirement that an IAH beneficiary could be matched to a given comparison beneficiary only if the two beneficiaries had a similar amount of time since their most recent hospitalization.

Second, we considered further the implications of differential attrition we had observed in the preliminary assessment of the samples. Differential attrition might be a problem for each year's new IAH entrants, with the potential for comparability issues to arise in the following

[42] In our expenditures and utilization regressions, we control for categories of the number of Medicare FFS-eligible months during a predemonstration or demonstration year (specifically, categories for 4–6, 7–9, and 10–12 months, vs. 1–3 months), starting from the first month of IAH eligibility. This method helps to account for any systematic difference in the claims-based eligibility dates for treatment versus matched comparison beneficiaries.

year. As a result, rather than follow beneficiaries and their initial matches across the years, we instead separated IAH beneficiaries into two groups for matching. The first group consisted of everyone who met the IAH qualifying criteria for purposes of Year 2, regardless of their inclusion in the sample in Year 1. All beneficiaries, including newly eligible ones and continuing ones, were tested against the complete set of IAH qualifying criteria and, if they met the IAH beneficiary criteria, all were matched in a single procedure using their health and demographic characteristics for Year 2. The decision to merge requalifying and new beneficiaries into a single group for matching, and implement matching based on characteristics in Year 2, placed members of both groups on an equal footing with respect to their health and demographic characteristics at the outset of the performance year.[43] Also, the decision conformed with precedents in other settings. For example, in the Medicare Advantage program, beneficiary characteristics are updated each year when establishing the applicable insurance premium.

The second group consisted of all continuing beneficiaries in Year 2 who had not "requalified" for the demonstration, i.e., "carryovers," or Year 1 sample members who did not meet the IAH eligibility criteria when tested against the criteria again, for purposes of Year 2. These patients were allowed to continue as participants under demonstration rules, as long as they continued to receive treatment from IAH practices. But being a carryover beneficiary often meant that no qualifying hospital or rehabilitation stay occurred during the historical period examined for purposes of Year 2. Therefore, administrative data lacked some history information used in matching. Given the likelihood of less-recent data for some variables in the matching list, we matched carryover beneficiaries in a separate procedure, dropping several matching variables unlikely to have been recorded recently. Like the carryover beneficiaries, the comparison beneficiaries matched to them met the IAH qualifying criteria in Year 1; however, for all carryovers and their comparisons, we generally used characteristics measured for Year 2 for matching. When conducting the regression analysis for incentive payments, we pooled the carryover and non-carryover beneficiaries—and their matches—into a single data set. Carryover beneficiaries comprised 18 percent of the data set in Year 2.[44]

*Persistence of mortality differences and resulting expansion of the sample*

Despite the changes to matching variables and the subdivision of the sample into two matching groups, implausibly large mortality differences persisted between the enrolled participants and the comparison group. To understand the reasons for the persistent difference in mortality between the IAH beneficiaries and their comparisons, we considered information about the enrollment process obtained during site visits to the IAH practices. We concluded that the enrollment process might have systematically discouraged enrollment of beneficiaries at high risk of mortality at the time they would have enrolled in the demonstration. Failure to enroll patients near death may be a legitimate and clinically appropriate decision. For example, practices might have rejected enrollment for especially sick patients who were close to death in the belief that they could not be helped by the program at such a late stage in life. Yet the matching approach using administrative data could not adequately replicate this type of selection decision. Additionally, enrollment verification procedures themselves apparently inadvertently led to excluding some dying beneficiaries; reasons tracked by the implementation contractor for sites not enrolling beneficiaries included numerous instances in which beneficiaries died before practices notified them of the demonstration or completed the enrollment process.

Because we could not replicate the enrollment process using administrative data, we could not create a parallel process for selecting comparisons in the administrative data. Many more beneficiaries in the comparison group than beneficiaries in the IAH demonstration would die within the first few months of demonstration eligibility. The consequence was likely to be an erroneous comparison between the treatment and comparison beneficiaries, due to the typically high costs in the weeks and months leading up to death.

---

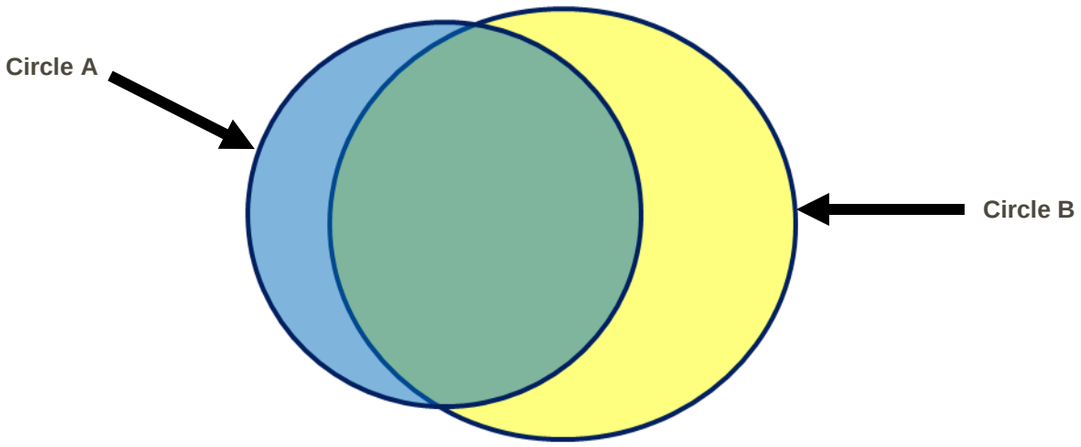
[43]Rematching continuing beneficiaries was a departure from the perspective of the original comparison group design, in which each treatment and matched comparison would continue under observation without allowing for any change in matched comparisons, other than that due to attrition. Any theoretical benefit to a practice of rematching a continuing beneficiary may be mitigated by the IAH incentives to avoid preventable hospitalizations, in the form of the quality measures tied to payment.

[44] Carryover beneficiaries were not used in the evaluation samples. See Appendix C for further information on exclusions from the evaluation sample.

Upon review of the site-visit information explaining the persistence of a large mortality differential between the treatment and the comparison groups, we determined that we needed a treatment group free of the unobserved exclusionary mechanisms that came with the enrollment process. To move away from relying on discretionary enrollment to create our sample, we expanded the IAH treatment group to include all Medicare FFS beneficiaries whom we identified as eligible for the demonstration in the administrative data and who had received HBPC from an IAH practice. Therefore, the expanded sample consisted of enrollees whose eligibility could be confirmed in administrative data, as well as other patients of the IAH practices who met IAH requirements in administrative data but whom the sites did not confirm as enrollees. Earlier in the evaluation's qualitative data collection we heard from practices that generally they did not treat IAH patients differently from the way they treated others. This information lent support to the decision to expand the sample beyond patients enrolled by the practices.

The revised comparison-group-based method for Year 2 incentive payments, offered as a choice that practices could agree to use for calculation of Year 2 and Year 3 incentive payments, incorporated all of the changes outlined above and used the expanded IAH treatment sample.[45] Specifically, the regression analysis of the expanded sample was used to adjust the analysis that was generated from the enrolled sample. Figure B2 illustrates the difference in sample selection between the enrolled sample, which was used for Years 1[46] and 2, and the expanded sample, which was used to adjust enrolled-sample results only in Year 2. In both Year 1 and Year 2, the intersection of Circle A and Circle B (green oval) was used for analysis of the enrolled sample —that is, the enrolled sample included only enrolled beneficiaries who met the IAH requirements in administrative data. However, in Year 2, we also used results of the analysis of Circle B (yellow circle) to adjust results from the analysis of the Year 2 green oval sample. That is, for deriving the adjustment factor, we added to the intersection area by including non-enrolled patients of the practices who met the IAH requirements in administrative data (Circle B without Circle A, shown as yellow crescent). The expanded sample had a larger sample size than the enrolled sample but still omitted the beneficiaries enrolled by the practices whose eligibility could not be confirmed in administrative data (Circle A without Circle B, shown as blue crescent). As noted earlier, the official beneficiary-months of enrollment were used to calculate the practice's gross savings or loss.

**Figure B2. Groups of IAH beneficiaries based on different identification processes**




Key:

Circle A (blue circle) = beneficiaries who were on the implementation contractor's list of IAH enrollees (IAH enrollees).

Circle B (yellow circle) = beneficiaries whom the evaluation identified as eligible (IAH beneficiaries).

Intersection of Circle A and Circle B (green oval) = IAH enrollees whom the evaluation also identified as eligible.

Circle A without Circle B overlap (blue crescent) = IAH enrollees whom the evaluation did not identify as eligible.

Circle B without Circle A overlap (yellow crescent) = beneficiaries whom the evaluation identified as eligible but who were not enrolled in IAH.

[45]The incentive payments results for Year 1 were not adjusted in light of improvements in the comparison-group methodology developed later.

[46] The enrolled sample for Year 1 consisted of IAH enrollees who met the demonstration eligibility requirements in administrative data. This is the same approach used for identifying the enrolled sample of Year 2. However, as explained in the text, several technical changes to matching procedures for the enrolled sample were implemented for Year 2.

Figure B2 is concerned only with patients who meet the IAH-qualifying requirements, whether identified by practices or in administrative data. It is worth noting that the numbers of patients that qualified for the IAH demonstration and met criteria for inclusion in the sample were only a subset of the practices' wider Medicare clientele. For the size of practices in the demonstration, the IAH enrolled subset comprised relatively small samples. Small samples can make estimates vulnerable to statistical instability because the smaller the number of patients, the more likely that, if an outlier occurs, the average expenses will be strongly affected. This situation had real implications for several practices in Year 2. After we disbursed incentive payments for Year 2, we found errors in our procedure for reconciling beneficiary lists across shared savings initiatives. These procedures were designed to ensure that no more than one program or model paid shared savings for the same patient.[47] Correcting the mistakes added modest numbers to the treatment and comparison samples (4 percent) and also caused a substantial reduction in the amount of incentive payments payable to several practices, even though the specific sites that earned an incentive did not change and the average change in savings for the practices was $33 PBPM.[48] Consistent with our obligation to make accurate payments and to protect the trust funds, practices were required to refund the overpayments that had been made. This development highlighted instability in PBPM savings estimates that may occur with practices of the size that participated in IAH.

Under the revised comparison-group-based method for Year 2, Medicare savings averaged $89 (2.3%) PBPM.[49] Practices that had selected the comparison-group method for Year 1 incentive payments were given the option to agree to switch to the revised methodology for Year 2, and most did so. As noted earlier, differential attrition in the treatment and comparison groups rendered samples with somewhat different risk profiles by Year 2. This generally led to large estimated losses when applying the Year 1 methodology to Year 2, but not in every case. Of the 14 practices that completed Year 2 and had chosen the comparison-group-based method in 2015, 11 agreed to switch to the revised comparison-group approach over the Year 1 methodology; the other three practices showed more savings under Year 1 methods than under Year 2 methods and opted to stay with the Year 1 approach.

**Implications of the experience with IAH incentive payment methodologies.** The experience with the evolution of the methodology for calculating incentive payments illustrates the challenges in determining savings estimates for the participating practices. These changes in the methodology led to delays in releasing results and disbursing incentive payments.

Omitting from the treatment group a good many IAH patients--enrollees whose IAH qualifying characteristics could not be confirmed in administrative data--was another mutual concern of both CMS and the practices. Those exclusions were warranted by the scientific needs of the evaluation to create an appropriately matched comparison group, as the only common source of data to confirm IAH eligibility for all beneficiaries in all samples was administrative records. Excluding this portion of the treatment group resulted from the decision to offer a comparison-group-based methodology as an alternative to the actuarial approach. Alternative methods that preserve as much of the practices' patient base as possible for use in calculations are desirable, if they could be found. This need is particularly important in view of the limited patient panel sizes common in HBPC. As noted earlier, small samples make results subject to instability. Experience with the evolution of the IAH incentive payment methodology also demonstrated that the high mortality found in the target population specified in statute can affect comparability of expense estimates for treated and comparison samples. The extraordinarily high expenses that may attend some deaths could bias the results, even when matching appears to be effective, because there is a large random element determining the timing of death. Incentive payment methods for high-mortality populations will need more development to overcome this kind of problem.

---

[47] The IAH statute requires the Secretary to ensure that no enrolled IAH beneficiary is participating in the Medicare Shared Savings Program (defined in Section 1899 of the Social Security Act). In addition, CMS implemented a policy for the IAH demonstration to avoid simultaneous enrollment or attribution of an IAH beneficiary to other Medicare shared savings initiatives.

[48] Individual practices had between 1,800 and 14,000 beneficiary months that, when multiplied by the difference between the uncorrected and corrected savings estimate, could amount to a substantial overpayment.

[49] Based on the 15 practices that completed Year 2.

Experience also showed that the sites did not enroll each and every patient who met the IAH criteria. Clinical decisions and administrative processes together contributed to the absence of some patients from the enrollment lists. This meant that we could not adequately identify patients for the comparison group similar to the patients that the practices actually enrolled, notwithstanding that the matching results appeared to be excellent. In the third methodology, the use of a savings estimate from the expanded sample to adjust the enrolled-sample estimate was intended to compensate for the inability to replicate enrollment procedures at the practices in the comparison group extraction process. This approach, however, raised an issue with some practices; they were concerned that some of the IAH patients attributed to their incentive payments sample were inappropriate as demonstration beneficiaries. For example, some practices took on the care of new patients at home temporarily after hospital discharge, and they did not consider such patients to be part of their IAH practice.

Even if equivalency could be attained between enrollment processes at the practices and comparison group extraction procedures, challenges to achieving comparable groups for IAH beneficiaries are likely to remain. Administrative data have limitations for identifying comparison beneficiaries who are at the same stage in their illness and have non-health-related characteristics similar to the IAH beneficiaries. Some potentially important data may be missing, such as characteristics leading to the decision to choose HBPC; the same characteristics could cause systematically lower costs for reasons having nothing to do with the delivery model itself. For example, many HBPC patients may decide that receiving primary care at home is the best option because they prefer conservative medical management. Additionally, important pieces of data—particularly the ADL limitations—are difficult to reconcile between administrative data and clinical-record data as found in the practices. One problem is that ADL status may shift in individuals over time, and it is impossible to equalize the timing of measurement between the two sources of ADL information. Another is that standardization of ADL measurement may be difficult to achieve across the two sources.

Finally, while each step in developing the incentive payment methodology used for Year 2 was important and increased our confidence in the accuracy of the result, a seeming volatility in results came with the sequence of methodology changes and the correction of results after Year 2 disbursements. From a business perspective, practices have reported that they strongly prefer predictable benchmarks to assist with revenue planning.

### IAH incentive payment results

Table B2 shows the incentive payments results for the years available at this writing, Years 1 and 2, listing only the 15 practices that completed Years 1 and 2. Nine practices earned incentive payments in Year 1 (two methodologies in use) and seven earned incentives in Year 2 (three methodologies in use). In Year 1, one practice used the actuarial methodology and all other practices used the Year 1 regression methodology. In Year 2, the same practice that used the actuarial methodology continued with it in Year 2. Three practices used the Year 1 regression methodology, and the remaining 11 practices used the Year 2 revised regression methodology.

Under all methodologies, a test for a minimum savings achieved was applied to the estimated savings amount, after the savings were converted to a savings rate (i.e., savings amount divided by target spending amount). A minimum savings requirement (MSR) was used in the payment methodology to comply with the authorizing legislation. The law required that the spending targets include a risk corridor that takes into account normal variation in expenditures (Social Security Act 1866E(c)(1)). The purpose of this provision was to avoid payments based upon savings occurring by statistical chance. Therefore, to determine the incentive payment, CMS tested each practice's savings rate—as long as the estimate indicated savings--to ensure it met or exceeded the MSR. To determine each practice's MSR, 1,000 samples, each consisting of the same number of IAH beneficiaries as the practice enrolled, were drawn from a matched comparison group of beneficiaries who met the eligibility criteria for the IAH demonstration but did not receive primary care in the home. For estimating the average savings rate of the sample, CMS calculated total Medicare Part A and Part B expenditures and a spending target for each member of this matched comparison group, using the original "actuarial" approach. The average savings rates for each of the 1,000 samples were ranked in ascending

order. The MSR was equal to the 90th or 95th percentile of the 1,000 simulations. Practices with average savings rates that equaled or exceeded the 95th percentile in the comparison sample were eligible to receive up to 80 percent (based on quality performance) of savings remaining after CMS retained the first five percent, while practices with average savings rates that equaled or exceeded the 90th percentile were eligible to receive up to 50 percent.

Table B3 shows examples of the two confidence criteria for the minimum savings requirement, which depended on the number of IAH beneficiaries in the practice. For example, in a practice with 200 IAH beneficiaries, if the calculated savings percent was at least 14.1 percent, then up to 80 percent of the savings remaining after CMS retained the first five percent was available as an incentive payment. However, if the calculated savings percent was at least 11.4 percent and below 14.1 percent, then the shareable savings could be no larger than 50 percent of the savings remaining after CMS retained the first five percent. In both alternatives, the incentive payment could be adjusted downward further, depending on the number of quality measures achieved.

### Worked example of incentive payment calculation

To explain the calculations determining the incentive payment, we use Long Island, NY, as an example. The Long Island IAH practice was paid under the original actuarial method, which used the risk-based formula shown earlier to determine the target amount. All other practices chose the Year 1 comparison-group-based method for Year 1. Those practices' respective target and actual amounts for Year 1 were determined from analysis of treatment and comparison samples before we developed the revisions to the comparison-group-method used by most practices for Year 2. Given (1) target and actual amounts determined according to the practice's chosen method, (2) results of the test for the minimum savings requirement (described below), and (3) the number of beneficiary months recorded by the implementation contractor (shown in Table B2), all additional calculations were carried out as explained below, regardless of performance year and methodology for determining the target (comparison-group or actuarial). Note that results reflect rounded values.

Long Island's actuarial-method target in Year 1 was $3,547 PBPM. The practice's actual average monthly expense, $3,024, was $524 (after rounding), or 14.8 percent, lower. The savings rate of 14.8 percent met the minimum savings requirement (MSR) at the 95 percent confidence level (Table B3[50]). The practice's beneficiary months in Year 1 totaled 2,395, implying a gross savings of $1,254,946 ($524 times 2,395 ≈ $1,254,946). Practices could share in the savings by keeping up to 80 percent of the savings beyond the first 5 percent saved; CMS retained the first 5 percent saved, which was calculated as 5 percent multiplied by the spending target times the total beneficiary months. In the case of the Long Island practice, $424,859 was the first five percent retained by CMS (0.05 times $3,547 times 2,395 = $424,859). Long Island's remaining savings were $830,086 ($1,254,946 - $424,859 = $830,086), and the maximum available payment (80 percent of that amount) was $664,069 (0.80 times $830,086 = $664,069). Because Long Island achieved 5 of the 6 quality measures tied to incentive payments (see Table B1), it was not entitled to the entire available maximum. Instead, it was entitled to a payment of 83.3 percent of the available maximum, or $553,391 (0.833 times $664,069=$553,391). After the sequestration discount of two percent, Long Island's final incentive payment was $542,323.

The Long Island practice's maximum available payment was 80 percent of the savings remaining after CMS retained the first five percent because the MSR confidence level for the practice was 95 percent. In the case of Long Island, its savings rate of 14.8 percent exceeded the required savings rate at the 95 percent confidence level, given its practice size.

[50]Table B3 was not necessarily used in this case. MSRs depend on the exact sample size of the practice.

**Table B2. Incentive payment results using the methodology chosen by each practice: Years 1 and 2**

| Practice Name | Year 1 Spending Target PBPM | Year 1 Actual Expenses PBPM | Year 1 Number of Beneficiary Months | Year 1 Incentive Payment* | Year 2 Spending Target PBPM | Year 2 Actual Expenses PBPM | Year 2 Number of Beneficiary Months** | Year 2 Incentive Payment* |
|---|---|---|---|---|---|---|---|---|
| Austin, TX | $5,210 | $5,384 | 4,750 | $0 | $4,128 | $4,698 | 4,578 | $0 |
| Boston, MA | $4,781 | $4,741 | 1,309 | $0 | $3,862 | $3,862 | 1,719 | $0 |
| Brooklyn, NY | $5,756 | $5,547 | 2,819 | $0 | $4,747 | $4,610 | 3,927 | $0 |
| Cleveland, OH | $4,778 | $4,434 | 1,886 | $0 | $3,558 | $3,574 | 2,727 | $0 |
| Dallas, TX | $4,857 | $4,088 | 6,281 | $1,727,392 | $4,266 | $3,940 | 7,586 | $446,872 |
| Durham, NC | $3,638 | $3,415 | 8,548 | $275,427 | $3,094 | $2,787 | 11,236 | $1,341,649 |
| Flint, MI | $5,471 | $4,404 | 7,029 | $2,915,062 | $4,204 | $4,119 | 8,267 | $0 |
| Jacksonville, FL | $4,673 | $4,213 | 6,014 | $711,527 | $3,647 | $3,645 | 7,126 | $0 |
| Lansing, MI | $4,886 | $4,134 | 3,840 | $1,018,857 | $4,094 | $3,757 | 5,001 | $345,795 |
| Louisville, KY | $4,477 | $4,753 | 8,628 | $0 | $3,986 | $3,930 | 13,013 | $0 |
| Mid-Atlantic Consortium | $5,076 | $4,060 | 3,021 | $1,805,208 | $4,066 | $3,580 | 3,844 | $851,948 |
| Milwaukee, WI | $3,953 | $3,059 | 3,967 | $1,443,964 | $3,305 | $2,983 | 6,344 | $519,772 |
| North Shore, NY | $3,547 | $3,024 | 2,395 | $542,323 | $3,276 | $2,708 | 2,753 | $874,151 |
| Portland, OR | $3,568 | $2,434 | 1,639 | $1,228,263 | $3,018 | $2,298 | 2,112 | $942,156 |
| Wilmington, DE | $5,192 | $5,421 | 1,757 | $0 | $3,912 | $4,454 | 3,143 | $0 |

Source: Demonstration implementation contractor.

*Incentive payment amount, after test for minimum savings requirement, quality measure adjustment, and sequestration. **Months adjusted to stay within 10,000 beneficiary cap.

Notes: Dallas, Lansing, and Mid-Atlantic Consortium used the Year 1 regression methodology for Year 2; target and expense for Year 2 come from Year 1 methodology. North Shore used the actuarial methodology in both years; targets and expenses come from actuarial methodology in both years. Among practices with calculated savings, Boston, Cleveland, Brooklyn, and Jacksonville did not meet MSR in Year 1, and Brooklyn, Flint, Jacksonville, and Louisville did not meet MSR in Year 2.

## Table B3. Example of minimum savings requirements

| Practice Size | Minimum Savings Requirement | |
|---|---|---|
| | 95% confidence level (if met, up to 80% of savings shared by the practice, after allocation of first five percent* of savings to CMS) | 90% confidence level (if met, up to 50% of savings shared by the practice, after allocation of first five percent* of savings to CMS) |
| 200 | 14.1% | 11.4% |
| 400 | 8.0% | 6.0% |
| 500 | 9.3% | 7.5% |
| 800 | 7.4% | 5.7% |
| 1000 | 6.4% | 5.0% |
| 2000 | 4.4% | 3.5% |

*CMS retains the first five percent saved, which is equal to .05 times the product of the monthly spending target and total beneficiary months.

# Appendix C: Evaluation Design

The evaluation of the IAH demonstration employs mixed methods, combining analysis of quantitative and qualitative data to assess the impact of the demonstration incentive structure on the process of care, health care utilization, and expenses. The incentive structure of the demonstration was the opportunity for practices to earn incentive payments adjusted for performance on quality of care. In establishing the incentive structure, the demonstration introduced an additional source of revenue for participating practices, and the evaluation is concerned with impacts of being exposed to the changes.

The main issue for the evaluation is not the effects of home-based primary care (HBPC) itself, because data are insufficient to address the question. In this appendix, we explain why and provide additional information on the methodologies underlying the results in this Report to Congress. We first summarize the evaluation research questions. We then describe the basic design used with each source of quantitative data--both Medicare administrative data and survey data--and qualitative data. Finally, we describe the sample selection methods for each source of data.

### Research questions

The overarching research question addressed in this report is whether the demonstration reduced expenditures and improved health outcomes for IAH beneficiaries. The question is further elaborated in the following research questions.

- Did the demonstration, in which practices were offered payment incentives tied to quality measures, lead to savings for Medicare?
- Did the demonstration lead to changes in health care utilization, such as reductions in acute-care stays and emergency department use?
- Did the demonstration lead to changes in health outcomes?
- Were the IAH beneficiaries satisfied with their primary care?
- Did the IAH practices change their approach to care delivery in response to the demonstration and, if so, how did they change?

The evaluation used quantitative data to answer the questions on cost, use, and health outcomes. Quantitative data sources consisted of Medicare administrative data (e.g., claims, Medicare enrollment files, routine assessments such as the Minimum Data Set [MDS] used in nursing facilities). Surveys were another source of quantitative data.

The evaluation also sought to learn whether any changes occurred during the demonstration that would suggest impacts on the following key issues of process-of-care quality:

- effective, efficient, and timely health services;
- care coordination; and
- patient and caregiver satisfaction with quality of care.

The evaluation relied primarily on qualitative data to address process-of-care effects. Qualitative data consisted of information collected during site visits with IAH practices and interviews with practice representatives.

Two important caveats about this report bear emphasizing. First, the report addresses the impact of the IAH demonstration on the outcomes of patients from the specific set of practices that participated. The information generated by our sample cannot be projected nationally. The practices were not a random draw of practices that specialize in HBPC. Instead, the IAH participants consisted of a relatively small group (18 practices at the start of the demonstration, including consortia, each considered a single practice for demonstration purposes) selected from more than one-hundred applicants. Chosen to ensure that a broad range of experienced practitioner groups would be included in the test of the demonstration program, the practices' representativeness of the national population of practices is not well understood. In short, no conclusions about the impact of the demonstration in a national program can be drawn from the demonstration.

Second, as noted earlier, the main focus of this report is not the question, what are the effects of HBPC in general? On that question, a survey of IAH enrollees contributed some limited information, primarily relating to care process impacts. However, later in this appendix we explain why conclusions from the survey are tentative. While the question of expenditures, health outcomes, and quality of care under HBPC is of great importance, data are unavailable to support a strong analytical approach for answering it. We explain the reasons in the next section, which is devoted to the basic comparison strategies used for the evaluation.

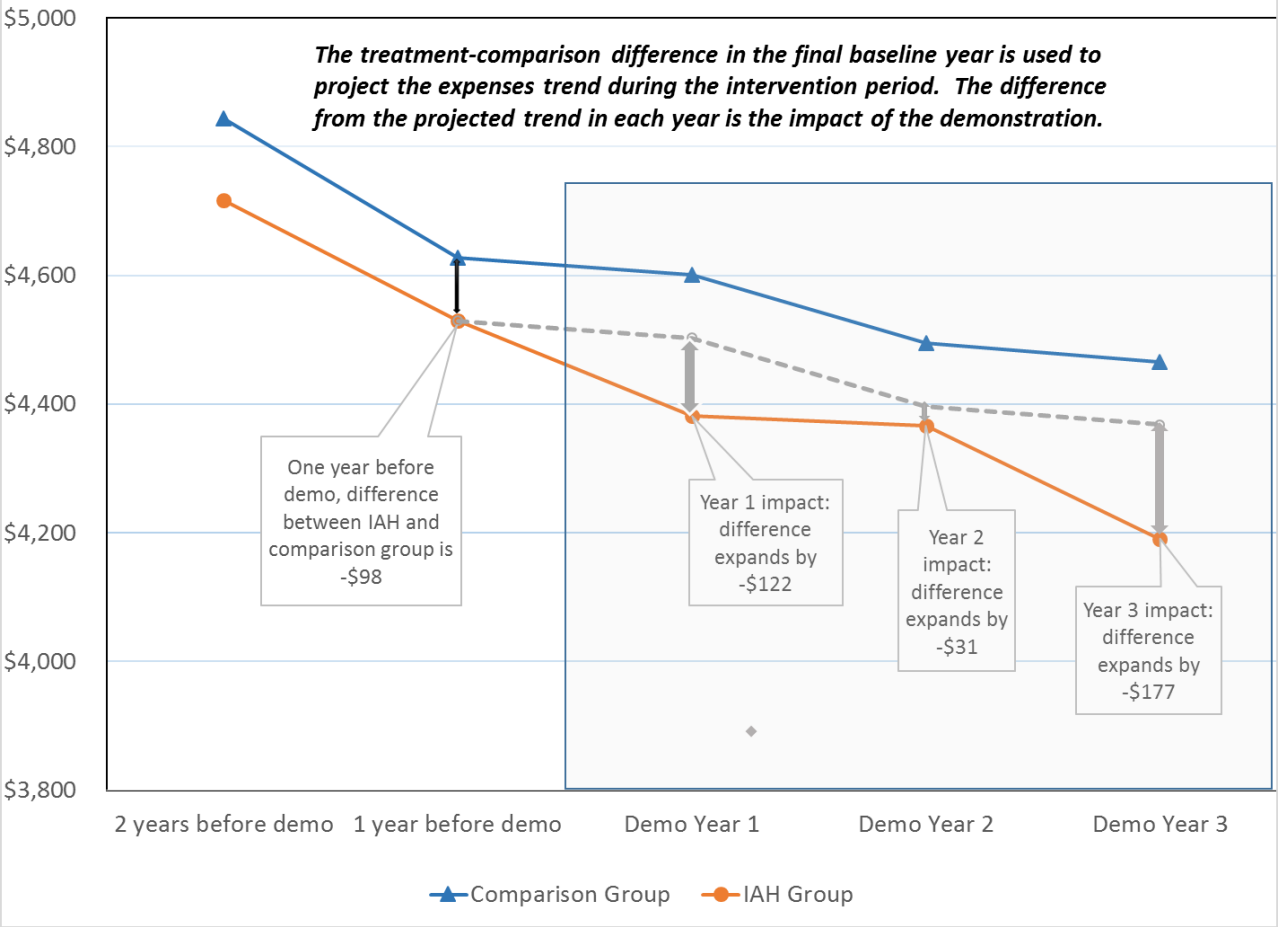
## Research designs for quantitative and qualitative analyses

The IAH evaluation uses three basic designs, one for each source of data—administrative, survey, and qualitative data. It is important to understand the comparisons being made in each basic design for making inferences about the demonstration. In this section of Appendix C, we explain each comparison strategy and key assumptions required for accepting comparisons as valid. Limitations of comparisons related to the survey and qualitative data are also explained.

**Research design for quantitative analysis of Medicare administrative data.** For analyzing the impact of the demonstration using Medicare administrative data, we employed a strategy of comparing the expense and utilization averages of the IAH practices' patients with averages for a matched comparison group across time, i.e., the year before the demonstration began versus the three-year period after the demonstration began. In this design we can interpret a change in the groups' difference relative to a baseline-year difference as the demonstration's impact (see Figure C1). This comparison strategy excludes from the impact any change due to factors outside of the demonstration, change which is assumed to be accounted for by the trend of the

comparison group.[51,52] Regression analysis was used to control for any group differences in observable variables that remained after matching comparison beneficiaries to treatment beneficiaries.

## Figure C1: Comparison strategy




[51] With this strategy, we assume that the patient mix of the practices doesn't change over time in ways we cannot measure and account for. We found similarity in patient profiles over time when examining various health conditions, demographics, and other variables derived from administrative data. However, the assumption that no important changes occurred in unmeasured characteristics of both and comparison groups is impossible to assess.

[52] Mathematically, the impact estimate is equivalent to comparing the before/after difference in the treatment group to the before/after difference in the comparison group. The before/after design with a comparison group is a difference-in-difference analysis. Technically, we used the last year before the demonstration to measure the groups' average difference in the "before" period, and we generated the Year 1 impact by testing the statistical significance of the change in the average difference in Year 1. We used the same "before" period as the baseline when generating impact for Years 2 and Year 3. We also generated the average difference in the entire three-year "after" period with respect to the last year before the demonstration. We also tested for sensitivity to the choice of baseline year by generating a two-year average baseline; our results for most outcomes were not sensitive to whether we used a two-year average baseline or the last year before the demonstration as the baseline.

In carrying out this analysis, an important step was finding beneficiaries whose information in CMS administrative data showed they had characteristics similar to the treatment group. Administrative data include claims histories, assessments such as MDS records submitted by nursing facilities, as well as enrollment information for demographics and Medicare entitlement. Variables from these sources indicate health conditions, functional status, and demographic characteristics that can predict Medicare expenses and other evaluation outcomes. The information was used to select comparison beneficiaries who closely resembled IAH beneficiaries, with the important difference that comparison beneficiaries did not use HBPC. As shown in Figure C1, the comparison group's function was to indicate how expenses and outcomes would have evolved in the absence of the demonstration. That key assumption—that the comparison group indicates the trend in the outcome during the study period—is tested by examining the two groups' trends in the baseline period for parallelism. Evidence of parallelism supports the assumption that the parallel trends would have continued in the absence of the demonstration, though the assumption is fundamentally untestable. Figure C1 indicates that the expenses trend for the IAH beneficiaries was parallel to the expenses trend for the comparison beneficiaries. We checked for parallelism in the pre-demonstration years and, for most outcomes, found evidence to support it.

**Inference problems with the question of HBPC impact.** Using the demonstration to draw conclusions about the effects of HBPC would have risked drawing the wrong conclusion. By design, the matched comparisons did not use HBPC. However, a comparison of the average expenses and utilization of these two groups, even after statistically adjusting for any differences that remain after matching, has limitations for inferring an effect of HBPC per se.[53]

If one wished to study the effectiveness of HBPC per se in achieving savings, the causal factor of interest would be the initiation of HBPC. One would want to observe a change between two time points—before and after treatment initiation. Observation of before/after change was not possible for many IAH beneficiaries, because longstanding patients of the IAH practices were permitted to participate in the demonstration. Thus, for these patients, at the time they were selected into the sample and matched to comparison beneficiaries, we could not ensure that their characteristics (observable and unobservable) were unaffected by the treatment.[54] This situation implies that a before/after comparison of the individual's outcomes might not accurately reflect the treatment's impact.

Furthermore, if we were to compare the IAH and comparison beneficiaries' average Medicare expenses during the demonstration period, we would have to assume that the available matching variables are sufficient to control for all relevant differences between the groups. The problem with this comparison is that, in our observational setting, patients may choose to enter

[53] In response to concerns raised about the original actuarial methodology, for purposes of estimating the expenditure target in the second and third incentive payment methodologies, CMS determined the incentive payment from a comparison of average expenses of IAH beneficiaries and matched beneficiaries who did not use primary care at home. See Appendix B for further description of incentive payment analysis in the demonstration.
[54] The demonstration was too small to identify and match sufficient numbers of beneficiaries before they entered HBPC and follow the two groups over time.

HBPC based in part on unobserved characteristics that may be associated with outcomes. Unobserved characteristics, including motivation-related ones, can be difficult to measure. For example, for an outcome measure such as use of specialists, many patients who chose HBPC may have decided that involving numerous specialists in their care is not worth the effort, given their frailty and limited mobility. And conceivably practices or their referral sources might also be selective based on characteristics we cannot measure. Without the ability to account for unobserved characteristics as in a randomized design, analysis of specialist use might erroneously credit the treatment with effects from missing motivational factors, and possibly produce misleading results.

Therefore, we concluded that comparing treatment and comparison group averages in this report is not appropriate to identify effects of HBPC. Instead, we implemented the comparison strategy described earlier in this section, which allowed us to study the policy intervention described in the law, i.e., the payment incentive structure. A key assumption of the policy intervention analysis is that unobservable characteristics influencing selection into HBPC do not change between the period before payment incentives were available and the three-year period after the demonstration began. We consider this assumption realistic over the relatively short period of the study.

**Research design for quantitative analysis of survey data.** In the evaluation of the IAH demonstration we examined beneficiary survey data to draw tentative conclusions about quality of care associated with HBPC. A one-time survey of IAH enrolled beneficiaries was conducted during 2013-2015. We employed two approaches in the quantitative analysis of the IAH beneficiary survey data, one of which used a comparison strategy while the other did not. In the comparative approach, we identified a group of respondents to the 2013 Medicare Current Beneficiary Survey (MCBS) who approximately met the IAH qualifying criteria but did not use home visits. The comparison strategy was to compare survey responses of IAH beneficiaries and the MCBS beneficiaries on the same questions. Questions were drawn from the MCBS questionnaire and dealt with the beneficiary's care-process experiences and self-reported health status. Comparisons of responses for IAH and MCBS groups were adjusted using characteristics available in the survey files (for survey respondents) and the Medicare administrative data (IAH respondents). Unlike the quantitative analysis of administrative data, we did not use matching techniques to identify comparisons.

The design of the survey analysis is to compare responses obtained on a single occasion of measurement. The main limitation of comparing responses from a single-occasion measurement is the groups were not assessed for their comparability before the demonstration took place; data for a pre-assessment were not available. Thus, we lack evidence to rule out pre-existing group differences as a reason for differences we may find in beneficiaries' reports of care quality. Because of this limitation, we interpret the results conservatively as suggestive of care quality differences that may exist between HBPC and usual care.

The second approach to the quantitative analysis of survey data was a descriptive analysis of IAH beneficiaries' responses to our survey questions created specifically for the evaluation. The

questions examined dealt with overall satisfaction with HBPC and benefits of HBPC delivery. Despite the lack of a comparison group, the analysis provided some insight into IAH beneficiaries' perspectives on HBPC, and we believe the information is worth reviewing.

**Research design for analysis of qualitative data.** We used qualitative information to understand how practices may have responded to the demonstration. Qualitative data sources consisted of site visits and telephone interviews. The evaluation conducted an initial round of site visits to the IAH practices in Year 1 (February 2013 to May 2013), a second round of visits in Year 2 (February 2014 to July 2014), and a third round in Year 3 (April 2015 to October 2015). During these visits, we interviewed staff members as well as practice leaders. The focus of data collection was how the practices were delivering care, including changes between years, as well as barriers and facilitators to meeting the quality metrics of the demonstration (Figure 2). In February and March 2017, we conducted telephone interviews with practice representatives to update information about practice characteristics for purposes of this report (see Table 3 and Appendix D). Finally, we conducted interviews with care partners of the practices in late 2016 and early 2017. Care partners are individuals and organizations involved in the care of beneficiaries but are not part of the IAH practices. We collected data from care partners to gain further information about how IAH practices coordinate care with them, as well as their perceptions of changes the IAH practices undertook during the demonstration.

We employed qualitative-data analytic software to support analysis of the information we gathered. Consistent with the topics of the interview questions, we focused on identifying key barriers and facilitators to mounting practice-driven strategic and operational changes focused around demonstration incentives. We also captured other implementation-related themes and issues such as characteristics of the external environment. We then identified common themes across sites. After identifying themes, we took another look at the data to review any cases that didn't fit the themes to identify any other issues that might be noteworthy.

Our analysis of qualitative data entails a description of what happened during the demonstration. A limitation of the design used with qualitative data is that it does not include comparative information from primary care practices not exposed to the demonstration incentive structure. Therefore, we cannot be certain whether changes in IAH practices' operations or structure occurred because of the demonstration or because of influences operating more broadly during the same period as the demonstration. We look for consistency between our understanding of what happened and the changes indicated by the analysis of quantitative data to provide mutually supportive information between the two sources. A separate caveat related to the care partner interviews stems from the approach to sample selection. We found care partners by asking IAH practices to provide contact information for several partners. Using the sites themselves to identify care partners was an efficient way to find care partners but it may not have resulted in a representative sample of IAH practices' care partners. A representative sample of care partners might have returned different responses to our questions.

## Samples used in the evaluation

In this section, we provide details about the way samples were identified to support comparisons made in each research design. The evaluation used different samples for the different purposes of (1) quantitative analysis of administrative data, (2) quantitative analysis of survey data, and (3) care partner interviews. First, we introduce the samples drawn from administrative data for purposes of the quantitative analysis of the impact of the demonstration. We provide details about the matching procedure used to select a comparison group. Then we explain reasons for adding beneficiaries to the treatment sample beyond demonstration enrollees, and reasons for making exclusions from the enrollee lists necessitated by design and measurement considerations. Although these sample modifications led to limitations on generalizability of the results relative to the demonstration's intended target population, they were important to the scientific soundness of the conclusions. Next, we describe the survey samples, which were used to provide information about possible differences in the care experiences between IAH beneficiaries and Medicare beneficiaries similar to IAH beneficiaries who did not use HBPC. We conclude with a description of the approach to sampling in the care partner interviews, which were used to add to our understanding of how IAH practices coordinate care and any changes they made during the demonstration.

### Samples used for quantitative analysis of Medicare administrative data

Without exception, the sample of treatment beneficiaries came from the patients treated by each IAH practice, as identified in claims. We also identified a separate comparison sample for each practice by extracting person-level information from Medicare administrative data for beneficiaries in the practice's geographic area. After the practice-specific samples were finalized, we combined them for use in a single impact analysis.

**Treatment sample.** The treatment beneficiaries were only a subset of each IAH practice's patients, because the legislation mandated only certain patient characteristics could qualify beneficiaries for the demonstration. We created a set of rules to use with administrative data to test beneficiaries against the qualifying criteria and to attribute beneficiaries to an IAH practice. The rules implementing the qualifying criteria were that the beneficiary had to be enrolled in fee-for-service (FFS) Medicare; require human assistance with two or more activities of daily living (ADLs) as recorded in Medicare administrative data; have two or more chronic conditions; and have had an acute hospital stay and used rehabilitation services[55] in the previous 12 months. In addition, the beneficiary could not have been in hospice or a long-term nursing facility for the duration of his or her eligibility for the demonstration. Therefore, quantitative analysis of Medicare administrative data to evaluate the demonstration's impact is based on beneficiaries who were:

- seen by the practices for HBPC, and
- met the IAH beneficiary requirements of the law in Medicare administrative data, and

[55] IAH-qualifying rehabilitation services consisted of a stay in either a skilled nursing facility, inpatient rehabilitation hospital, or long-term-care hospital, or utilization of a home health agency.

- met those IAH requirements in administrative data in each year of the demonstration.

*We did not limit the treatment sample to beneficiaries officially enrolled in IAH by the practices.* The sample included all beneficiaries treated by the practices at home, as long as they met the IAH qualifying criteria every year. Our experience implementing the incentive payment methodology indicated likely omissions from the enrollment lists that could introduce error into the evaluation analysis. (See Appendix B for more information about how our experience resulted in revisions to incentive payment methods in Year 2.)

**Comparison sample.** The initial steps in sample construction did not attribute beneficiaries to either the treatment or comparison group. We first identified all beneficiaries in the IAH practices' geographic areas who met the IAH eligibility criteria. The purpose of using a geographic restriction was to control for locale-related practice patterns, access conditions, and cost levels. We then performed the practice attribution. After identifying the IAH practices' beneficiaries, we dropped from the remaining pool of qualifying beneficiaries the small number who used more than one home visit during the year, which left us with a pool of potential comparisons representing "usual care" received from other practices. Using propensity score matching, we selected from the potential comparisons pool persons who very closely resembled IAH beneficiaries, as indicated by traditional diagnostics for matching.

**Sample sizes and observation periods.** Table C1 shows the final matched sample numbers in each year of the evaluation impact analysis, beginning with two years before the demonstration. We matched up to five comparison beneficiaries for each treatment beneficiary. The individual practices had varied sample sizes, ranging in Year 3 from 138 (Portland) to 1,267 (Durham) (data not shown).

From claims histories and Medicare enrollment data, we established the date by which each beneficiary met all the IAH eligibility requirements, using the same procedures for treatment and comparison beneficiaries. We began observing outcomes with the first month after eligibility was established[56]; all subsequent months of the year were used for outcome measurement, provided that the beneficiary remained alive and stayed enrolled in Medicare Parts A and B. Table C1 shows a slightly larger average number of eligibility months, i.e., months observed, for the treatment group relative to the comparison group each year. Longer average eligibility periods for IAH beneficiaries are due to differences in mortality rates and the fact that relatively more IAH beneficiaries than comparison beneficiaries were eligible before the start of the performance year. We controlled for differences in eligibility months in the matching procedure and in the analysis. The outcome variables in the analysis each year were computed as the beneficiary's average over the eligibility months. We weighted the analysis in

[56] Since IAH qualifying criteria involved looking back up to 12 months for acute hospital and rehabilitation utilization, some beneficiaries qualified prior to the performance year. In those cases, we began observing outcomes from the first month of the performance year.

accordance with each beneficiary's number of eligibility months in the year.[57] Weighting allows beneficiaries under observation longer to have proportionate influence on the analysis

[57] Measures for annual outcomes, such as number of acute-care stays per year, were annualized.

## Table C1. Evaluation analysis sample, by year

| | Two years before demonstration | One year before demonstration | First demonstration year | Second demonstration year | Third demonstration year |
|---|---|---|---|---|---|
| Number of confirmed eligible* beneficiaries served by IAH sites | 6,837 | 7,367 | 8,216 | 7,266 | 7,564 |
| Total months of IAH eligibility** for treatment group | 65,781 | 70,591 | 79,396 | 69,768 | 72,215 |
| Average months of IAH eligibility* per treatment beneficiary | 9.6 | 9.6 | 9.7 | 9.6 | 9.5 |
| Number of matched comparison beneficiaries | 29,517 | 31,888 | 33,916 | 32,248 | 31,259 |
| Total months of IAH eligibility* for comparison beneficiaries | 264,588 | 286,314 | 303,770 | 293,081 | 278,015 |
| Average months of IAH eligibility* per comparison beneficiary | 9.0 | 9.0 | 9.0 | 9.1 | 8.9 |

Source: Medicare claims and enrollment data for 2010–2015

*For each beneficiary, we established the date by which all IAH eligibility criteria were met in administrative data.

**The months of IAH eligibility were counted from the month after eligibility was established until the end of the performance year, as long as the beneficiary remained alive and enrolled in Parts A and B. Analysis was based on expenses and utilization incurred during all months of IAH eligibility in the performance year.

**Propensity score matching variables.** The comparison-group sample sizes shown in Table C1 were achieved after we found good matches for treatment beneficiaries. The propensity score matching procedures relied on numerous characteristics variables to implement matching, with the goal of attaining the same or nearly the same percent of treatment and comparison beneficiaries with each characteristic. Table C2 lists the matching variables used in the propensity score procedure for finding good comparisons for IAH beneficiaries from the pool of potential matches. For each IAH practice annually, including each member of the Mid-Atlantic consortium individually, we implemented a separate propensity score procedure. We then examined the results to check for similarity of the two groups for each practice by year.[58] Our examination showed that the groups resulting from the matching procedure were very similar, at least as indicated by the characteristics we could observe.

Matching variables were categorized into four groups: prior healthcare utilization, demographics, functional status (activities of daily living limitations), and health status indicators. The prior healthcare utilization variables included one that was matched exactly: the number of months since the beneficiary's last inpatient admission prior to the date on which the beneficiary met all IAH eligibility criteria (one, two to three, or four or more months). This variable was important to model the expense trajectory that tends to follow acute care; beneficiaries with more-recent acute care would tend to have a monthly average that reflects a period of high expenses associated with the aftermath of acute care. Exact matching meant that the matched comparison for a treatment beneficiary had the same value for this variable,

[58] Beneficiaries of the three practices in the consortium participant, Philadelphia, Washington, and Richmond, were matched in separate matching procedures.

and the proportion of the sample with each value would be exactly equal for the treatment and comparison samples. For all other variables, the propensity score matching procedure resulted in a very similar proportion of each sample having the characteristics used in matching, without the need for an exact match.

## Table C2. Matching variables used to identify comparison beneficiaries in Medicare administrative data

| |
|---|
| **Eligibility and utilization** |
| Number of months since last inpatient admission (one, two to three, four or more)[a] |
| Month of the demonstration year that beneficiary met eligibility criteria (1, 2–6, 7–12)[b] |
| Whether beneficiary had an observation stay and no inpatient admission 12 months prior to IAH eligibility date |
| **Demographic characteristics** |
| Age: younger than 65, 65–79, 80 or older |
| Gender |
| Race: white, black or African American, other or unknown |
| Dually eligible for Medicare and Medicaid |
| Original reason for Medicare entitlement: old age, ESRD or ESRD and disability, disability only |
| **Activities of Daily Living (ADL) limitations** |
| Number of ADLs for which beneficiary requires human assistance: two, three to four, five to six |
| Whether information about the feeding ADL was missing[c] |
| **Health status** |
| HCC risk score |
| Specific HCCs |
| HCC8: Metastatic cancer and acute leukemia |
| HCC9–10: Lung and other severe cancers; lymphoma and other cancers |
| HCC11–12: Colorectal, bladder, and other cancers; breast, prostate, and other cancers and tumors |
| HCC18: Diabetes with chronic complications |
| HCC21: Protein-calorie malnutrition |
| HCC27: End-stage liver disease |
| HCC28–29: Cirrhosis of liver; chronic hepatitis |
| HCC46: Severe hematological disorders |
| HCC48: Coagulation defects and other specified hematological disorders |
| HCC51: Dementia with complications |
| HCC52: Dementia without complications |
| HCC54–55: Drug/alcohol psychosis; drug/alcohol dependence |
| HCC57–58: Schizophrenia; major depressive, bipolar, and paranoid disorders |
| HCC70–71: Quadriplegia; paraplegia |
| HCC72: Spinal cord disorders/injuries |
| HCC85: Congestive heart failure |
| HCC96: Specified heart arrhythmias |
| HCC103–104: Hemiplegia/hemiparesis; monoplegia, other paralytic syndromes |
| HCC106: Atherosclerosis of the extremities with ulceration or gangrene |
| HCC107–108: Vascular disease with complications; vascular disease |
| HCC111: Chronic obstructive pulmonary disease |

HCC134: Dialysis status

HCC136–138: Chronic kidney disease, stages 3–5

HCC139–140: Chronic kidney disease, stages 1–2 or unspecified; unspecified renal failure

HCC157–159: Pressure ulcer of skin with necrosis through to muscle, tendon, or bone; or with full or partial thickness skin loss

Depression

Anemia[d]

Fluid and electrolyte disorders[d]

Whether beneficiary had a complicating condition or major complicating condition during the most recent inpatient admission[d,e]

Chronically critically ill or medically complex diagnosis[d,f]

Number of chronic conditions recorded in the Chronic Conditions Data Warehouse (2–5, 6–9, 10 or more)

---

[a] The count of months between the date that the beneficiary met all IAH eligibility criteria and the date of the last inpatient admission prior to eligibility. Used as an exact matching variable. All other variables were ordinary matching variables used in the propensity score procedure.

[b] For sites that began the demonstration in June 2012, Month 1 is June. For sites that began the demonstration in September 2012, Month 1 is September.

[c] Feeding assessments are generally not available on home health OASIS assessment data at the time of home health recertification. If the beneficiary had a previous assessment during the study year that was recorded at the time of discharge from home health, we used the feeding values from that assessment; however, sometimes there was no previous discharge assessment.

[d] Measured using the claim from the most recent hospitalization in the 12 months before the beneficiary met IAH eligibility criteria for the year.

[e] Complicating condition or major complicating condition based on the claim's Medicare Severity Diagnosis Related Group (MS-DRG) code.

[f] Chronically critically ill (CCI) or medically complex (MC) according to the diagnosis in the claim; this classification was developed to identify hospitalized patients marked by extended stays of high acuity (CCI) or medical complexity such that acute nursing care is necessary in non-intensive-care beds (MC) (see Kandilov et al. 2014 found at https://innovation.cms.gov/Files/reports/ChronicallyCriticallyIllPopulation-Report.pdf).

ADL = activity of daily living; ESRD = end-stage renal disease.

A second prior utilization measure was the month the beneficiary met eligibility criteria (Month 1 [June or September], 2 to 6, or 7 to 12), which could account for any general influences on outcomes related to the time by which all IAH requirements were met and for differences in length of observation for individuals. The third prior utilization measure was whether the beneficiary had an observation stay but not an acute-care admission in the prior year. An observation stay could meet the IAH hospitalization requirement under the demonstration operational policies, but such stays could differ in acuity from the typical hospital stay, so this variable controlled for that possibility. Aside from the three prior utilization measures, we avoided matching on utilization history variables such as the number of recent inpatient stays prior to the eligibility date. The reason was that many IAH beneficiaries had been under the care of the IAH practice before the IAH eligibility date. The IAH practice could have influenced the utilization history by, for example, reducing the likelihood of acute stays. In that case, a match on the number of acute stays could actually result in matching IAH beneficiaries with comparison beneficiaries who were healthier.

The second category of matching variables, demographic variables, comprised age (younger than 65, 65 to 79, or 80 or older), gender, race (white, black/African American,

other/unknown), whether the beneficiary was dually eligible for Medicare and Medicaid, and the beneficiary's original reason for Medicare eligibility (old age, end-stage renal disease [ESRD] or ESRD and disability, disability only). All variables in this category are predictors of Medicare expenses.

For achieving comparability of the treatment and comparison groups on functional status, we matched on three categories for the number of ADLs (two, three to four, or five to six). We also flagged records that were missing information about whether the person needed human assistance with feeding.[59]

For the final category of matching variables, health status indicators, we used various kinds of diagnosis-related measures. First, we drew from the condition groups defined in the CMS hierarchical condition categories model (CMS-HCC). Medicare uses the CMS-HCC model for risk adjusting Medicare Advantage payments. We measured individual HCCs using diagnostic information from a beneficiary's claims history for the 12 months prior to the date of eligibility for the demonstration in a given performance year. We selected HCCs to use in matching based on their relationship to risk of mortality.[60] Unequal mortality rates in samples resulting from an earlier matching approach suggested a need to control for mortality risk to achieve comparable groups, so we emphasized high-mortality conditions in selecting HCCs as matching variables. Second, we flagged selected diagnosis information from the most recent hospitalization to identify particularly severe acute-care episodes. If the diagnosis-related group (DRG) for that hospitalization indicated a complicating condition or major complicating condition, or if the claim mentioned diagnoses associated with chronically critically ill (CCI) or medically complex (MC) episodes[61], we used that information in matching. We also looked for diagnoses from the last hospitalization indicating anemia or fluid or electrolyte disorders. Finally, we matched on the number of chronic conditions from the Chronic Conditions Data Warehouse. The regression models used for the outcomes analysis controlled for many of these same matching variables to account for the generally small differences between the IAH and comparison groups that remained after matching.

**Beneficiaries added to the samples.** As explained above, our evaluation design for quantitative analysis of administrative data measured baseline-period annual average Medicare expenses and other outcomes to implement the before/after analysis. To promote a valid analysis of the demonstration, we sought to use consistent rules for selecting each of the five annual samples, beginning two years before the demonstration's first year. There was no enrollment possible in the baseline period, which pre-existed the demonstration, but consistency required that we select baseline and demonstration-period beneficiaries using the same eligibility rules and data sources. Table C3 indicates that 3,686, 2,702, and 3,066 non-enrolled beneficiaries were

[59] Some home health assessments are not required to report on feeding limitations.
[60] See Gagne J et al., Journal of Clinical Epidemiology 2011; 64(7):749–759.
[61] Kandilov et al. 2014, "Chronically Critically Ill Population Payment Recommendations (CCIP-PR)." Research Triangle Park, NC: RTI International). This classification was developed to identify hospitalized patients marked by extended stays of high acuity (CCI) or medical complexity such that acute nursing care is necessary in non-intensive-care beds (MC).

included in the evaluation samples in Years 1, 2, and 3, respectively, because they met the eligibility criteria based on the same rules used for selecting observations for the baseline period.[62] (Similarly, comparisons were selected using the same rules.) We included them notwithstanding that the practices did not enroll them. Often the difference in eligibility determination between the practices and the administrative data related to the required ADL limitations. Therefore, one implication of adding these beneficiaries to the evaluation samples is that it is unclear whether the study pertains precisely to the target population that has two or more ADL limitations, or to a broader population. Data sources may provide conflicting information, suggesting that definitive targeting based on ADLs is elusive.

**Beneficiaries excluded from the samples.** Two types of restrictions led to exclusions from the evaluation sample used in quantitative analysis of Medicare administrative data:

1) IAH beneficiaries who qualified for the demonstration (based on administrative data) in their year of entry but did not meet the IAH eligibility requirements in succeeding years; demonstration rules allowed such continuing beneficiaries, or "carryovers," to remain as participants. Although they were excluded from the evaluation sample, carryovers were included in the incentive payment analysis for Year 2 and accounted for 18 percent of the analysis sample. (By definition, carryovers did not exist in Year 1.)
2) Enrolled beneficiaries whose administrative records lacked functional limitations information (i.e., ADL limitations) or otherwise failed to confirm that they met all IAH requirements, such as having two or more chronic conditions. Table C3 shows that the enrollees excluded for not meeting requirements in administrative data accounted for 35 percent, 47 percent, and 51 percent of the annual demonstration enrollment of the 14 study practices in Years 1, 2, and 3, respectively (see Column F, Table C3). After Year 1, carryovers who began the demonstration as enrollees were excluded from the evaluation sample, contributing to the increase in the percent excluded in Years 2 and 3.

[62] Carryovers who began the demonstration as non-enrollees eligible for the demonstration were excluded from the evaluation sample upon becoming carryovers (data not shown).

**Table C3. Relationship between numbers of beneficiaries enrolled* and evaluation sample**

| Demonstration Year | A Eligible** and enrolled | B Eligible** and not enrolled | C Total evaluation treatment group (A + B) | D Not eligible** and enrolled | E Total enrolled* (A + D) | F Excluded enrollees as percent of all enrollees (D/E) |
|---|---|---|---|---|---|---|
| 1 | 4,530 | 3,686 | 8,216 | 2,405 | 6,935 | 35% |
| 2 | 4,564 | 2,702 | 7,266 | 4,059 | 8,623 | 47% |
| 3 | 4,498 | 3,066 | 7,564 | 4,718 | 9,216 | 51% |

Note: Table created from data for the 14 practices in the evaluation study's analysis of quantitative data.
*Total enrolled as reported by evaluation contractor.
**Eligibility as determined from evaluation's analysis of administrative data.

**Rationale for the sample exclusions.** At the same time that we added beneficiaries beyond the official enrolled population, it was necessary to exclude some enrolled beneficiaries. One problem was that demonstration rules in Year 1 allowed only beneficiaries who met the qualifying criteria in their 12-month utilization history to participate but, thereafter, beneficiaries were allowed to continue in the demonstration beyond their first year without meeting the IAH requirements again. These "carryover" beneficiaries were liable to introduce a change in the composition of the sample after Year 1 (exclusion group 1). In the interests of avoiding erroneous findings from shifting to samples that included patients who do not meet the demonstration criteria, we excluded beneficiaries who did not "requalify" in a year after their initial qualifying year. Furthermore, a second problem was that our administrative data sources did not confirm eligibility for all new beneficiaries enrolled by practices in each year. So, to maintain consistency between the IAH beneficiaries and the comparison group, we excluded new entrants whom the practices enrolled each year but who lacked confirmed eligibility (exclusion group 2 above). More information about the two excluded groups follows.

***Carryovers (exclusion group 1).*** IAH beneficiaries who failed to "requalify" were those who, upon our examination of their administrative data, no longer met the IAH eligibility requirements for the year following one in which they did qualify. The demonstration rules allowed for the realistic assumption that not all beneficiaries would incur another hospital stay and another rehabilitation stay[63] after initially qualifying for the demonstration, and that not all beneficiaries would continue to meet the ADL status criteria. These carryover beneficiaries were excluded from the evaluation samples but not the incentive payment analysis samples. In fact, since the demonstration incentive structure was designed to reduce unnecessary utilization and improve health outcomes, excluding from the analysis patients whose experience indicated successful outcomes would not be a fair reflection of performance. (The method for this calculation needed special consideration, which is described in Appendix B.)

However, from an evaluation perspective, retaining this group in the evaluation samples would have meant that we could not implement a rigorous before/after design to analyze the

[63] Rehabilitation stays include stays in skilled nursing facilities and rehabilitation facilities, as well as home health episodes.

incentive structure. Including carryovers would have permitted a change in the treatment group's composition during the "after" period, beginning with Year 2, when carryovers came into existence. A change in the treatment group's composition would arise in Year 2 from adding to the sample of newly qualified and requalified beneficiaries an additional group who did not strictly meet the IAH qualifications any longer. The only feasible treatment and comparison groups in the pre-demonstration baseline consisted of beneficiaries who met the IAH qualifying criteria, as did all the IAH beneficiaries of Year 1, in conformance with first-year rules. Therefore, we restricted the analytic sample to beneficiaries who met the IAH eligibility criteria every year. This approach provides a consistent sample definition each year and avoids confounding the effect of the demonstration incentives with a change in the treated population beginning in Year 2. Furthermore, had we simulated carryovers in the baseline years of the evaluation, in an attempt to construct a treatment group consisting of a mix of new and continuing beneficiaries, it would have contradicted the Year 1 demonstration rule that allowed only newly eligible participants. One implication of excluding carryovers is that the evaluation results apply to frail and chronically ill beneficiaries with a recent history of acute and rehabilitation care and may not pertain to the frail and chronically ill who, for various reasons, avoided acute and rehabilitation care in the recent past.

Notwithstanding that it was necessary to drop carryovers from the evaluation, we sought some understanding of their characteristics and cost trajectory over a two-year period. Therefore, we conducted an analysis of the carryovers by comparing them to their original, Year 1 matched comparisons (i.e., comparisons they had in Year 1, before they were rematched for Year 2 for purposes of incentive payment analysis), and we also compared them to Year 1 IAH beneficiaries who were not carryovers, namely, requalifying beneficiaries, beneficiaries deceased during Year 1, and beneficiaries who exited the analysis sample by Year 2 (largely due to ending their association with the practice).

Our comparison of carryovers and their original matched comparisons showed that in Year 1, carryovers had expenses that were 41 percent lower and, not surprisingly, a mortality rate more than 80 percent lower (2.9 percent vs. 17.3 percent). A large difference was expected because, by definition, carryovers generally did not experience a repetition of the two utilization-related qualifying events for IAH (hospitalization and rehabilitation in the prior 12 months) and they survived into Year 2. In contrast, their original matches may have been hospitalized and they experienced high mortality typical of the IAH study participants. We saw other indications that carryovers were healthier than their original matches. Carryovers had about one-sixth the rate of hospice or long-term-care entry that their matches had, and carryovers had fewer ADL limitations (e.g., 46 percent with five or six ADLs vs. 54 percent) and fewer chronic conditions (7.7 vs. 8.4), although they had the same mean HCC risk score, a slightly higher proportion were age 85 or older, and they were more likely to have dementia.

In general, observing both groups (carryovers and their surviving original matches) over the longer term, we found that the expense differences narrowed, and carryovers in Year 2 were slightly more likely to enter hospice or long-term care than their surviving matches (18 percent vs. 16 percent). By the end of 24 months of followup for every individual, total mortality was 20.8 percent for carryovers and 31.8 percent for their original matched comparisons, signifying a large increase in mortality for carryovers during the latter 12 months of the 24-month

followup period but, overall, still a lower long-term death rate. Another indication of the carryovers' standing: A comparison of health characteristics between carryovers and their surviving matches during Year 2 suggested that carryovers had widened the gap that had favored them in Year 1, in terms of several major chronic diseases such as diabetes, heart disease, and chronic obstructive pulmonary disease.

The comparisons of Year 1 characteristics between carryovers and other groups of beneficiaries in the Year 1 IAH sample suggested carryovers had generally better health status--again, as might be expected, given the definition of carryover status. While they were healthier, they also had the highest dementia rates of any group in this analysis and, except for decedents, they had the largest proportion (39 percent) age 85 and older. Also, a relatively high proportion of carryovers had a long period (at least 7 months) between their last hospital stay and their IAH eligibility date.

In summary, the carryover analysis suggested that carryovers represent a segment of IAH beneficiaries with a tendency to be in comparatively stable health while being relatively old and associated with one marker of advanced age, dementia. The analysis might also suggest that HBPC can defer mortality or long-term-care entry, as might be inferred from our comparison of mortality for carryovers and their Year 1 matched comparisons over a two-year period, and the pattern in rates of LTC and hospice entry between Year 1 and Year 2. However, such a hypothesis should be explored in stronger research designs, e.g., randomized trials.

The carryover analysis also confirmed the need to rematch carryovers for incentive payment calculations, inasmuch as similarity of variables to predict expenses had deteriorated over time. For example, not only did the HCC scores differ by Year 2, but also 7.7 percent of carryovers vs. 19.5 percent of their surviving comparisons had the most recent acute stay two to three months before the start of the second performance year. This important variable, used for modeling the expense trajectory in the succeeding period, would have signaled a significant source of non-comparability by Year 2.

***Enrolled beneficiaries whose administrative records lacked confirmatory information on IAH eligibility (exclusion group 2).*** The demonstration rules allowed practices to enroll beneficiaries based on their own assessment of the chronic conditions and functional limitations requirements. With no clinician assessments available before the demonstration or for comparison group members, the only source of data all had in common was administrative data. Administrative data contain chronic condition diagnoses from claims and mandatory assessments of functional limitations for skilled nursing stays, home health stays, and inpatient rehabilitation hospitals. A disadvantage of administrative assessments was that their timing depended on when the beneficiary used services that generate assessments, and functional status could change by the time observation for the study began. Therefore, some assessments could have reflected a time when the beneficiary's disablement differed from the one that existed for the performance period.[64]

We conducted some analyses to understand the implications had we included in the sample all enrolled beneficiaries without confirmed eligibility in administrative data. We examined

[64] We selected assessments whose date best coincided with the performance period in question (Years 1, 2, or 3).

available administrative data and found that beneficiaries in the second exclusion group appeared to be generally healthier than the beneficiaries we did include. (See Figure C2 for illustrative information about this group of excluded beneficiaries.) These findings implied that adding beneficiaries who lacked confirmatory information to the evaluation sample might have produced misleading results; any decrease in expenditures could have been due to including

## Figure C2: Enrolled beneficiaries with unconfirmed eligibility in administrative data (exclusion group 2)

We found indications that enrollees with administrative data that did not confirm their IAH eligibility had better health status than other enrollees. They had, on average, fewer ADLs for which they needed human assistance (where ADL data were not missing), fewer chronic conditions, and lower expenditures than enrolled beneficiaries with confirmed eligibility. The table below shows differences using Year 3 as illustration. We compared characteristics of two groups of non-confirmed beneficiaries—newly enrolled IAH patients (subgroup 1) and carryovers from the prior year who remained enrolled (subgroup 2)—with two groups of IAH beneficiaries with confirmed eligibility who were members of evaluation sample—enrolled beneficiaries (subgroup 3) and other patients of IAH practices who were not enrolled (subgroup 4).

Comparisons indicate generally better health status for excluded beneficiaries (subgroups 1 and 2) compared to included beneficiaries (subgroups 3 and 4). For example, about one in ten excluded beneficiaries had less than 2 ADLs according to administrative sources, the minimum required. They had lower average HCC risk scores (3.2 and 2.7 vs. 3.7 and 3.5), and averaged less than one hospitalization per year, compared to at least 1.6 per year for included beneficiaries. Excluded beneficiaries had generally lower total Medicare expenses (data not shown). An important exception to this pattern of more-favorable indicators for excluded enrollees was mortality, which was highest (29.1%) in the newly enrolled with unconfirmed eligibility in Year 3 (subgroup 1).

**Differences among subgroups of beneficiaries who did and did not meet IAH eligibility criteria: Year 3**

| | Subgroup 1 N=2,990 | Subgroup 2 N=1,728 | Subgroup 3 N=4,498 | Subgroup 4 N=3,066 |
|---|---|---|---|---|
| Eligibility confirmed in administrative data | No | No but confirmed in prior year | Yes | Yes |
| Enrolled by IAH practice | Yes, newly enrolled | Yes | Yes | No but received HBPC from IAH practice |
| Age <65 | 13.6% | 12.3% | 15.5%* | 16.4%* |
| Age > 85 | 45.4% | 45.1% | 37.7%* | 35.0%* |
| 0 or 1 ADLs | 12.0% | 10.2% | 0%* | 0%* |
| Dually eligible | 37.2% | 39.8% | 40.5% | 35.3% |
| Average HCC risk score | 3.2 | 2.7 | 3.7* | 3.5* |
| Number of hospital admissions per person[a] | 0.9 | 0.5 | 1.6* | 1.8* |
| Mortality[b] | 29.1 | 18.0 | 13.2* | 18.3* |

Source: Medicare claims and enrollment data for 2012–2015
[a]Average number of hospital stays annually.
[b]12-month mortality rate.
*The difference is significantly different from zero at the .10 level.

relatively more beneficiaries in better health in the follow-up year samples than in the baseline year samples.

### Samples used for beneficiary survey

The IAH beneficiary survey measured experiences of care, satisfaction, provider attributes, health outcomes, and beneficiary characteristics. We took some survey questions from the MCBS, and we used the MCBS survey files to provide national benchmark data from a subgroup of 2013 MCBS respondents similar to IAH participants. To the IAH survey, we added questions for several topics related to IAH beneficiaries' experiences with HBPC.

The IAH beneficiary sample consisted of 7,293 beneficiaries who enrolled in the demonstration between June 1, 2012, and June 30, 2014. The survey was administered between 2013 and 2015. Data presented in this report are based on responses from 3,870 beneficiaries (a response rate of 63.3 percent) weighted for non-response.

Our analytic approach compared IAH enrollees' responses to those of respondents to the 2013 MCBS on questions common to both surveys. The questions dealt with the care-process experiences of the beneficiary and self-reported health status. To allow for valid comparison, we selected a subset of respondents to the MCBS. The selected respondents provided an appropriate benchmark because they met demonstration eligibility requirements, according to analysis of Medicare claims and enrollment data and according to their self-reported functional status in the MCBS. Out of 14,874 respondents to the 2013 MCBS, we identified 360 beneficiaries who met the IAH eligibility criteria. To accommodate time lags between IAH enrollment and IAH survey administration dates, we made some adjustments in the measurement of IAH qualifying criteria for MCBS beneficiaries (e.g., MCBS beneficiaries were eligible if they had an inpatient hospitalization or observation stay in the two years prior to the interview date, rather than the IAH requirement of a one-year look-back period). Also, the definition of ADLs differed slightly from the IAH ADL requirements.

As noted earlier in this appendix, findings from our comparisons of IAH and MCBS survey respondents should be considered tentative. In other words, results are considered suggestive of care process differences that *may* exist between HBPC and usual care. The samples of IAH beneficiaries and IAH-eligible MCBS respondents were similar in many respects. However, the samples differed substantially on a few key variables that could also be related to reported care process quality. For example, IAH beneficiaries were more likely than MCBS respondents to live alone, and have dementia with complications, quadriplegia or paraplegia, and pressure ulcers. A significantly smaller share of IAH beneficiaries had lung, lymphoma and other cancers, and coagulation defects. Compared with MCBS respondents, more IAH beneficiaries reported needing help from another person with five or six ADLs. Although we controlled for these differences using regression analysis, differences in responses are not strong evidence of an effect of HBPC. One important reason is that our data comparing responses at a single point in time do not rule out pre-existing differences between IAH enrollees and MCBS respondents, before the former chose to use HBPC; that is, unobserved reasons for self-selection into HBPC

rather than to usual care may also be responsible for differences in average responses, as opposed to any effect of HBPC or the demonstration (or both).

### Samples used for care-partner interview

In interviews with care partners of IAH practices, we sought to learn how the care partners perceive their work with the HBPC practitioners and staff, asking them about experiences with communication and information sharing, care coordination, accessibility, and continuity. We also asked respondents to compare these experiences to their work with office-based practices and to identify any changes in their experiences with IAH practices since the demonstration began in 2012.

To recruit interview respondents among IAH practices' care partners, we asked each IAH practice to identify and provide contact information for up to seven care partners. Care partners may include home health agencies, hospices, specialists, durable medical equipment (DME) suppliers, pharmacists, social workers, and social service organizations. Because our data are limited to care partners identified by the IAH practices, their views may or may not represent the views of all care partners working with the practices. We interviewed up to five care partners for each practice, selecting a range of partner types from the lists provided by practices. A total of 48 care partners participated in interviews between November 2016 and January 2017. Table C4 shows the types of care partners that participated and the practices that identified them as sources.[65] The most common type of care partner came from a home health agency (n=21). The number of different care partner sources was largest for the Richmond practice (which was part of the mid-Atlantic Consortium). For Visiting Physicians Association (VPA) practices, in addition to interviewing one to three partners each referred from the five VPA sites (Dallas, Flint, Jacksonville, Lansing, and Milwaukee), we also interviewed three care partners in the VPA headquarters, based in Troy, Michigan. These interviewees were the VPA corporate vice president of case management and leaders of Grace Hospice and Pinnacle Senior Care—two organizations that work closely with local VPA practices and, like VPA, are owned by U.S. Medical Management, LLC.

[65] The Brooklyn practice did not identify any of their care partners.

## Table C4. Care partner sample: Type and IAH affiliation of interviewees

| | |
|---|---|
| Total number of individuals interviewed | 48 |
| **Number of care-partner interviewees by type of care partner** | |
| Home health | 21 |
| Hospice | 7 |
| Specialist | 5 |
| Assisted living facility/adult foster care | 3 |
| Pharmacist | 2 |
| DME/oxygen | 6 |
| Other[a] | 4 |
| **Number of care-partner interviewees by care partner's IAH practice affiliation** | |
| Austin | 1 |
| Boston | 3 |
| Brooklyn[b] | 0 |
| Cleveland | 3 |
| Dallas | 3 |
| Durham | 3 |
| Flint[c] | 3 |
| Jacksonville | 2 |
| Lansing | 1 |
| Milwaukee | 2 |
| Philadelphia | 2 |
| Portland | 4 |
| Richmond | 5 |
| VPA Corporate (Troy) | 3 |
| Washington, DC | 5 |
| North Shore, Long Island | 4 |
| Wilmington | 4 |

Source: Care partner interviews, November 2016–January 2017.

[a]Includes staff from an imaging center, emergency medical services program, Area Agency on Aging program, and a Medicaid waiver program.

[b]Brooklyn site did not submit a list of care partners.

[c]All three of the Flint care partners reported information relevant to both the Flint and Lansing practices, as well as differences between the two practices.

# Appendix D: Supplementary Information on IAH Practices

Information in this appendix discusses structural and operational characteristics of the three practice categories in the IAH demonstration. Tables D1 and D2 summarize information for a variety of attributes collected during interviews with practice representatives.

### Visiting Physicians Association

The five VPA practices (Dallas, Flint, Jacksonville, Lansing, and Milwaukee) have similar structural and operational characteristics. VPA is a corporation with multiple home-based primary care (HBPC) practices operating in multiple states; five of those practices are in the demonstration. Each practice has one or more patient care coordinators who serve as the main point of contact for patients and have access to the VPA corporate infrastructure for finance, human resources, data analytics, and data support. Patients (both IAH beneficiaries and others) are assigned to a mobile care team consisting of one physician and one medical assistant; visits occur at least once every four weeks.[66] Medical assistants provide administrative and clinical support to physicians. Each office has one clinical educator for case management support. Physicians provide most evaluation and management (E&M) visits to IAH beneficiaries.

VPA uses a centralized call center to provide standardized 24-hour support to local practice staff, including care coordination; prescription refills; patient needs assessment; and orders for home health, hospice, and durable medical equipment. Coordinators in the center receive notification of patient hospitalizations from HIEs in VPA FL and MI offices.

E&M visits occurred an average of 15.7 to 19.2 times annually for each beneficiary, depending on the VPA site (data not shown). In four of the VPA sites, more than 60 percent of these visits were conducted in homes; in the Milwaukee site, 71.5 percent of visits took place in assisted living facilities or other group living facilities (Table 4). Clinicians in VPA practices reportedly have an average panel size of 175 total patients (Table D1). VPA staff provide some weekend and non-billable visits. Each VPA practice reported conducting weekend visits for routine care as well as for urgent reasons, such as meeting the 48-hour requirement post-discharge. Although these visits are not billable, VPA clinical educators will often conduct home visits to patients.

All VPA practices have electronic systems that enable clinicians to collect data, communicate with the care team, and submit orders during a home visit (Table D2). Each VPA also risk-stratifies patients based on their history of hospitalization and ED visits to determine both the level of care and the frequency of proactive phone calls to patients and caregivers. Two practices have developed relationships with hospitals and their staff; those staff notify the practice directly when one of its patients is hospitalized or visits the ED, whereas the others receive automated notices from hospitals.

[66] The term *patients* in this section refers to all patients of the practice regardless of IAH enrollment status.

### Academic medical centers

Seven IAH practices are part of nonprofit academic medical centers (Boston, Philadelphia, Richmond, Washington, and Wilmington) or health systems with academic missions (Cleveland and North Shore)[67]. This status gives them access to institutional resources and information technology systems and support. Providers in these settings are typically responsible for training and education in addition to clinical care, so many see patients only part time. Across these practices, patients are assigned to a care team based on geographic service area, with some adjustment to ensure that clinicians have panels of roughly equal size. In Boston, Cleveland, and North Shore, physicians conducted all or most E&M visits; in Philadelphia, Richmond, and Washington, NPs conducted most of the visits.

The care teams at the academic medical centers comprise physicians, NPs, PAs, and social workers. Social workers are key members of the care team for many academic medical center practices, as they coordinate home health services and refer patients to social services and supports.

Three practices (Boston, Cleveland, and Wilmington) conducted fewer than 11 E&M visits per beneficiary per year**,** fewer than any of the other four academic medical center practices or any of the practices in the other two groups. However, all seven of the academic medical center practices conducted most E&M visits in home settings, and three (North Shore, Philadelphia, and Richmond) conducted no visits in assisted living facilities. Academic medical centers reported average panel sizes ranging from 40 to 200 patients per clinician. All but one academic medical center provides non-billable visits, such as those conducted by social workers or nurses (data not shown). Most also provide weekend visits, but only for urgent issues or to meet the 48-hour follow-up requirement. Two of the academic medical centers conduct regular visits after hours, and one provides after-hours visits only for urgent issues.

Academic medical centers vary in their use of technologies to facilitate care delivery and planning. Most rely on clinical judgment to determine the level of care, rather than a formal risk-stratification system (that groups the beneficiaries into high and low risk groups to aid in care planning) and, similarly, check in on patients as needed and as determined by clinicians' recommendations. Nearly all academic medical centers are notified automatically of patients' hospitalizations or ED visits from at least some hospitals with which they have built relationships. Unlike the other practices, which all have remote access to electronic medical records in the field, one academic medical center is unable to remotely access patients' data, collect new patient data, or submit orders during a home visit.

### Independent practices

The demonstration included four independent practices (Austin, Brooklyn, Durham, and Portland), diverse in their sizes, structures, and operating practices.

[67] Three practices (Philadelphia, Richmond, Washington) participated as part of one consortium, which the demonstration considers as one site.

The number and type of providers conducting home visits in independent practices differ across each site, with some practices having an approximately equal mix of physicians, NPs, and PAs conducting home visits, and others primarily relying on one or two types of providers. Some practices assign patients to one provider; others assign them to a team of two or more providers. In the Brooklyn and Durham practices, physicians provided most of the E&M visits to Medicare beneficiaries, whereas NPs provided most of the visits in Portland.

Non-medical support staff among the independent practices serve multiple functions. All have staff dedicated to coordinating care for patients; however, the type of staff used varies across the sites. For example, some have nurse care managers and others train medical assistants or similar staff to be patient care coordinators. Several independent practices have a social worker on staff to assist with care coordination and address behavioral health issues; others rely on the social work staff at home health agencies to connect patients to needed services. Additional support staff in independent practices, such as patient liaisons, are responsible for a variety of activities, including scheduling home visits, connecting patients to specialists, referring patients to social services and resources, and communicating with home health agencies.

Visit frequency varied substantially within this group, from 11.8 to 26.9 visits per beneficiary per year. The location of these visits varied across the sites—from all visits conducted in home settings to a high of 84.4 percent conducted in assisted living facilities. Average panel size also varied widely, ranging from 80 to 200 patients per clinician. Most of the independent practices reported conducting weekend or after-hours visits for both urgent and non-urgent reasons. Three of the independent practices provide non-billable visits by social workers and nurse care managers.

One independent practice reported risk-stratifying patients to determine the intensity of care, whereas the rest reported relying on clinicians' judgment for these determinations. Independent practices reported using different methods for learning of hospitalizations and ED visits, with one relying on patients and caregivers to notify practice clinicians and others receiving notice through health information exchanges.

## Table D1. Table of practice visit characteristics

| Practice site | Visits per clinician per day | Clinician panel size | Non-billable visits | Weekend visits | After-hours visits | Visits in settings other than home (office or clinic) |
|---|---|---|---|---|---|---|
| **Independent practices** | | | | | | |
| Austin, TX | 10 | 200 | No | No | Yes: for urgent visits only, uncommon | No |
| Brooklyn, NY | 8 to 10 | 120 to 130 | Yes: visits to uninsured patient; uncommon | Yes: for both regular and urgent or post-discharge visits | Yes: for urgent and regular visits, common | No |
| Durham, NC | 10 to 15 | 150 to 200 | No | Yes: for urgent or post-discharge only | No | No |
| Portland, OR | 4 or 5 | 80 to 120 | Yes: RN, social worker, or chaplain visit as needed | Yes: for urgent or post-discharge only | No | No |
| **Academic medical center practices** | | | | | | |
| Boston, MA | 4 | 90 | Yes: visits from nurse care manager | No | Yes: for urgent visits only, uncommon | No |
| Cleveland, OH | 6 or 7 | 150 to 200 | No | Yes: for urgent or post-discharge only | Yes: for urgent and regular visits, uncommon | No |
| North Shore, NY | 6 | 170 | Yes: community paramedic visit for urgent issues only | No | No | No |
| Philadelphia, PA* | 6 | 140 | No | Yes: for urgent or post-discharge only | Yes: for urgent visits only, uncommon | No |
| Richmond, VA* | 3 to 6 | 40 | Yes: nurse visit, but uncommon | No | No | No |
| Washington, DC* | 6 | 150 to 170 | Yes: social worker visit as needed | Yes: for urgent or post-discharge only | Yes: for regular visits, uncommon | Yes |
| Wilmington, DE | 6 | 90 to 120 | Yes: RN or social worker visit as needed or requested by patients and caregivers | Yes: for urgent or post-discharge only | No | Yes |
| **VPA practices** | | | | | | |
| Dallas, TX | 8 or 9 | 175 | Yes: clinical educator visits | Yes: for both regular and urgent or post-discharge visits | No | No |
| Flint, MI | 8 or 9 | 175 | Yes: clinical educator or social worker visits | Yes: for both regular and urgent or post-discharge visits | No | No |
| Jacksonville, FL | 8 or 9 | 175 | Yes: clinical educator visits | Yes: for both regular and urgent or post-discharge visits | No | No |
| Lansing, MI | 8 or 9 | 175 | Yes: clinical educator or social worker visits | Yes: for both regular and urgent or post-discharge visits | No | No |
| Milwaukee, WI | 8 or 9 | 175 | Yes: clinical educator visits | Yes: for both regular and urgent or post-discharge visits | No | No |

Source: Telephone interviews with practices in January 2017.

*Consortium member. ACO = accountable care organization; ALF = assisted living facility; RN = registered nurse; VPA Visiting Physicians Association.

## Table D2. Table of practice care delivery and operational characteristics

| Practice site | Formal risk-stratification classification | Remote access to patient's record and remote submission of orders | Notification of hospitalization or ED visit | Participate in ACO (years) | Proactive outreach to patients or caregivers |
|---|---|---|---|---|---|
| **Independent practices** | | | | | |
| Austin, TX | Yes: assessment scores, hospitalization history, and clinical judgment used to assign level-of-risk score, which determines level of proactive outreach and care team involvement | Yes | Rely on patient or caregivers to notify practice | No | Yes: call weekly or biweekly based on acuity of patient |
| Brooklyn, NY | No: clinical judgment only | Yes | Automated notice from some institutions | Yes (2) | No |
| Durham, NC | No: clinical judgment only | Yes | Notification from ALFs where majority of patients live | No | Yes: call as needed |
| Portland, OR | No: clinical judgment only | Yes | Automated notice from HIE | No | Yes: call as needed based on acuity of patient and if patient was recently hospitalized |
| **Academic medical centers** | | | | | |
| Boston, MA | No: clinical judgment only | Yes | Automated notice from some sites | Yes | Yes: call as needed based on care plan |
| Cleveland, OH | No: clinical judgment only | No | Automated notice from some institutions | Yes (1) | Yes: call twice weekly |
| North Shore, NY | Yes: determines level of proactive outreach and care team involvement | Yes | Automated notice from some institutions | No | Yes: call as needed based on acuity of patient |
| Philadelphia, PA* | No: clinical judgment only | No | Automated notice from some institutions | No | Yes: call as needed based on provider's judgment |
| Richmond, VA* | No: clinical judgment only | Yes | Automated notice from some institutions | No | No |
| Washington, DC* | No: clinical judgment only | Yes | Automated notice from some institutions | No | Yes: call monthly |
| Wilmington, DE | Yes: software assesses patients and assigns level of acuity score, which determines level of proactive outreach and care team involvement | Yes | Automated notice from all institutions | Yes (1) | Yes: call as needed for high-need patients; for those recently hospitalized, weekly in first month post-discharge and biweekly for second month post-discharge |
| **VPA** | | | | | |
| Dallas, TX | Yes: based on hospital or emergency department admissions; if patient has 2 or more visits in 60-day period, the patient is enrolled in an intensive care management program | Yes | Rely on hospital staff to notify practice | Yes (2) | Yes: call as needed based on acuity of patient |
| Flint, MI | Yes: based on hospital or emergency department admissions; if patient has 2 or more visits in 60-day period, the patient is enrolled in an intensive care management program | Yes | Automated notice from all institutions | Yes (5) | Yes: call as needed based on acuity of patient |

| Practice site | Formal risk-stratification classification | Remote access to patient's record and remote submission of orders | Notification of hospitalization or ED visit | Participate in ACO (years) | Proactive outreach to patients or caregivers |
|---|---|---|---|---|---|
| Jacksonville, FL | Yes: based on hospital or emergency department admissions; if patient has 2 or more visits in 60-day period, the patient is enrolled in an intensive care management program | Yes | Automated notice from all institutions | Yes (2) | Yes: call as needed based on acuity of patient |
| Lansing, MI | Yes: based on hospital or emergency department admissions; if patient has 2 or more visits in 60-day period, the patient is enrolled in an intensive care management program | Yes | Automated notice from all institutions | Yes (2) | Yes: call as needed based on acuity of patient |
| Milwaukee, WI | Yes: based on hospital or emergency department admissions; if patient has 2 or more visits in 60-day period, the patient is enrolled in an intensive care management program | Yes | Rely on hospital staff to notify practice | Yes (2) | Yes: call as needed based on acuity of patient |

[a] Length of ACO involvement not available.

Source: Telephone interviews with practices in January 2017.

*Consortium member. ACO = accountable care organization; ALF = assisted living facility; ED = emergency department; HIE = health information exchange; VPA = Visiting Physicians Association.

NA = not available