Evaluation of the Independence at Home Demonstration

An Examination of the First Five Years - Appendices

March 2020

Evelyn Li, Laura Kimmey, Valerie Cheh

Submitted to:

Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
7500 Security Blvd.
Baltimore, MD 21244
Project Officer: David Nyweide
Contract Number: HHSM-500-2014-00034I / HHSM-500-T0018

Submitted by:

Mathematica
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Valerie Cheh

This page has been left blank for double-sided copying.

CONTENTS

Appendix A

Data and methods technical appendix

This page has been left blank for double-sided copying.

I. OVERVIEW

Congress mandated the Independence at Home (IAH) demonstration to test a payment incentive and service delivery model for home-based primary care. Under the IAH demonstration, physicians and nurse practitioners (NPs) direct home-based primary care teams. These teams aim to reduce expenditures and improve health outcomes of Medicare beneficiaries with multiple chronic conditions and substantial functional limitations. As discussed in Chapter I, the IAH demonstration introduced both an incentive to reduce Medicare expenditures (incentive payments) and a service delivery model (home-based primary care led by physicians or NPs). As we described in Chapter II, the Mathematica study team estimated a difference-in-differences model to determine whether the demonstration payment incentive affected Medicare expenditures, service utilization, and quality of care (measured as potentially avoidable hospital use). We also examined whether the demonstration payment incentive had unintended consequences on mortality or entry into institutional long-term care. In this appendix, we present the sample, data, and methods we used for the analyses in Chapter II.

The quantitative evaluation design of the demonstration was a difference-in-differences analysis using repeated cross-sections of eligible beneficiaries within demonstration practices with a propensity score-matched comparison group. We had two years of pre-demonstration data and five years of post-demonstration data (the first five years of the demonstration). Beneficiaries analyzed by the evaluation were observed for the number of months they were eligible for IAH for each demonstration year. We used three key pieces of information to determine the effect of the demonstration on expenditures, service utilization, quality of care, and unintended consequences in a given year. To determine the effect of the demonstration on expenditures (and other outcomes) in a given year, such as Year 5, we did the following:

- Estimated the difference in Medicare expenditures per beneficiary per month (PBPM) between the year before the demonstration (the baseline year) and Year 5 for IAH beneficiaries. We restricted claims to those occurring between the date of eligibility for the demonstration in a given year and the end of that year (and date of death). We controlled for beneficiary characteristics, such as time since most recent hospitalization; demographic characteristics; activities of daily living (ADLs); and several measures of health status, including the Centers for Medicare & Medicaid Services (CMS) Hierarchical Condition Categories (HCC) risk score. We provide a complete list of control variables later in this appendix.
- Estimated the difference in Medicare expenditures during the same period for comparison beneficiaries. As with the IAH group, we restricted claims to those that occurred between the date of eligibility and the end of the year, controlling for beneficiary characteristics.
- Obtained the estimated effect of the demonstration by calculating the difference between the change in expenditures for IAH beneficiaries and the change in expenditures for comparison beneficiaries.

We refer to this model as a difference-in-differences model because it measured the change between two differences (the pre- and post-demonstration differences). This method isolated the effect of the demonstration by accounting for two factors. First, it accounted for the difference in expenditures between IAH and comparison beneficiaries before the demonstration. Second, it accounted for changes in expenditures during the demonstration caused by factors unrelated to the demonstration and that affected both IAH and comparison beneficiaries. This before-and-after design provided a strong assessment of the demonstration's effect, assuming that the difference in expenditures between IAH and comparison beneficiaries was stable before the demonstration. As we describe later, we tested this assumption. However, the difference-in-differences model was not without limitations; we address our evaluation's limitations at the end of this appendix.

Our total sample consisted of 14 practices because we treated the consortium in Richmond as one practice (Exhibit I.5). Our quantitative analyses excluded three practices (Atlanta, Chicago, and Stuart) that withdrew from the demonstration before Year 3 and one practice (Louisville) that CMS terminated for cause after completing the first three years.

In this appendix, we begin by describing how the IAH practices operate and characteristics of their patients. Next, we describe how we identified the IAH and comparison groups to evaluate the effect of the demonstration, which is designed to have IAH sites assess and enroll participants; however, we could not use data from the IAH sites to identify our sample because we needed to use the same source of data to identify the IAH and comparison groups. As a result, the sample of beneficiaries enrolled by the practices in the demonstration differed from the beneficiaries in the IAH group we used for the evaluation. For example, about 60 percent of the IAH group we used for the evaluation in Year 5 was enrolled in the demonstration in Year 5. As we describe in the next section, the IAH group for the evaluation consisted of beneficiaries who Mathematica identified as eligible for the demonstration and attributed to an IAH site. Next, we describe how we selected the comparison group. We then present the sources of data and measures for our quantitative analyses. We then describe the estimation of demonstration effects. Next, we present the methods and sources of data for our qualitative analysis and conclude with a discussion of the study's limitations.

II. DESCRIPTION OF IAH PRACTICES AND BENEFICIARIES

To understand the features of the IAH practices and identify the changes they made to improve care, we collected and analyzed interview data from the practices and analyzed their claims data.[1] Every IAH site had substantial experience in providing home-based primary care before the demonstration. However, the sites differed substantially in their approaches to care, such as who was included on the care team; whether they tracked patients across care settings; whether they focused on serving in private homes or assisted living facilities; and whether they used a formal risk-stratification system, which groups the beneficiaries into high- and low-risk groups to aid in care planning. We grouped each IAH practice on the basis of its structural and operational characteristics. In this section, we summarize care delivery patterns according to each of the three types of practices: (1) Visiting Physicians Association (VPA) practices, (2) academic medical center practices, and (3) independent practices. We obtained information about the sites of care from claims data in Years 2 and 4 of the demonstration. Exhibit A.1 provides site-by-site information on practices' structural and operational characteristics.

A. Visiting Physicians Association

The five VPA practices (Dallas, Flint, Jacksonville, Lansing, and Milwaukee) had similar structural and operational characteristics. VPA is a corporation with multiple home-based primary care practices operating in several states; five of those practices were in the demonstration. Each practice had a patient care coordinator who was the main point of contact for patients and had access to the VPA corporate infrastructure for finance, human resources, data analytics, and data support.

Patients (both IAH beneficiaries and others) were assigned to a mobile care team consisting of one physician and one medical assistant.[2] VPA clinical educators often conducted home visits to patients, although those visits were not billable. In four of the VPA sites, about two-thirds of visits occurred in private homes. In Milwaukee, about two-thirds of visits occurred in assisted living or other group living facilities.

Each VPA risk-stratified patients on the basis of their history of hospitalization and ED visits to determine the needed level of care and the frequency of proactive phone calls to patients and caregivers. Two practices developed relationships with hospitals and their staff; those staff notified the practice directly when one of its patients was hospitalized or visited the ED, whereas the remaining three received automated notices from hospitals.

[1] Information in this section is drawn from site visits we conducted from February to May 2013 and February to July 2014. In January and February 2017, we conducted telephone interviews to confirm and update information for all practices.

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

(continued)

B. Academic medical centers

Seven IAH practices (Boston, Cleveland, Long Island, Philadelphia, Richmond, Washington, and Wilmington) were part of nonprofit academic medical centers or health systems with academic missions.[3] This status gave them access to institutional resources and information technology systems and support. Clinicians in these settings were typically responsible for training and education in addition to clinical care, so many saw patients only part time.

In Boston, Cleveland, and Long Island, physicians conducted all or most visits; in Philadelphia, Richmond, and Washington, NPs conducted most of the visits. In Wilmington, NPs and physicians conducted most of the visits. Social workers were key members of the care team for many academic medical center practices because they coordinated home health services and referred patients to social services and supports. All but one academic medical center provided nonbillable visits, such as those conducted by social workers or nurses not acting under a physician's direction or as part of a home health episode. All academic medical center practices conducted most visits in private home settings; three (Long Island, Philadelphia, and Washington) conducted no visits in assisted living facilities.

Academic medical centers varied in their use of technologies to facilitate care delivery and planning. Most relied on clinical judgment to determine the level of care rather than using a formal risk-stratification system. Nearly all were notified automatically of patients' hospitalizations or ED visits from at least some hospitals with which they had built relationships.

C. Independent practices

The demonstration included four independent practices (Austin, Brooklyn, Durham, and Portland) that were diverse in size, structure, and operating practices. All four independent practices had staff dedicated to coordinating care for patients; however, the type of staff used to coordinate care varied across the sites. For example, some had nurse care managers, whereas others trained medical assistants or similar staff as patient care coordinators.

In Brooklyn and Durham, physicians provided most of the visits, whereas in Portland, NPs provided most of those visits. In Austin, about half of visits were made by NPs, and about one-quarter each by physicians and PAs. The sites of care for independent practices varied across them—Brooklyn and Austin conducted most visits in private home settings, whereas Durham and Portland conducted most visits in assisted living or other group living facilities. Some of the independent practices provided nonbillable visits by social workers and nurse care managers.

[3] Three practices (Philadelphia, Richmond, and Washington) participated as one consortium, which the demonstration considers as one site for the purpose of calculating incentive payments.

One independent practice reported risk-stratifying patients as a way to determine the intensity of care the practice would provide, whereas the remaining three reported relying on clinicians' judgment for these determinations. These practices reported using different methods for learning about patient hospitalizations and ED visits; one relied on patients and caregivers to notify the practice, and others received notice through health information exchanges.[4]

Exhibit A.1. Characteristics of IAH practices, as of 2017

Site	Affiliation	Full-time clinicians making house calls	Part-time clinicians making house calls	Visits per clinician per day	Other staff involved in care team
Dallas, TX	US Medical Management	17 clinicians[a]	None	8 or 9	18 MAs, 2 clinical educators on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[b]
Flint, MI	US Medical Management	23 clinicians[a]	None	8 or 9	24 MAs, 5 clinical educators on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[b]
Jacksonville, FL	US Medical Management	14 clinicians[a]	2 clinicians	8 or 9	10 MAs, 1 clinical educator on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[b]
Lansing, MI	US Medical Management	10 clinicians[a]	None	8 or 9	11 MAs, 2 clinical educators on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[b]
Milwaukee, WI	US Medical Management	12 clinicians[a]	None	8 or 9	11 MAs, 1 clinical educator on site, 1 scheduler, 1 patient care coordinator, 1 practice manager[b]
Boston, MA	Boston Medical Center	None	6 physicians	4	5 nurses, 1 office manager, 3 ambulatory service representatives, 1 project coordinator
Cleveland, OH	Cleveland Clinic	7 physicians, 3 NPs	1 PA	6 or 7	3 RNs, 4 MAs, 1 nurse manager, 1 social worker, 3 schedulers, 1 pharmacist

[4] Health information exchanges allow the electronic sharing of health care information. They can be implemented at different levels, including a region (such as the greater Washington, DC area), community, or hospital system.

Exhibit A.1 (*continued*)

Site	Affiliation	Full-time clinicians making house calls	Part-time clinicians making house calls	Visits per clinician per day	Other staff involved in care team
Long Island, NY	Northwell Health	4 physicians, 2 NPs	2 physicians	6	6 nurses, 6 medical coordinators, 5 social workers, 1 clinical data analyst, 1 DME coordinator
Philadelphia, PA[c]	University of Pennsylvania	1 NP	3 physicians, 1 NP	6	1 social worker
Richmond, VA[c]	Virginia Commonwealth University	2 physicians, 6 NPs	2 physicians, 1 NP	3 to 6	2 RNs, 1 consulting pharmacist, 3 social workers, 1 office manager, 3 patient access representatives
Washington, DC[c]	MedStar Health	6 physicians, 5 NPs	1 NP	6	1 RN, 1 LPN, 5 MAs, 1 social worker, 1 outcomes analyst
Wilmington, DE	Christiana Care Health Systems	1 physician, 3 NPs	4 physicians, 1 PA, 1 NP	6	1 phlebotomist, 4 RNs, 4 MAs, 3 social workers, 1 office manager
Austin, TX	Kindred Health Care	4 physicians, 9 NPs, 4 PAs	2 physicians	10	5 LPNs, 2 MAs serving as patient service coordinators, 2 intake coordinators, 1 office manager, 1 medical record personnel
Brooklyn, NY	None	10 physicians, 15 PAs, 9 NPs[d]	None[d]	8 to 10	1 quality assurance nurse, 1 patient liaison[d]
Durham, NC	None	33 physicians, 35 PAs, 7 NPs	None	10 to 15	6 podiatrists; 2 psychologists; 1 social worker; 130 additional office support staff, 40 of whom are MAs serving in clinical service, management, and scheduling capacities
Portland, OR	None	4 physicians, 3 NPs, 1 PAs	1 physician, 1 PA, 3 NPs	4 or 5	17 RNs, 4 LPNs, 7 social workers, a team of care coordinators, 1 care coordinator supervisor, 1 DME specialist

Source: Information from interviews with practice staff conducted in 2015 and 2017.

[a]VPAs did not provide a breakdown of physicians, NPs, and PAs.

Exhibit A.1 (*continued*)

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

[c]These three sites (Philadelphia, Richmond, and Washington, DC) are considered one practice for purposes of the demonstration.

[d]The Brooklyn, New York site did not provide information in 2017 on the number of full- and part-time clinicians making house calls or other staff involved in the care team.

DME = durable medical equipment; IAH = Independence at Home; LPN = licensed practical nurse; MA = medical assistant; NP = nurse practitioner; PA = physician assistant; RN = registered nurse; VPA = Visiting Physicians Association.

D. Characteristics of IAH beneficiaries

In the year before the demonstration, more than half of IAH beneficiaries were age 80 or older, and 40 percent were dually eligible for Medicare and Medicaid (Exhibit A.2). The demonstration eligibility criteria focused on Medicare beneficiaries who were chronically ill and disabled. As a result, about 43 percent of IAH beneficiaries had 10 or more chronic conditions, and 55 percent required human assistance with at least five ADLs. On average, IAH beneficiaries incurred nearly $4,400 in Medicare expenditures per beneficiary per month (PBPM) in the year before the demonstration. They had an average of 1.8 hospital admissions and 2.9 ED visits per year. About 18 percent of IAH beneficiaries died within twelve months of meeting demonstration eligibility criteria.

IAH beneficiaries were more likely to be dually eligible, older, have more chronic conditions, and have a higher mortality rate than the average Medicare beneficiary. Among the IAH states, the average percentage of Medicare beneficiaries who were dually eligible in 2013 was 20 percent (Kaiser Family Foundation). In 2012, 36 percent of beneficiaries who resided in the community (not a facility) were older than age 75, 26 percent had five or more chronic conditions, and 3 percent died over the course of the survey year (CMS 2012). IAH beneficiaries also struggled with daily activities at a higher rate than the average Medicare beneficiary. Only 12 percent of Medicare beneficiaries who resided in the community reported difficulties in performing three or more ADLs without human assistance or special equipment, such as a walker or grab bar (CMS 2012).

Exhibit A.2. IAH beneficiaries' demographic characteristics and health status, Medicare expenditures, and service utilization in the year before the demonstration

Variable name	Value for IAH beneficiaries in the year before the demonstration
Demographic characteristics and health status	
Percentage age 80 or older	51.7
Percentage dually eligible for Medicare and Medicaid	40.1
Average HCC score	3.5
Percentage with 10 or more chronic conditions	42.7
Percentage requiring human assistance with at least five activities of daily living	55.0
Average Medicare expenditures per beneficiary per month	
Total	$4,397
Inpatient hospital services	$1,741

Exhibit A.2 (*continued*)

Variable name	Value for IAH beneficiaries in the year before the demonstration
SNF services	$605
Home health services (Parts A and B)	$781
Hospice services	$153
Outpatient services	$253
Physician/supplier services	$715
Durable medical equipment	$150
Average numbers of key utilization events per beneficiary per year	
Number of hospital admissions[a]	1.8
Number of potentially avoidable hospital admissions[b]	0.5
Number of ED visits	2.9
Visits by primary care clinicians[c]	11.2
Visits by specialists	5.7
Probability of key utilization events	
Probability of having a qualifying hospital discharge and an unplanned readmission within 30 days of discharge (percentage)[d]	19.6
Probability of home health use (percentage)	91.3
Probability of hospice use (percentage)	17.9
Probability of SNF use (percentage)	41.0
12-month mortality (percentage)	18.1

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

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

[b]The number of potentially avoidable hospital admissions includes observation stays. A potentially avoidable hospital admission is one in which appropriate primary and specialty care might prevent or reduce the need for a hospital admission.

[c]Primary care clinicians are defined as primary care physicians, nurse practitioners, and physician assistants. Nonacute settings are defined as home, office, outpatient clinic, federally qualified health center, or rural health clinic.

[d]The probability of an unplanned readmission for a beneficiary is measured over the IAH-eligible months during each demonstration year. The probability equals zero for beneficiaries who did not have a qualifying hospital discharge or an unplanned readmission within 30 days of a qualifying hospital discharge during the measurement period.

ED = emergency department; HCC = Hierarchical Condition Category; IAH = Independence at Home; SNF = skilled nursing facility.

III. IDENTIFYING THE IAH BENEFICIARIES

To comply with the legislation that established the IAH demonstration, the demonstration used a site-based enrollment process. Sites were responsible for ensuring that the enrollees met health status and other clinical and programmatic requirements, such as providing consent. The implementation contractor used both administrative data and information provided by the sites to construct the list of enrolled beneficiaries as part of its work to calculate spending by IAH beneficiaries in each practice.

Although the implementation contractor used Medicare claims data, other administrative data, and information provided by the sites to construct the list of enrollees, Mathematica used only Medicare claims and other administrative data to identify the IAH group for the evaluation. (See Section IV of this appendix for more information about the data sources we used to determine eligibility.) To measure the effect of the demonstration, we had to use the same data sources and approach to identify the IAH and comparison groups across all pre-demonstration and demonstration years. Information provided by the sites to construct the list of IAH enrollees was available for the demonstration years only, not the pre-demonstration years. Also, no information other than administrative data was available for the comparison group. As a result, we used only administrative data to define the IAH group in each pre-demonstration and demonstration year, rather than using the information the sites provided to the implementation contractor. We describe our process for defining the IAH group in this section. We describe our process for identifying the comparison group in Section III of this appendix.

The approaches of Mathematica and the implementation contractor to identifying eligible beneficiaries yielded different counts of IAH practices' beneficiaries in Years 1 to 5. After explaining these approaches in Sections II.A and II.B, we provide details about reasons for differences in the counts of IAH practices' eligible beneficiaries in Section II.C.

A. Process the IAH implementation contractor used to determine the sample of enrolled beneficiaries

The IAH sites identified beneficiaries they thought were eligible to participate in the demonstration; we list the eligibility requirements in Chapter I. After providing these beneficiaries with information about the demonstration and conducting home visits to explain it, the IAH sites enrolled willing participants and uploaded a list of potential enrollees to a reporting system created for the demonstration using a process established by the implementation contractor. The contractor then used administrative data to verify that each enrolled beneficiary had a qualifying hospitalization and had used rehabilitation services in the previous 12 months, was covered by Medicare Parts A and B, and was not enrolled in a Medicare Advantage plan as of the date of IAH enrollment.

In addition to verifying whether the beneficiaries enrolled by the practices had a qualifying hospitalization and had used rehabilitation services, the implementation contractor also assisted

IAH sites in identifying potential beneficiaries for enrollment into the demonstration based on the eligibility criteria. The contractor identified beneficiaries who had received at least one home visit by the demonstration practice and had had qualifying hospitalization and rehabilitation service events but whom the sites had not yet enrolled in the reporting systems; these beneficiaries were called "potential enrollees." The contractor provided the sites with information on the potential enrollees, and the sites then reviewed their records and assessed additional information about the beneficiaries' eligibility (such as whether they met the ADL and chronic condition criteria). Clinicians followed up with potential enrollees who met all demonstration criteria and enrolled them in the demonstration.

The implementation contractor set the enrollment date as the first day of the month after the beneficiary had a qualifying hospitalization, used post-acute rehabilitation services, and received a home visit by the IAH practice within the previous 12 months. The home visit by the practice may have occurred before or after the qualifying hospitalization and rehabilitation services as long as all three occurred within 12 months before the enrollment date.

If the beneficiary did not meet the demonstration eligibility criteria, the sites provided the implementation contractor with the reason for the beneficiary's ineligibility. Reasons sites reported for not enrolling beneficiaries whom the contractor identified as potential enrollees included the following: (1) the beneficiary did not meet the ADL or chronic condition criteria; (2) the beneficiary received primary care from another practice and the IAH practice was not considered the beneficiary's primary practice; (3) the beneficiary began receiving hospice care, moved into a nursing home, or died before receiving notification of his or her eligibility for the demonstration; and (4) the beneficiary refused to participate in the demonstration. If the IAH practice did not provide any reasons for ineligibility for a potential enrollee, the implementation contractor assumed that the beneficiary was eligible and added that person to the official demonstration enrollment records.

We refer to all beneficiaries confirmed as IAH participants in the implementation contractor's records as "enrolled beneficiaries." CMS allowed beneficiaries who enrolled in the demonstration in a given year to continue in the demonstration, whether or not they requalified in subsequent years.

B. Process Mathematica used to identify the sample of eligible and attributed beneficiaries for the evaluation

To identify beneficiaries eligible for the demonstration and attributed to a demonstration practice, Mathematica used different processes and data sources than those used by the implementation contractor and the IAH sites. As explained earlier, our method for measuring the effect of the demonstration required us to use the same data sources and approach to identify the IAH and comparison groups across all pre-demonstration and demonstration years. We could not use enrollment in the demonstration as part of determining who would be in our sample, because enrollment was based in part on information from the IAH practices. Therefore, the IAH group

consisted of all beneficiaries eligible for the demonstration in that year according to our analysis of Medicare enrollment, claims, and assessment data.

We applied the following criteria to identifying beneficiaries for the IAH group:

- Enrollment in fee-for-service (FFS) Medicare
- Two or more ADLs that required human assistance
- Two or more chronic conditions
- Inpatient hospitalization or observation stay in the previous 12 months[5]
- Use of acute or subacute rehabilitation services in the previous 12 months[6]
- Not in hospice or long-term care for the entire time they were eligible for the demonstration in a given year and not on hospice on the first day of demonstration eligibility

For beneficiaries enrolled in the demonstration, the eligibility date determined by Mathematica based on administrative data sometimes differed from the enrollment date determined by the implementation contractor. Mathematica set the eligibility date as the first day of the month following the last service use required to qualify for the demonstration. For example, if a beneficiary had a hospitalization in July 2016 and home health care in October 2016, that person would be eligible for demonstration Year 5 as of November 1, 2016.

In the following section, we explain how we used assessment data to measure limitations in ADLs. After that, we explain how we used Medicare claims to attribute eligible beneficiaries to the IAH group.

1. Eligibility and assessment data

We measured ADL limitations in accordance with the guidelines the IAH implementation contractor gave to IAH practices. Those guidelines stated that a beneficiary qualified as having an ADL limitation if he or she needed any type of human assistance with the activity. The exception to this general guideline was for wheelchair use; use of a wheelchair as the primary mode of mobility with or without human assistance qualified as an ADL limitation for enrollment in the IAH demonstration.

To measure limitations in ADLs for the evaluation sample, we used assessment data from the given pre-demonstration or demonstration year. We used three sources of assessment data: (1) the Outcome and Assessment Information Set (OASIS), collected when a beneficiary receives home health care; (2) the Minimum Data Set (MDS), collected when a beneficiary receives skilled nursing facility (SNF) care; and (3) the Inpatient Rehabilitation Facility Patient

[5] Includes acute care, critical access, and psychiatric hospitals.

[6] Includes discharge from inpatient rehabilitation hospitals and rehabilitation units or skilled nursing facilities (SNFs), and use of home health (but not necessarily discharge). We did not include long-term care hospitals.

Assessment Instrument (IRF-PAI), collected when a beneficiary receives inpatient rehabilitation facility care. All three data sets provided information about the extent to which a beneficiary could complete the six standard ADLs—dressing, bathing, toileting, transferring, ambulating, and feeding. Transferring includes transfer between bed and chair and excludes transferring to or from the bath or toilet. Each assessment instrument has one or more data elements that indicate the extent of limitations, if any, for each of the six ADLs. If a beneficiary did not have any assessment data in a given year, that person was ineligible for the demonstration in that year, and we did not include him or her in our sample.

We faced three challenges when measuring limitations with the six ADLs. First, each ADL is coded differently in each of the three data sets. Second, different providers collect ADL data at different points in time. Third, a beneficiary can have multiple assessments in a given year. Next, we discuss how we handled each of those three challenges.

a. Each ADL is coded differently in each data set

Each ADL limitation is coded differently in each data set, and the codes do not always clearly define the person's need for human assistance to do the activity. We reviewed all of the values of each variable that measured ADL functioning. If the value for a particular beneficiary indicated that the person needed human assistance to do the activity safely, we classified him or her as needing human assistance with that ADL; we had to measure the need for human assistance as best we could.

In cases in which the level of functioning did not make clear that the beneficiary required human assistance to complete the activity, we erred on the side of not including patients. For example, one of the possible values for the transferring data element in an OASIS assessment was "able to transfer with minimal human assistance or with use of an assistive device," such as a walker. If a beneficiary had an OASIS assessment with that value for the transferring data element, we did not consider that person to have a limitation that required human assistance for transferring based on that particular assessment. This conservative approach excluded from our sample beneficiaries who required a device but not human assistance, such as beneficiaries who could get out of bed alone when using a walker. However, it might also have excluded some people who required human assistance and therefore could be IAH eligible.

Although we usually did not score a beneficiary as having a limitation if he or she needed human assistance or an assistive device, we applied one exception to that rule. In accordance with the guidelines given to IAH practices by the implementation contractor, use of a wheelchair as the primary mode of mobility with or without human assistance qualified as an ADL limitation.

b. Different providers collect ADL data at different points in time

CMS requires that health care providers conduct OASIS, MDS, and IRF-PAI assessments at specific points in time. For example, a beneficiary who received skilled nursing services for a 60-day period may have had MDS data from assessments at admission, at discharge, and at the

time of any significant changes in status. Because providers conduct each of these assessments at multiple points in time, we had to determine which assessments we would use in measuring ADL limitations to determine IAH eligibility. We used discharge assessments from all three data sets, as well as interim assessments from the OASIS data set. We did not use admission or interim assessments from the MDS and IRF-PAI because a beneficiary must be discharged from an SNF or inpatient rehabilitation facility before becoming eligible for IAH. Unlike with skilled nursing and inpatient rehabilitation services, a beneficiary can receive Medicare-funded home health care on the date he or she becomes eligible for IAH. Therefore, we included interim OASIS assessments[7] in addition to discharge assessments to ensure we had the latest information in the study year.

c. A beneficiary can have multiple assessments in a given year

A beneficiary could have had more than one assessment in a given year. For example, in one demonstration year, a beneficiary could have had three sets of assessment data: an interim OASIS assessment from home health care, a discharge OASIS assessment from home health care, and a discharge MDS assessment from skilled nursing care. When beneficiaries had more than one assessment in a given year, we kept the most recent assessment in which a beneficiary had at least two ADL limitations. We selected the most recent ADL assessment in which a beneficiary had at least two ADL limitations because we sought to identify beneficiaries who were least likely to recover from the ADL limitation. If a beneficiary had assessment data during a given year but not at least two ADL limitations in any of those assessments, that person was ineligible for the demonstration in that year, and we did not include him or her in our sample. Also, if a beneficiary did not have any assessment data in a given year, that person was ineligible in that year, and we did not include him or her in our sample.

2. Attribution and enrollment data

In addition to determining eligibility for the demonstration, in each year we applied the following criteria for attributing a patient to a demonstration site (we used Medicare claims data for visits to the IAH practice that occurred between the date of eligibility for the demonstration and the end of the demonstration year):

- Residence in the same state as the demonstration practice
- At least one evaluation and management[8] (E&M) or non-E&M home visit from the demonstration practice; "home" included private homes, assisted living facilities, group homes, and custodial care facilities

[7] Interim home health (OASIS) assessments do not include scoring on one activity: feeding. Because this item's effect on overall eligibility determination is small, we did not apply any adjustments to interim assessments.

[8] "E&M visit" refers to a patient-provider encounter during which the provider assesses the patient's medical history, conducts an evaluation, and engages in medical decision making.

- For beneficiaries eligible for the demonstration for more than three months, at least one additional visit from the demonstration practice in the home, an assisted living facility, or an office

The demonstration rules required that all patients of the IAH practice eligible for the demonstration be enrolled in the demonstration. Therefore, we required only one home visit for attribution to the IAH practice for beneficiaries eligible for the demonstration for three months or less. Some beneficiaries eligible for the demonstration for many months in a given year may have had only one visit with the IAH practice before returning to office-based primary care. To reduce the chance that the analysis sample would include beneficiaries who received only a single visit from the IAH practice, we required at least one additional visit from the practice for beneficiaries eligible for the demonstration for more than three months.

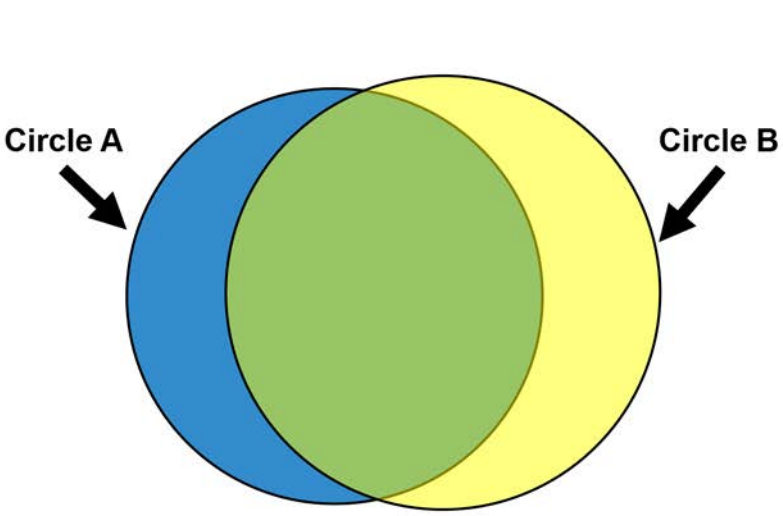
In each of the seven pre-demonstration and demonstration years, we refer to the beneficiaries who met eligibility criteria for IAH in administrative data and were attributed to a demonstration site as "Mathematica-eligible IAH beneficiaries" (or simply "IAH beneficiaries"). IAH beneficiaries were the treatment group for the evaluation. For a beneficiary to be in the IAH group for the evaluation, he or she had to meet the eligibility and attribution criteria outlined above according to Mathematica's analysis of Medicare enrollment, claims, and assessment data.

A beneficiary's enrollment (or non-enrollment) in the demonstration did not affect whether that person was in the IAH group for the evaluation. As described above, demonstration enrollment was based in part on data from the IAH practices, such as ADL limitations, chronic conditions, and residence in a long-term nursing home. In contrast, we excluded beneficiaries from the evaluation IAH group who were not eligible for the IAH demonstration and attributed to the IAH site according to administrative data (the part of Circle A excluding Circle B, or the blue crescent). We excluded those beneficiaries from the IAH group for two reasons: (1) we needed to identify the IAH group consistently in all study years, but demonstration enrollment data existed for the demonstration years only, not the pre-demonstration years; and (2) we could not replicate the enrollment process for comparison group members. In other words, we had no practice-reported data for identifying IAH beneficiaries in the pre-demonstration years, nor did we have such data for comparison group members in any year. Because our study design required that we use the same data sources to identify IAH and comparison beneficiaries in all years, we could not use practice-reported data to identify IAH beneficiaries in the demonstration years.

As shown in Exhibit A.3 and in the rest of this appendix, we use the term green oval to refer to beneficiaries who were enrolled in the demonstration and met its eligibility and attribution criteria in administrative data according to Mathematica's analysis of those data. We use yellow circle to refer to beneficiaries who met the eligibility and attribution criteria for the demonstration, regardless of whether they were enrolled in the demonstration. The yellow circle is the group we refer to as IAH beneficiaries (the treatment group for the evaluation). Enrollees who were not in the evaluation IAH group (the blue crescent) were those who were enrolled but

not confirmed as eligible for the demonstration or attributed to the IAH site according to administrative data.

Exhibit A.3. 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 who Mathematica identified as eligible and attributed to an IAH practice (IAH beneficiaries used for the evaluation)

Intersection of Circle A and Circle B (green oval) = IAH enrollees Mathematica identified as eligible and attributed to an IAH practice

Circle A without Circle B overlap (blue crescent) = IAH enrollees Mathematica did not identify as eligible and attributed to an IAH practice

Circle B without Circle A overlap (yellow crescent) = beneficiaries Mathematica identified as eligible and attributed to an IAH practice but who were not enrolled in IAH

Note: Figure not drawn to scale.

After we identified an IAH beneficiary, that beneficiary remained in the sample for the rest of the demonstration or pre-demonstration year unless the person died or left Medicare FFS. For example, if an IAH beneficiary became eligible for the demonstration in November 2016 (Month 2 of Year 5) and moved out of the IAH practice's geographic area or entered long-term care in April 2017, we continued to follow that beneficiary through the end of the study year (September 30, 2017 for all practices in Year 5).

Demonstration Year 1 (June 2012 – May 2013).[9] Mathematica identified 8,216 beneficiaries who met the demonstration eligibility criteria and were attributed to the demonstration practice during the first year (Exhibit A.4). This group represented the IAH group in the first year of the demonstration. It included 4,530 beneficiaries who were enrolled in the IAH demonstration according to the implementation contractor (the intersection of Circles A and B, the green oval in Exhibit A.3) and 3,686 beneficiaries not enrolled in the IAH demonstration in Year 1 (Circle B excluding Circle A, the yellow crescent). The analysis sample did not include the 2,405 beneficiaries whom the implementation contractor identified as enrollees but who we did not find eligible for the demonstration using administrative data (Circle A excluding Circle B, the blue crescent).

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

Exhibit A.4. Numbers of beneficiaries based on different identification processes

Demonstration year	Mathematica-eligible IAH beneficiaries[a]			IAH-enrolled only (blue crescent)
	Mathematica-eligible and IAH-enrolled (green oval)	Mathematica-eligible only (yellow crescent)	Total IAH group (all Mathematica-eligible regardless of enrollment, yellow circle)	
1	4,530	3,686	8,216	2,405
2	4,564	2,702	7,266	4,059
3	4,498	3,066	7,564	4,718
4	6,019	3,485	9,504	5,663
5	5,950	4,008	9,958	6,407

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

[a]Corresponds to the yellow circle in Exhibit A.3, which encompasses all Mathematica-eligible IAH beneficiaries (that is, those who met the demonstration eligibility criteria and were attributed to the demonstration practice).

IAH = Independence at Home.

Demonstration Year 2 (June 2013 – May 2014).[10] In Year 2, Mathematica identified 7,266 beneficiaries who met the demonstration eligibility criteria and were attributed to the demonstration practice. This group represented the IAH group in the second year of the demonstration. Of these 7,266 IAH beneficiaries, 4,564 were enrolled in the IAH demonstration in Year 2 (the green oval in Exhibits A.3 and A.4), and 2,702 beneficiaries were not enrolled (the yellow crescent). As in Year 1, the analysis sample for the evaluation did not include the 4,059 beneficiaries enrolled in the demonstration in Year 2 but who we did not find eligible for the demonstration using administrative data (the blue crescent).

Beneficiaries enrolled but not eligible and/or attributed according to Mathematica in Year 2 (the blue crescent in Year 2) included people who enrolled for the first time in Year 2. They also included two groups of beneficiaries who initially enrolled in Year 1 and continued to be enrolled in Year 2: those who were eligible and attributed according to administrative data in Year 1 (that is, those included in the yellow circle in Year 1) and those not eligible according to administrative data in Year 1 (the blue crescent in Year 1). The enrollment process did not require an individual who was enrolled in Year 1 to meet the qualifications for enrollment in Year 2.

The IAH group for the Year 2 analysis sample consisted of the 7,266 beneficiaries identified as eligible and attributed by Mathematica (the yellow circle in Exhibits A.3 and A.4). As explained previously, our method for measuring the effect of the demonstration required us to use the same data sources and approach to identify the IAH and comparison groups across all pre-demonstration and demonstration years. When we identified the Year 2 IAH beneficiaries, we

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

did not consider whether a beneficiary was in the IAH group, comparison group, or neither group in Year 1. Therefore, the Year 2 IAH group included beneficiaries who were in the analysis sample in Year 1 and requalified in Year 2 by meeting eligibility and attribution requirements, as well as people not in the analysis sample in Year 1. It did not include beneficiaries who were in the IAH group in Year 1 but did not requalify for the IAH group in Year 2 because they failed to meet eligibility or attribution requirements. Including beneficiaries who qualified for the IAH group in Year 1 but did not requalify in Year 2 would potentially bias our estimates of the effect of the demonstration in Year 2 because non-requalifying beneficiaries in Year 2 could differ from the IAH beneficiaries in Year 1 and the pre-demonstration years, all of whom were selected without regard to which beneficiaries were in the IAH group in the prior year.

Demonstration Year 3 (June 2014 – May 2015).[11] In Year 3, Mathematica identified 7,564 beneficiaries who met the demonstration eligibility criteria and were attributed to the demonstration practices. This group represented the IAH group in the third year of the demonstration. Of these 7,564 IAH beneficiaries, 4,498 were enrolled in the IAH demonstration in Year 3 (the intersection of Circle A and Circle B, or green oval, in Exhibits A.3 and A.4), and 3,066 were not enrolled (Circle B, excluding Circle A, or the yellow crescent). These 7,564 beneficiaries included people in the analysis sample in Years 1 or 2 and who requalified in Year 3 by meeting eligibility and attribution requirements, as well as people not in the analysis sample in either of the first two years. These beneficiaries could be new patients who met the eligibility criteria or patients who previously received care from the IAH practice and did not meet the eligibility criteria for the demonstration until Year 3.

As in demonstration Year 1, the analysis sample for the evaluation did not include the 4,718 beneficiaries on the implementation contractor's enrollment list in Year 3 but who we did not find eligible for the demonstration using administrative data (Circle A excluding Circle B, the blue crescent). Beneficiaries enrolled but not eligible according to Mathematica in Year 3 (the blue crescent) included those who enrolled for the first time in Year 3. Beneficiaries enrolled but not eligible according to Mathematica in Year 3 also included beneficiaries who initially enrolled in Years 1 or 2, continued to be enrolled in Year 3, but did not requalify for the demonstration in Year 3 because they failed to meet eligibility or attribution requirements.

Demonstration Year 4 (October 2015 – September 2016). In Year 4, Mathematica identified 9,504 beneficiaries who met the demonstration eligibility criteria and were attributed to the demonstration practices. This group represented the IAH group in the fourth year of the demonstration. Of these 9,504 IAH beneficiaries, 6,019 were enrolled in the demonstration (the intersection of Circles A and B, the green oval, in Exhibits A.3 and A.4), and 3,485 were not (Circle B excluding Circle A, the yellow crescent). These 9,504 beneficiaries included people who were in the analysis sample in Years 1, 2, or 3 and requalified in Year 4 by meeting

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

eligibility and attribution requirements, as well as people not in the analysis sample in any of the first three years.

As in Years 1 through 3, the analysis sample for the evaluation did not include the 5,663 beneficiaries enrolled in the IAH demonstration in Year 4 but who we found ineligible for and/or attributed to the demonstration using administrative data (Circle A excluding Circle B, or the blue crescent). The Year 4 IAH group also did not include beneficiaries who initially enrolled in Years 1, 2, or 3, continued to be enrolled in Year 4, but did not requalify for the demonstration in Year 4 because they failed to meet eligibility or attribution requirements.

The IAH group was substantially larger in Year 4 than in previous demonstration years. For all sites combined, this group increased 26 percent from Year 3 to Year 4. Five sites had increases of more than 20 percent from Year 3 to Year 4: Brooklyn, Durham, Dallas, Flint, and Portland. This increase may reflect the expansion of existing IAH practices. Brooklyn merged with another home-based primary care practice, and the Durham practice has expanded throughout North Carolina since the demonstration began. In Year 4, Dallas expanded into a new geographic area, and Flint added clinicians in its existing geographic area. Finally, Portland's sample size in Year 4 was larger than Year 3 but was about the same size as Year 1. The increase from Year 3 to Year 4 also could have been caused in part by some IAH practices participating in accountable care organizations (ACOs) in Year 4. Several practices did so in Year 4, including three of the five practices with the largest increases in sample sizes: Brooklyn, Dallas, and Flint. Other providers in the ACO may have referred some patients to the IAH practice. We discuss the implications of ACO participation in the limitations section of this appendix, Section VIII.

Demonstration Year 5 (October 2016 – September 2017). In Year 5, Mathematica identified 9,958 beneficiaries who met the demonstration eligibility criteria and were attributed to the demonstration practices. Of these 9,958 IAH beneficiaries, 5,950 were enrolled in the demonstration (the intersection of Circles A and B, the green oval, in Exhibits A.3 and A.4), and 4,008 were not (Circle B excluding Circle A, the yellow crescent). These 9,958 beneficiaries included people in the analysis sample in Years 1 through 4 who requalified in Year 5 by meeting eligibility and attribution requirements, as well as those not in the analysis sample in any of the first four years.

As in Years 1 through 4, the analysis sample for the evaluation did not include the 6,407 beneficiaries enrolled in the IAH demonstration in Year 5 but who we found ineligible for and/or attributed to the demonstration using administrative data (Circle A excluding Circle B, or the blue crescent). The Year 5 IAH group also did not include beneficiaries who initially enrolled in Years 1 through 4, continued to be enrolled in Year 5, but did not requalify for the demonstration in Year 5 because they failed to meet eligibility or attribution requirements.

The noticeable increase in size between Year 3 and Year 4, in which the IAH group increased by 26 percent, was not repeated in Year 5. Rather, the sample size increased by only 5 percent, consistent with the observed increase between Years 2 and 3 (4 percent). This stability suggests

that the observed increase in the overall size of the IAH sample in Year 4 was likely the result of events that may not reoccur in subsequent years, such as Brooklyn merging with another home-based primary care practice. This finding would be consistent with the proposed reasons for sample size increases outlined in the Year 4 summary above.

C. Reasons for the differences between demonstration enrollment and evaluation analysis cohorts

The evaluation analysis group identified by Mathematica and the enrolled group identified by the implementation contractor differed for two overarching reasons: (1) the use of different data sources and (2) the use of different analytic techniques. The primary reason Mathematica used different data sources and analytic techniques was that the implementation contractor was required to identify only an IAH group, whereas Mathematica was required to identify both an IAH group and a comparison group. Because Mathematica had to use the same procedures to identify both groups, and we could not obtain clinical data from the comparison group's primary care providers, we relied on administrative data alone when identifying the IAH group for the evaluation. This approach was in contrast to the implementation contractor, which used both administrative data and data from IAH practices to identify IAH enrollees. This difference resulted in Mathematica excluding some beneficiaries identified as enrollees by the implementation contractor and including some beneficiaries in the IAH group for the evaluation who were excluded by the implementation contractor. Kimmey et al. (2019) presents a detailed discussion regarding the differences in the samples for the evaluation versus enrollment; we highlight key findings in this section.

1. Reasons some IAH enrollees did not meet Mathematica's eligibility and/or attribution criteria

The use of different data sources was the primary reason some beneficiaries were excluded from the IAH group by Mathematica but were identified as enrollees by the implementation contractor. In each demonstration year, a majority of IAH enrollees not identified by Mathematica did not meet the ADL criterion because they had missing or insufficient ADL information in the assessment data. In contrast, the contractor used information provided by the IAH practices to determine whether a beneficiary required human assistance with at least two ADLs. In addition, the number of enrollees that Mathematica did not find eligible for the demonstration increased over time because beneficiaries remained on the IAH enrollment list from one year to the next regardless of whether they met IAH eligibility criteria again.

2. Reasons some beneficiaries found eligible and attributed by Mathematica were not enrolled

Among those who were in the IAH group for the evaluation but were not IAH enrollees, Mathematica identified three groups of beneficiaries:

- **Beneficiaries not found to be eligible by the implementation contractor based on administrative data:** As part of determining eligibility for enrolling in the demonstration, the contractor considered the dates that the beneficiary had a hospitalization, a rehabilitation services stay, and a home visit from the IAH practice; however, Mathematica considered only the dates of the qualifying hospitalization and rehabilitation services stay. Mathematica did not rely on the date of a home visit when measuring the 12-month period and setting the demonstration eligibility date because we could not replicate that requirement for the comparison group, who did not receive home-based primary care and therefore received no home visit.
- **Beneficiaries excluded from enrollment based on information from IAH practices:** The reason sites offered most frequently for excluding a beneficiary from enrollment was that the beneficiary did not meet the ADL criterion. The implementation contractor used information provided by the IAH practices to determine whether a given ADL required human assistance, which provided a more nuanced picture of ADL severity. Mathematica used only administrative data when identifying ADLs that required human assistance because information from clinicians was not available for the comparison group.
- **Beneficiaries who disenrolled from the demonstration:** An enrollee may voluntarily disenroll from the demonstration when he or she changes clinicians within the practice service area, is discharged by the practice, declines home care, or elects hospice and changes clinicians. If the beneficiary voluntarily disenrolled within six months of enrollment in the demonstration, the implementation contractor did not identify that beneficiary as an enrollee in the final enrollment list for a given year. Mathematica did not exclude a beneficiary who voluntarily disenrolled within six months because we could apply no such restriction to the comparison group.

IV. IDENTIFYING THE COMPARISON GROUP

In this section, we begin by describing how we used Medicare administrative data to identify a potential comparison group of beneficiaries who were eligible for the demonstration, lived in the same area as the IAH beneficiaries, and did not receive home-based primary care. Next, we present the methods and results of propensity-score matching. Finally, we present the number of IAH and comparison beneficiaries and eligible months in the evaluation sample.

A. Identifying the potential comparison group

To identify the potential comparison group beneficiaries, we relied on administrative data. We identified a set of potential comparison beneficiaries from each state in each year. We based our analyses on data for two pre-demonstration years and five demonstration years. Beneficiaries who had no visits to any of the demonstration practices in the study year and met all demonstration eligibility criteria were eligible to be in the potential comparison group for all sites in that state in that year. For example, a beneficiary who lived in Michigan, had no visits from any IAH practice, and met all demonstration eligibility criteria in Year 2 was in the potential comparison group for Flint and Lansing. We refer to these comparison groups as potential because we identified the final comparison groups using propensity-score matching (described later in Section III.B). Because we sought to compare beneficiaries who primarily received in-home physician care with those who did not receive such care, we excluded from the potential comparison group all beneficiaries who had two or more home visits from any clinician during or after their first month of eligibility through the end of the study year. In addition, we excluded all beneficiaries who had any visit from an IAH practice in the study year. As with the IAH beneficiaries, we did not assess whether potential comparison beneficiaries had home visits before the first month of eligibility.

In addition, to control for possible geographic variation in practice styles, access to services, and costs, we restricted our comparison groups to beneficiaries who lived in the zip codes served by the demonstration practices. The list of zip codes served by a demonstration practice in a given year reflected all zip codes where the practice's IAH beneficiaries lived in that year according to beneficiary address information in Medicare administrative data. For example, if a site operated in one state and had at least one IAH beneficiary who lived in each of 57 zip codes in that state during demonstration Year 1, the potential comparison group for that site in Year 1 included all beneficiaries who met demonstration eligibility requirements, had no visits to any demonstration practice in that year, had no more than one home-based primary care visit in that year, and lived in one of those zip codes. We used this zip code-based restriction for all practices in all years.

For the six practices located in states that had two demonstration practices (Brooklyn and Long Island, New York; Austin and Dallas, Texas; and Flint and Lansing, Michigan), some zip codes contained IAH beneficiaries for two practices. We could not simply restrict potential comparison beneficiaries only to those living in the zip codes represented by beneficiaries served by the IAH practice in a given year because it would have allowed a single potential comparison beneficiary

to be selected as a matched comparison for two IAH beneficiaries in different practices. In those cases, we identified the potential comparison group by conducting a preliminary propensity-score matching (using the same model to predict treatment status, as described below) to split the comparison sample into two potential comparison groups.[12] For each pair of sites located in the same state, we included in the preliminary model all of the IAH beneficiaries in those two sites as well as all beneficiaries in the comparison pool for both sites after applying the zip code restriction. Each comparison beneficiary was matched to an IAH beneficiary in one of the two sites; this process determined the site potential comparison pool to which the beneficiary was assigned. After using preliminary matching to split the overlapping comparison sample into two potential comparison groups (one group per practice), we matched IAH beneficiaries to the potential comparison group for each practice using the same approach as for other sites.

As with IAH beneficiaries, we again identified beneficiaries in the matched comparison group in demonstration Years 1, 2, 3, or 4 as potential comparison beneficiaries if they met all IAH eligibility requirements in Year 5.

B. Propensity-score matching methods

For each analysis year before and after the demonstration began, we used propensity-score matching to create a comparison group of nonparticipants similar in observable characteristics to IAH beneficiaries but who did not receive home-based primary care. The goals of matching were twofold. First, we sought to minimize nonrandom selection of individuals in the IAH group by constructing a matched comparison group that appeared similar to the treatment group on key observable characteristics that affect treatment status (receipt of home-based primary care from an IAH practice) and outcomes. Subject to that constraint, we then sought to maximize the size of the comparison group to increase statistical efficiency. For the IAH demonstration, key characteristics for matching included those that determined eligibility for the demonstration, as well as measures of health status, health trajectory, and other personal characteristics observable in administrative data that are predictive of health care expenditures. Limiting the comparison group to Medicare beneficiaries who closely matched the observed characteristics of the IAH group might also have reduced differences between the two groups on unobserved characteristics if those characteristics were correlated with matching variables.

We conducted matching for the entire IAH group, which consisted of beneficiaries who met the eligibility and attribution criteria based on administrative data (the yellow circle in Exhibits A.1 and A.2). For Year 5, for example, we matched 9,958 IAH beneficiaries on observable characteristics with beneficiaries who were similar and lived in the same geographic area but did not receive home-based primary care. We matched each site separately, including each member of the Mid-Atlantic Consortium. We created a comparison group for each practice by estimating

[12] Conducting a preliminary match typically provides a better match in both sites than using a random split because it ensures that the covariate distribution for the pool of eligible comparison beneficiaries is closely aligned with the covariate distribution for the treatment beneficiaries at each of the two practices.

a propensity-score equation using data for the IAH group and the potential comparison group, and then using the results to find the best matches for each IAH beneficiary.

We used demographics and health-related variables for matching beneficiaries in the IAH group with comparison beneficiaries. We used only one measure for exact matching: the number of months since the beneficiary's last inpatient admission (one, two or three, or four or more months). "Exact matching" means that an IAH beneficiary could be matched only to potential comparison beneficiaries who had the same value of that variable. We chose this measure for exact matching because expenditures and utilization—our key outcomes of interest—tend to be substantially higher in the months following a hospitalization. Preliminary data analyses indicated that adding other exact matching variables would likely result in dissimilarities on other key characteristics, such as disability. Therefore, we chose not to add other exact matching variables. We used two other measures related to eligibility for the demonstration as ordinary matching variables: (1) because a beneficiary can enter the sample at any time in a given year, we used a categorical measure of the month the beneficiary met eligibility criteria (Months 1, 2 to 6, or 7 to 12); and (2) because beneficiaries who had an observation stay may have been less acutely ill than those with an inpatient admission, we used whether the beneficiary had an observation stay but not an inpatient admission in the prior year (Exhibit A.5). We included the following demographic variables in the matching model but did not seek exact matches for them: age (younger than 65, 65 to 79, or 80 or older), gender, race, whether the beneficiary was dually eligible for Medicare and Medicaid, original reason for Medicare eligibility, and number of ADLs (two, three or four, or five or six). We used an indicator variable to identify beneficiaries with missing information for feeding assistance.

We used various measures of health status. We measured individual HCCs using each beneficiary's claims history for the 12 months before the date of eligibility for the demonstration in a given year. Beneficiaries who meet IAH eligibility criteria are at much higher risk of mortality in a given year than the average Medicare FFS beneficiary, and mortality can substantially affect expenditures in the year before death. To increase the likelihood that the comparison group was as similar as possible to the IAH beneficiaries in health status measures that predict mortality, we matched the IAH and comparison beneficiaries on risk factors for mortality. After reviewing the literature on mortality among Medicare beneficiaries, we selected chronic conditions or diagnoses that were significant predictors of mortality for use in matching. We included an HCC in the matching equation if Gagne et al. (2011) had identified a diagnosis code as predictive of mortality among elderly low-income Medicare beneficiaries. We collapsed several of the individual HCCs based on the type of condition, frequency in the IAH group, and a relative factor, the last of which represents the contribution of that HCC to the overall HCC risk

score.[13, 14, 15] We also used the risk score itself as a matching variable. Additional details about how we calculated the HCC score and indicators are available in Section IV of this appendix.

Exhibit A.5. Variables used in propensity–score matching equation

Variable
Eligibility and utilization
Number of months since most recent inpatient admission (1, 2 or 3, 4 or more)
Month of the demonstration year beneficiary met eligibility criteria (1, 2–6, 7–12)[a]
Whether beneficiary had an observation stay and no inpatient admission in prior 12 months
Demographic characteristics
Age: younger than 65, 65–79, 80 or older
Gender
Race: white, black, other or unknown
Dually eligible for Medicare and Medicaid
Original reason for Medicare entitlement: old age, ESRD or ESRD and disability, disability only
ADLs
Number of ADLs for which beneficiary requires human assistance (2, 3 to 4, 5 to 6)
Whether information about the feeding ADL was missing[b]
Health status
HCC risk score
Specific HCCs
HCC8: Metastatic cancer and acute leukemia[c]
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[c]
HCC52: Dementia without complications[c]
HCC54–55: Drug/alcohol psychosis; drug/alcohol dependence
HCC57–58: Schizophrenia; major depressive, bipolar, and paranoid disorders

[13] For example, we combined cirrhosis of the liver (HCC28) and chronic hepatitis (HCC29) into a single indicator for matching but did not combine them with end-stage liver disease (HCC27). Less than 2 percent of the treatment group had cirrhosis of the liver or chronic hepatitis; the relative factor for those conditions was less than half of the relative factor for end-stage liver disease.

[14] Table 9 of the Announcement of Calendar Year (CY) 2012 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies and Final Call Letter lists the relative factor for each HCC. Available at https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Announcement2012.pdf.

[15] We used software version V2117 to calculate HCC scores for beneficiaries in Year 5, which incorporated version 10 of the *International Classification of Diseases*.

Exhibit A.5 (*continued*)

Variable
HCC70–71: Quadriplegia; paraplegia
HCC72: Spinal cord disorders/injuries
HCC85: Congestive heart failure[c]
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[c]
HCC136–138: Chronic kidney disease, stages 3–5[c]
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[d]
Anemia
Fluid and electrolyte disorders
Number of chronic conditions (2–5, 6–9, 10 or more)[d]
Whether beneficiary had a complicating condition or major complicating condition during the most recent inpatient admission
Chronically critically ill or medically complex diagnosis

Note: "Exact matching" means that an IAH beneficiary can be matched only to a potential comparison beneficiary with the same characteristic. An ordinary matching variable is one used as an independent variable in the matching regression equation.

[a]For pre-demonstration years and Years 1–3, Month 1 was June or September. For sites that began the demonstration in June 2012, Month 1 was June. For sites that began the demonstration in September 2012, Month 1 was September. All sites began Years 4 and 5 in October 2015 and October 2016, respectively.

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

[c]Identified as a key predictor of mortality by Gagne et al. (2011); they are the measures of health status we prioritized most highly when determining which of several alternative matched comparison groups was most appropriate for a particular site in a particular year.

[d]Chronic condition categories measured by the Chronic Conditions Warehouse.

ADL = activity of daily living; ESRD = end-stage renal disease; HCC = hierarchical condition category; IAH = Independence at Home.

In addition to the HCCs included in the matching equation based on Gagne et al. (2011), we included an HCC indicator of pressure ulcers because a large share of the IAH population has poor functional status and may be at higher than average risk for a pressure ulcer. We included three other conditions not measured by HCCs: anemia, depression, and electrolyte disorders. Gagne et al. (2011) identified anemia and electrolyte disorders as predictive of mortality.

We included two other measures of health status using diagnosis codes from the beneficiary's most recent inpatient admission in the past year. The first measure indicated whether the Medicare Severity Diagnosis Related Group included a complicating condition or major complicating condition. The second measure indicated whether, according to the diagnosis in the claim, the beneficiary was chronically critically ill or medically complex (Kandilov et al. 2014).

C. Results of propensity-score matching

The standardized difference in means is a standard statistic used to assess similarities between the treatment group and the final matched comparison group (Stuart 2010). The literature suggests that a standardized difference of less than 0.25 is an appropriate threshold for determining that the treatment and comparison groups are well matched on a particular variable (Rubin 2001). We applied a more stringent standard of 0.10 for our matching. We examined the matching results for both of the variables used in the matching algorithm and the variables that might be important to control for but could not be included, such as individual HCCs aggregated with other HCCs in the matching equation (for example, cirrhosis of the liver and chronic hepatitis), and individual chronic conditions measured by the Chronic Conditions Warehouse.

Across all 14 sites together (treating the three Mid-Atlantic Consortium sites as one), the absolute value of the standardized difference in the fifth demonstration year was less than 0.10 on all matching variables and less than 0.10 on all but one nonmatching variable (Exhibit A.6). All 14 sites individually had standardized differences of less than 0.25 on all of the matching variables; for 12 of those sites, the standardized differences were also less than 0.10 on all of the matching variables (data not shown). Furthermore, 8 of the sites had standardized differences of less than 0.25 on all of the nonmatching variables.

Exhibit A.6. Characteristics of potential comparison beneficiaries, matched comparison beneficiaries, and IAH beneficiaries, Year 5

Variable	Potential comparison group	Matched comparison group	IAH beneficiaries	Standardized difference
Eligibility for the demonstration				
Proportion with number of months since most recent inpatient admission[a]				
One	0.575	0.389	0.389	0.000
Two or three	0.157	0.173	0.173	0.000
Four or more	0.268	0.438	0.438	0.000
Proportion with month of the demonstration year that beneficiary met eligibility criteria[b]				
Month 1	0.424	0.657	0.652	-0.009
Months 2–6	0.293	0.211	0.215	0.008
Months 7–12	0.283	0.132	0.133	0.004

Exhibit A.6 *(continued)*

Variable	Potential comparison group	Matched comparison group	IAH beneficiaries	Standardized difference
Proportion with observation stay and no inpatient admission in previous 12 months	0.056	0.095	0.086	-0.031
Demographic characteristics				
Female	0.613	0.657	0.662	0.010
Age				
Proportion younger than 65	0.129	0.156	0.161	0.014
Proportion 65–79	0.425	0.328	0.328	0.001
Proportion 80 or older	0.446	0.516	0.511	-0.011
Race and ethnicity				
Proportion white	0.753	0.707	0.708	0.002
Proportion black	0.186	0.241	0.240	-0.001
Proportion other	0.062	0.052	0.051	-0.004
Proportion dually eligible for Medicare and Medicaid	0.257	0.381	0.386	0.010
Original reason for Medicare entitlement				
Proportion whose original eligibility was due to age	0.736	0.670	0.668	-0.005
Proportion whose original eligibility was due to disability	0.243	0.317	0.320	0.008
Proportion whose original eligibility was due to ESRD or ESRD plus disability	0.022	0.013	0.012	-0.010
ADLs				
Proportion with two ADLs	0.124	0.073	0.072	-0.005
Proportion with three or four ADLs	0.309	0.314	0.307	-0.014
Proportion with five or six ADLs	0.566	0.613	0.621	0.016
Proportion missing information about feeding ADL	0.091	0.166	0.172	0.019
Health status				
HCC risk score	3.604	4.020	4.064	0.023
Proportion with HCCs				
HCC8: Metastatic cancer	0.046	0.017	0.016	-0.008
HCC9–10: Lung, lymphoma, and other cancers	0.059	0.035	0.035	0.002

Exhibit A.6 *(continued)*

Variable	Potential comparison group	Matched comparison group	IAH beneficiaries	Standardized difference
HCC11–12: Colorectal, bladder, breast, prostate, and other cancers	0.107	0.086	0.083	-0.011
HCC18: Diabetes with chronic complications	0.359	0.381	0.374	-0.014
HCC21: Protein-calorie malnutrition	0.139	0.197	0.209	0.031
HCC27: End-stage liver disease	0.018	0.012	0.011	-0.005
HCC28–29: Cirrhosis of liver and chronic hepatitis	0.027	0.025	0.024	-0.005
HCC46: Severe hematological disorders	0.018	0.013	0.013	0.001
HCC48: Coagulation defects and other specified hematological disorders	0.172	0.138	0.135	-0.009
HCC51: Dementia with complications	0.068	0.167	0.177	0.028
HCC52: Dementia without complications	0.192	0.322	0.309	-0.028
HCC54–55: Drug/alcohol psychosis and drug/alcohol dependence	0.072	0.077	0.078	0.004
HCC57–58: Schizophrenia, major depressive, bipolar, and paranoid disorders	0.184	0.279	0.283	0.010
HCC70–71: Quadriplegia, paraplegia	0.027	0.071	0.081	0.041
HCC72: Spinal cord disorders/injuries	0.024	0.016	0.016	-0.005
HCC85: Congestive heart failure	0.450	0.518	0.513	-0.011
HCC96: Specified heart arrhythmias	0.389	0.366	0.360	-0.014
HCC103–104: Hemiplegia/hemiparesis, monoplegia, other paralytic syndromes	0.102	0.141	0.140	-0.002
HCC106: Atherosclerosis of the extremities with ulceration or gangrene	0.046	0.046	0.047	0.004
HCC107–108: Vascular disease with or without complications	0.429	0.504	0.499	-0.010
HCC111: Chronic obstructive pulmonary disease	0.329	0.372	0.371	-0.003
HCC134: Dialysis status	0.058	0.040	0.038	-0.010

Exhibit A.6 *(continued)*

Variable	Potential comparison group	Matched comparison group	IAH beneficiaries	Standardized difference
HCC136–138: Chronic kidney disease, stages 3–5	0.091	0.111	0.113	0.008
HCC139–140: Chronic kidney disease stages 1–2, unspecified renal failure	0.048	0.066	0.067	0.006
HCC157–159: Pressure ulcer of skin with necrosis or skin loss	0.081	0.166	0.181	0.044
Number of chronic conditions measured by Chronic Conditions Warehouse				
Fewer than six	0.185	0.111	0.115	0.012
Six to nine	0.488	0.463	0.457	-0.012
More than nine	0.327	0.425	0.427	0.004
Proportion with anemia[c]	0.182	0.187	0.189	0.005
Proportion with depression	0.436	0.563	0.567	0.008
Proportion with fluid and electrolyte disorders[c]	0.385	0.409	0.404	-0.010
Proportion with diagnosis of chronically critically ill or medically complex[d]	0.320	0.321	0.322	0.002
Proportion with complicating condition or major complicating condition during the most recent inpatient admission	0.574	0.579	0.579	0.001

Source: Medicare claims and enrollment data for 2009–2017 obtained from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in demonstration Year 5.

Notes: The final sample sizes in Year 5 were 9,958 IAH beneficiaries and 41,387 matched comparison beneficiaries. The number of weighted matched comparison beneficiaries equaled the number of IAH beneficiaries.

[a]Variable used for exact matching.

[b]"Month" refers to the first month in the demonstration year after the beneficiary met eligibility criteria. For example, if a beneficiary had a qualifying admission and rehabilitation services in one or more months before the demonstration, the Month 1 group included that person. For all sites in Year 5, Month 1 was October.

[c]Measured using claims from the most recent inpatient and observation stays in the year before the demonstration eligibility date. Diagnosis codes for these conditions were drawn from Gagne et al. (2011).

[d]Measured using diagnoses from the most recent inpatient stay in the year before the demonstration eligibility date. Diagnoses were drawn from Kandilov et al. (2014).

ADLs = activities of daily living; ESRD = end-stage renal disease; HCC = hierarchical condition category; IAH = Independence at Home.

As in Year 5, the IAH and matched comparison groups were very similar in each of the first four demonstration years. Across all sites together in each of the first four years, the absolute value of the standardized difference was less than 0.10 on all matching variables and less than 0.25 on all variables not included in matching. At an individual site level, all 14 sites in Years 1 to 4 had standardized differences of less than 0.25 on all of the matching variables; most of the 14 sites had standardized differences of less than 0.10 on all of the matching variables.

D. Number of beneficiaries and eligible months

Beneficiaries in both the IAH group and comparison group were analyzed from the month they became eligible for the demonstration and observed for the remaining months in a given demonstration year. Over the five years of the demonstration, the number of IAH beneficiaries varied; for each IAH beneficiary, we matched up to five comparison beneficiaries. On average, each IAH beneficiary was matched to four comparison beneficiaries. Across the demonstration years, the average number of eligible months for the comparison beneficiaries was slightly smaller than among the IAH beneficiaries (Exhibit A.7). This difference arose because the comparison beneficiaries were more likely to die within one year of the eligibility date than the IAH beneficiaries, and the IAH beneficiaries were more likely to qualify for the demonstration earlier in the 12-month period than the comparison beneficiaries. To address any possible concerns that this difference might cause, we incorporated an eligibility fraction into the weighting design for regressions, where the eligibility weight reflected the number of months eligible for the demonstration in a given year. For example, a beneficiary eligible for the demonstration for 6 months in Year 5 had half the weight of a beneficiary eligible for the demonstration for 12 months in Year 5. Using an eligibility fraction in the weight ensured that each beneficiary's contribution to the estimation was proportionate to how long we observed that person during a given year. In addition, we added two control variables: number of months since most recent inpatient admission and month of the demonstration year that the beneficiary met the eligibility criteria. In this way, we controlled for differences in the time between when beneficiaries met the service utilization criteria required for demonstration eligibility and their eligibility date. Those who qualified in the first month may have met both of the service utilization criteria up to one year before the demonstration year began, whereas those who qualified in later months met at least one of the two service utilization criteria in the month immediately before the eligibility date. Section VI of this appendix provides additional details about weights and control variables.

Exhibit A.7. Analysis sample, by years

	Two years before the demonstration	One year before the demonstration	Year 1	Year 2	Year 3	Year 4	Year 5
Number of IAH beneficiaries	6,837	7,367	8,216	7,266	7,564	9,504	9,958
Total number of eligible months for IAH beneficiaries	65,781	70,591	79,396	69,768	72,215	90,223	95,003
Average number of eligible months per IAH beneficiary	9.6	9.6	9.7	9.6	9.5	9.5	9.5
Number of comparison beneficiaries	29,517	31,888	33,916	32,248	31,259	38,365	41,387
Total number of eligible months for comparison beneficiaries	264,558	286,314	303,770	293,081	278,015	335,250	363,251
Average number of eligible months per comparison beneficiary	9.0	9.0	9.0	9.1	8.9	8.7	8.8

Source: Medicare claims and enrollment data for 2009–2017 obtained from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in demonstration Year 5.

IAH = Independence at Home.

V. MEDICARE DATA AND MEASURES

In this chapter, we describe the data sources and measures we used in our analyses of the effect of the demonstration.

We constructed our yearly analytic files with observations at the beneficiary year level. We drew data for determining demonstration eligibility and measuring outcomes in the analytic files from several sources (Exhibit A.8). We accessed all data through the Chronic Conditions Warehouse Data Enclave.

Exhibit A.8. Data sources

Data	Demographic characteristics	Chronic conditions	Activities of daily living	Service use: Demon-stration eligibility	Service use: Outcome measures	Unintended consequences
Medicare enrollment database	X					X
Master beneficiary summary file		X				
Inpatient claims				X	X	
Outpatient claims				X	X	
Physician or supplier claims				X	X	
Home health agency claims				X	X	
Skilled nursing facility claims				X	X	
Hospice claims				X	X	
Durable medical equipment claims					X	
Inpatient rehabilitation facility–patient assessment instrument[a]			X			

Data	Demographic characteristics	Chronic conditions	Activities of daily living	Service use: Demon-stration eligibility	Service use: Outcome measures	Unintended consequences
Minimum data set			X		X	
Outcome and assessment information set			X			
Timeline file[b]						X

[a]Includes inpatient rehabilitation hospitals and rehabilitation units. Excludes long-term care hospitals.
[b]Used to measure whether a beneficiary entered institutional long-term care.

A. HCC score and indicators

To account for differences in health status and the differential risks of incurring high Medicare expenditures, we used the CMS-HCC risk-adjustment model to create HCC scores and indicators (Exhibit A.9). To estimate the HCC scores, we used a 12-month look-back period for Medicare claims to obtain diagnosis information. Because the claims-based eligibility dates for IAH and comparison beneficiaries can vary for a specific pre-demonstration or demonstration year, the 12-month look-back period also varied, depending on the beneficiaries' eligibility dates. For each beneficiary in the IAH and comparison groups, we estimated the HCC score by using the publicly available HCC software (CMS 2017) and information on demographics, Medicare eligibility, and dual eligibility status, as well as Medicare claims for the 12 months before the person's claims-based eligibility date. We used fewer than 12 months of Medicare claims if a beneficiary was not enrolled in Medicare for all 12 months. We used Version 21 of the HCC model, which was developed and calibrated for the Program of All-Inclusive Care for the Elderly (PACE) population, because that population resembles the IAH-eligible population in being sicker and frailer than the average Medicare beneficiary.

Exhibit A.9. Measures of Medicare expenditures, service utilization, and unintended consequences used in regressions

Measure
Medicare expenditures per beneficiary per month
Total
Inpatient
Home health service[a]
Outpatient
Skilled nursing facility
Physician or supplier
Hospice
Durable medical equipment
Service utilization
Number of hospital admissions per beneficiary per year[b]
Number of potentially avoidable hospital admissions per beneficiary per year (AHRQ PQI)b
Number of ED visits per beneficiary per year[c]
Number of potentially avoidable outpatient ED visits per beneficiary per year (AHRQ PQI)c
Percentage of beneficiaries with a qualifying index discharge and an unplanned readmission within 30 days of discharge in the year[d]
Unintended consequences
Death within the study year
Entry into institutional long-term care

Notes: Measures were constructed using data from the date the beneficiary became eligible in the demonstration year through the end of that demonstration year. Following the CMMI Priority Measures for Monitoring and Evaluation, we did not truncate expenditure measures; rather, we risk-adjusted, annualized, and weighted them to reflect partial year observations. We did not price standardize the expenditure measures.

[a]Total home health expenditures include all care provided under the home health benefit. Claims for therapy appear only in the outpatient file.

[b]Includes inpatient admissions and observation stays.

[c]Measured as specified in the CMMI Priority Measures for Monitoring and Evaluation.

[d]Eligible index discharges for the numerator of the readmission measure include index discharges for patients who were enrolled in Medicare FFS, discharged from nonfederal acute care hospitals, alive at the time of discharge, and not transferred to another acute care facility. Home-based primary care and the demonstration might affect whether a beneficiary has an eligible index discharge in a particular year. Such an effect could lead to estimating biased rates of readmission for the IAH and comparison groups if readmission is defined only for beneficiaries who had an eligible index discharge, as recommended by the CMMI Priority Measures for Monitoring and Evaluation. Thus, we defined the readmission measure using all beneficiaries in the denominator, rather than limiting it to beneficiaries with an eligible discharge. For example, if home-based primary care or the demonstration reduces the likelihood of having an eligible index discharge, IAH beneficiaries who have such a discharge might be sicker on average than comparison beneficiaries who have such a discharge. Being sicker could lead to an increased risk of readmission.

AHRQ = Agency for Healthcare Research and Quality; CMMI = Center for Medicare & Medicaid Innovation; ED = emergency department; FFS = fee-for-service; IAH = Independence at Home; PQI = Prevention Quality Indicator.

CMS has separate HCC models for beneficiaries residing in the community and those residing in an institution. We used the HCC score estimated by the community model for all beneficiaries in our sample. Beneficiaries cannot reside in an institution when they become eligible for the demonstration, so we did not use scores predicted by the institutional model for any beneficiary. We also did not use the demographics-only model for new enrollees. Given the service use requirements for the demonstration, all IAH-eligible beneficiaries had some claims history during the previous 12 months. Using any available diagnoses information in the HCC model should have provided a score that captured health status better than a demographics-only model. The specific scale of the HCC score should not have affected propensity-score matching if the score was estimated similarly for both IAH and potential comparison beneficiaries; thus, we did not normalize or rescale HCC scores. We did not apply any frailty factors to the HCC scores because (1) we did not have survey-based ADL measures that calculate plan-level frailty factors for the PACE population and (2) we could not apply plan- or practice-specific frailty factors to the comparison group in this case. However, we included indicators for the number of ADLs with which the beneficiary needed human assistance as control variables in all regressions.

B. Dual eligibility

When we did propensity-score matching for the full sample in all demonstration and pre-demonstration years, we measured dual eligibility using the monthly Part A and Part B state buy-in variables on Medicare enrollment data because Medicaid enrollment data were not available promptly enough for us to define dual eligibility using those data. If a beneficiary had state buy-in for Part A, Part B, or both in any month in a pre-demonstration or demonstration year, we identified that person as being dually eligible in that year. We used the same measure of dual eligibility as a control variable in the regression models for Medicare expenditures and other Medicare claims-based outcomes.

C. Outcome variables

We used three groups of measures for the regression analysis of outcomes in the demonstration based on Medicare Part A and Part B claims, as well as the Medicare enrollment database: (1) Medicare expenditures, (2) Medicare service utilization, and (3) unintended consequences (Exhibit A.9). We measured these outcomes for the number of months a beneficiary was observed in a study year starting with the first day of the first month after the beneficiary met all eligibility criteria in each year based on our analysis of Medicare enrollment and administrative data.

We measured all claims-based outcomes at the beneficiary level in that particular study year. For expenditures, we measured each outcome PBPM. For example, if a beneficiary was alive and in Medicare FFS for four months from the demonstration eligibility date through the end of the year, we divided expenditures during those four months by four to measure expenditures PBPM. We annualized claims-based outcomes other than expenditures and binary measures (such as the likelihood of unplanned readmission or mortality). For example, if a beneficiary had four hospital admissions and an eligibility weight of 0.5 (because he or she was eligible for the

demonstration for 6 of 12 months in the demonstration year), the annualized number of hospital admissions would be eight.

Potentially avoidable hospitalizations and outpatient emergency department (ED) visits. Potentially avoidable hospital use occurs when ambulatory care might have prevented or reduced the need for a hospital admission or ED visit. We measured a beneficiary as having a potentially avoidable hospitalization or ED visit if the principal diagnosis for the hospitalization or ED visit was an ambulatory care-sensitive condition (ACSC). We based our definition of ACSCs on the Agency for Healthcare Research and Quality Prevention Quality Indicator 90, which includes the following conditions: diabetes short-term complications, diabetes long-term complications, uncontrolled diabetes, lower-extremity amputation among diabetics, chronic obstructive pulmonary disease or asthma in older adults, hypertension, heart failure, angina without procedure, dehydration, bacterial pneumonia, and urinary tract infection. The measure of potentially avoidable ED visits included only outpatient ED visits—that is, ED visits not accompanied by an admission. We excluded ED visits that led to an inpatient admission because the principal diagnosis on the inpatient claim would not necessarily be the ACSC leading to the ED visit.[16]

ED visits. Our primary measure of emergency care was total number of ED visits. However, to better understand the results of the effect of the demonstration on total ED visits, we also used two other measures of such visits: (1) those that led to an inpatient admission and (2) outpatient ED visits (including those visits that led to an observation stay). We used these measures because the demonstration could have different effects on the two types of ED visits. One that led to an admission might suggest that the beneficiary was more seriously ill than when such a visit did not lead to an admission. The measure of outpatient ED visits included cases in which a beneficiary was transferred to a different hospital for admission and might have included some cases in which a hospital billed ED and inpatient services separately.

Unplanned readmission within 30 days of discharge. The unplanned readmission measure indicated whether the beneficiary had at least one unplanned readmission within 30 days of an eligible index discharge. Eligible index discharges for the readmission measure included index discharges from nonfederal acute care hospitals for patients who were enrolled in Medicare FFS, alive at the time of discharge, and not transferred to another acute care facility. The eligible index discharges included patients discharged to nonacute care settings. Index discharges did not include admissions to Prospective Payment System-exempt cancer hospitals or admissions for patients without at least 30 days of post-discharge enrollment in FFS Medicare Parts A and B

[16] ED visits appear in Medicare inpatient and outpatient claims. A beneficiary whose ED visit led to a hospital admission would not have a separate claim in the outpatient file; the ED claim would be part of the hospital claim in the inpatient file, which would have diagnoses that reflect the hospital stay. Using inpatient claims to measure potentially avoidable ED visits that led to hospital admission poses two problems. First, the diagnosis that led someone to the ED may be different from the diagnoses on the inpatient claim (for example, a beneficiary visits the ED because of shortness of breath but is later admitted due to another underlying factor). Second, hospital admissions with potentially avoidable diagnoses are counted in the potentially avoidable hospital admission measure. If we also counted them as potentially avoidable ED visits, we would double-count the utilization.

(unless a patient was enrolled in FFS but died within 30 days), patients discharged against medical advice, primary psychiatric diagnoses, rehabilitation, and medical treatment of cancer.

We excluded planned readmissions from this measure. To identify them, we followed the approach used by CMS's hospital-level 30-day risk-standardized readmission measure, developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (2015). Unlike the Yale measure, our list of procedure codes to identify planned readmissions did not include codes that apply only to all-payer populations.

All beneficiaries who had an eligible index discharge and an unplanned readmission within 30 days were identified as having an unplanned readmission. Therefore, the measure provided an estimate of the combined effect of the demonstration on whether a patient had an eligible index discharge and, if so, whether he or she had an unplanned readmission within 30 days.

Entry into institutional long-term care. The institutional long-term care measure identified beneficiaries who had at least one episode of long-term care spanning 90 or more days during a given study year. An episode of long-term care began when a beneficiary entered a skilled or unskilled nursing facility and ended when the beneficiary spent more than 14 consecutive days in the community, or the study year ended. We evaluated each study year (demonstration or pre-demonstration year) separately. In other words, we required a beneficiary to have a 90-day episode of institutional long-term care during a single study year to identify that person as entering long-term care in that study year. Beneficiaries in long-term care for the entire time they were otherwise eligible for the demonstration in a given year could not be in the IAH group for the evaluation in that year.

We created this measure using the Timeline file, which combines data from claims (inpatient, SNF, and home health) and assessment data (MDS and OASIS) to flag a beneficiary's residency status for each day of a calendar year. The daily residency flag can contain one of the following values: I (inpatient), S (SNF), M (MDS, which includes nursing home days not paid by Medicare), C (community, which includes days identified by home health claims or OASIS assessment data and days with no claims or assessment data), D (dead), or blank (not Medicare eligible). We considered all days flagged with a C, H, O, or blank to be community days. Timeline data were available only through 2016, so we calculated the long-term care measure through Year 4.

Potential episodes of long-term care began when the beneficiary entered a long-term care institution (a daily status of S or M) in a given study year. Inpatient days that occurred during a potential episode of long-term care were considered part of the long-term care episode. However, an episode of long-term care could not begin with an inpatient stay. For example, an inpatient day that immediately preceded the beneficiary's first SNF or MDS day did not count toward an episode of long-term care, but an inpatient day that occurred the day after a SNF day did. We counted days in the community that occurred during an episode of long-term care toward the 90-day requirement as long as there were no more than 14 consecutive community days and the

beneficiary re-entered an institution—a daily status of S, M, or I—on or before a 15th community day.

VI. ESTIMATION OF DEMONSTRATION IMPACTS

We used a difference-in-differences model to estimate the impact of IAH in each demonstration year and the average annual impact for the five-year demonstration period. Our difference-in-differences impact estimate measured the difference in a given outcome between the year before the demonstration started and any demonstration year for IAH beneficiaries relative to the difference during the same period for comparison beneficiaries. We implemented the difference-in-differences model using two approaches—a frequentist model and a Bayesian analysis. In this section, we describe the specifications and assumptions of each analysis, the model specifications we employed for different outcomes, and the methods we used to account for clustering.

A. Frequentist difference-in-differences model

1. Model specification for continuous and count outcomes

We estimated the impacts of the demonstration by comparing the regression-adjusted differences in outcomes between the IAH treatment and comparison groups in the pre- and post-demonstration periods. We used a difference-in-differences estimation strategy to test for significant differential changes in all claims-based outcomes between the IAH and comparison groups during the two pre-demonstration years and the first five years of the demonstration. Equation (1) shows the model we estimated for each outcome:

$$(1)\quad Y_{it} = \alpha + X_{it}\beta + \tau \cdot treatment_{it} + \gamma_{-1}\, PD_1 + \gamma_1\, DY_1 + \gamma_2\, DY_2 + \gamma_3\, DY_3 + \gamma_4\, DY_4 + \gamma_5\, DY_5 + \theta_{-1} treatment_{it} \cdot PD_1 + \theta_1 treatment_{it} \cdot DY_1 + \theta_2 treatment_{it} \cdot DY_2 + \theta_3 treatment_{it} \cdot DY_3 + \theta_4 treatment_{it} \cdot DY_4 + \theta_5 treatment_{it} \cdot DY_5 + \varepsilon_{it}$$

where Y_{it} is the claims-based outcome measured for a beneficiary i in year t; α is a constant term; X_{it} is a set of beneficiary characteristics measured in the index year; PD_1 is an indicator for pre-demonstration Year 1—that is, two years before the start of the demonstration, with the year immediately preceding the demonstration serving as the reference or omitted category; $DY_1 - DY_5$ are a set of indicators for each post-demonstration year; $treatment_{it}$ is an indicator variable for being in an IAH practice; and ε_{it} is a random error term. As we describe in Section VI.A.6 below, the set of beneficiary characteristics included in X_{it} were largely the same as the variables used for matching; they controlled for any remaining differences between the IAH and matched comparison groups in these characteristics.

The key parameters are $\theta_1 - \theta_5$, which constitute the difference-in-differences coefficients; they are the change of outcome from the year before the demonstration to each year after the demonstration for the IAH group, net of the change in outcome for the comparison group during the same period. Separate estimates for each year (that is, one θ per year) allowed for nonlinearities in such trends. Last, the parameter θ_{-1} captures the differential change in outcome between the IAH and matched comparison groups during the two pre-intervention years. We use

θ_{-1} to examine whether the two groups were on the same outcome trajectories before the demonstration (see Section VI.A.7 for details).

In cases where we estimated a linear model, such as total Medicare expenditures, the difference-in-difference coefficients ($\theta_1 - \theta_5$) equaled the difference-in-differences impact estimates. In cases where we used non-linear models, such as a negative binomial regression for the number of hospital admissions, we transformed $\theta_1 - \theta_5$ into difference-in-differences impact estimates using the following steps, using the estimated impact in Year 5 as an example:

1. Using the coefficients obtained from equation (1), we calculated the average outcomes for IAH treatment and comparison groups in each year. We adjusted the yearly average outcomes for both groups to reflect the covariate distribution of the IAH group in the latest demonstration year (Year 5). For example, we used the mean covariate values of the Year 5 IAH group to generate two estimates of predicted total Medicare expenditures in the year before the demonstration: one estimate assumed that beneficiaries received home-based primary care in that year (the IAH treatment group estimate), and one assumed that beneficiaries did not receive home-based primary care in that year (the comparison group estimate).
2. We calculated the difference of the regression-adjusted outcome for the IAH group and matched comparison group in Year 5.
3. We calculated the change in the difference between the IAH and matched comparison group in Year 5 relative to the difference in the year before the demonstration. We refer to this estimate as the difference-in-differences impact estimate.

Our difference-in-differences impact estimates measured the change between two differences: the pre- and post-demonstration difference for IAH beneficiaries, and the pre- and post-demonstration difference for comparison beneficiaries. This method isolated the impact of the demonstration by accounting for two factors that affected outcomes. First, it accounted for the difference in outcomes between IAH and comparison beneficiaries before the demonstration, controlling for differences in observed beneficiary characteristics. Second, it accounted for changes in outcomes during the demonstration caused by factors unrelated to the demonstration that affected both IAH and comparison beneficiaries over time.

In addition to estimating the yearly impact, we estimated a separate difference-in-differences model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with IAH status to obtain an average difference-in-differences estimate over the five post-demonstration years. As Equation (2) shows, DY_{it} is an indicator for the demonstration period, where $t = 1$ in demonstration Years 1 through 5 (and 0 otherwise). This model provided a measure of the impact of the demonstration, if any, during the entire demonstration period considered as a whole, by averaging across all of the yearly observations for the demonstration years, as shown in Equation (2), where the average difference-in-differences impact estimate over all demonstration years is given by θ_1.

$$(2)\quad Y_{it} = \alpha + X_{it}\beta + \tau \cdot treatment_{it} + \gamma_{-1}\, PD_1 + \gamma_1\, DY_{it} + \theta_{-1} treatment_{it} \cdot PD_1 + \theta_1 treatment \cdot DY_{it} + \varepsilon_{it}$$

In addition to reporting all difference-in-differences estimates in absolute terms, we also calculated the impacts in percentage terms by dividing the impact estimate for an outcome by the unadjusted IAH group mean for that same outcome in the year before the demonstration. The percentage impact helped us to interpret whether the absolute impact in a given year was likely to be meaningful.

We used linear regressions for expenditures. We used negative binomial regressions for the number of hospital admissions and ED visits to account for over-dispersion of counts, and zero-inflated negative binomial regressions for the number of potentially avoidable hospital admissions and outpatient ED visits to account for both over-dispersion and the large percentage of beneficiaries with no utilization during the time period.

For all outcomes, we adjusted standard errors for clustering at the practice level for the IAH group and at the beneficiary level for the comparison group (which we refer to below as the hybrid clustering). We estimated the effect of the demonstration on all outcomes using two weighting schemes, which we refer to as beneficiary weighting and practice weighting. Sections V.A.4 and V.A.5 describe clustering and weighting in detail.

2. Model specification for mortality

We used survival modeling techniques to estimate whether the demonstration had an effect on the probability of a beneficiary dying within the demonstration year. The advantage of this approach relative to a logistic regression model is that it allowed us to use a flexible functional form to account for some beneficiaries becoming eligible after the beginning of the demonstration year, and thus having shorter periods of observation relative to other beneficiaries. We used the accelerated failure time (AFT) hazard specification to estimate a survival-time model in Equation (3) as follows:

$$(3)\quad \log(T_{it}) = X_{it}\beta + \tau \cdot treatment_{it} + \gamma_{-1}\, PD_1 + \gamma_1\, DY_1 + \gamma_2\, DY_2 + \gamma_3\, DY_3 + \gamma_4\, DY_4 + \gamma_5\, DY_5 + \theta_{-1} treatment_{it} \cdot PD_1 + \theta_1 treatment_{it} \cdot DY_1 + \theta_2 treatment_{it} \cdot DY_2 + \theta_3 treatment_{it} \cdot DY_3 + \theta_4 treatment_{it} \cdot DY_4 + \theta_5 treatment_{it} \cdot DY_5 + z_{it}$$

where T_{it} denotes the number of days that individual i survived in demonstration year t subsequent to his/her eligibility date in that year; X_{it} includes the same set of beneficiary characteristics measured in the index year, as in Equation (1); PD_1 is an indicator for two years before the demonstration; $DY_1 - DY_5$ are a set of indicators for each post-demonstration year; and $treatment_{it}$ is an indicator variable for being in an IAH practice. The term z_{it} is an error with a distribution $f(\cdot)$.

The model in Equation (3) takes into account that the exact survival time was not observed for beneficiaries who did not die at the end of a given demonstration year (that is, right censoring)

and the survival time was not measured from the beginning of the demonstration year for beneficiaries who entered the study sample late (that is, left truncation). We estimated the model using the maximum likelihood method, with a generalized gamma distribution for $f(\cdot)$ to allow for the possibility of non-monotonic hazard functions.[17] We used matching weights to account for the number of matched comparisons per IAH beneficiary so that the two groups were the same size. We adjusted standard errors using the hybrid clustering approach, described in detail in Section VI.A.4 below.

After estimating the survival regression, we transformed $\theta_1 - \theta_5$ into difference-in-differences effect estimates, following steps similar to those we used for estimating impacts for other outcomes. Specifically, we obtained the regression-adjusted average mortality (that is, one minus the probability of survival by the end of the demonstration year) for IAH and comparison groups in each year. We adjusted the yearly average outcomes for both groups to reflect the covariate distribution of the IAH group in Year 5. Finally, we estimated the difference-in-differences impact by calculating the difference of the regression-adjusted mortality for the IAH group and matched comparison groups in that year relative to the difference between the two groups in the year before the demonstration. As with other outcomes, we estimated a separate difference-in-differences model that used a post-demonstration indicator and its interaction with the IAH status to obtain the five-year annual effect estimate.

As part of the outputs from the survival regression, we also obtained the predicted mortality for each beneficiary during a given demonstration year, based on the individual's treatment status and baseline characteristics. This predicted mortality, denoted as h_{it}, then fed into the estimation model for other binary outcomes, which we describe next.

3. Model specification for other binary outcomes

In addition to mortality, we estimated the impact of the demonstration on two other binary outcomes: the probability of having an unplanned readmission and the probability of entering institutional long-term care within the demonstration year. Our model specification for these outcomes was similar to that for continuous and count outcomes, but we used additional controls to account for differences between IAH and comparison group beneficiaries in the length of time they were exposed to the risk of the outcome.

We measured outcomes for the period that beneficiaries were eligible during a given demonstration year, which started from the date of eligibility through the end of the demonstration year or date of death. Therefore, the eligibility period differed across beneficiaries, depending on their eligibility start dates and death dates. In particular, death

[17] To inform our choice of the survival function, we compared the goodness-of-fit of models using different distributions. We considered five types of parametric survival distributions: (1) Weibull, (2) log logistic, (3) log normal, (4) generalized gamma, and (5) Gompertz. In choosing the final model, we analyzed the log likelihood, the Akaike information criterion, and the Bayesian information criterion across these different models.

(continued)

occurred less frequently in each demonstration year for the IAH group than for the matched comparison group; for example, 14.7 percent of IAH beneficiaries died during Year 5 compared to 18.3 percent of matched comparison beneficiaries.[18] Such a difference implied the importance of controlling for observation length because, all else being equal, IAH beneficiaries had more time during the demonstration year at risk for the outcome relative to the matched comparison beneficiaries. Further, mortality might directly affect the probability of readmission (or entry into long-term care) if the probability changes as individuals approach death. Thus, not controlling for death could bias the estimated effect of the demonstration.

For continuous and count outcomes, we accounted for differential observation lengths by annualizing the outcome and using eligibility weights in regressions (Section VI.A.5 describes weights in detail). However, because we could not annualize binary outcomes, we employed a modeling approach similar to the one used in Deb (2016). The basic idea behind Deb's model is to first estimate a survival model to derive the predicted probability of dying for each individual in each time period and include the predicted probability of dying in the second stage to account for the differences in outcomes due to differences in mortality rate across individuals.

Following Deb's approach, we estimated a survival-adjusted difference-in-differences model, controlling for the predicted probability of dying within the demonstration year (h_{it}), the interaction between treatment status and the probability of dying, and the proportion of time during the demonstration year that the beneficiary was eligible and alive ($survdays_{it}$).

Equation (4) shows our model specification:

$$(4)\quad P(Y_{it}=1) = \alpha + X_{it}\beta + \tau.treatment_{it} + \beta_h h_{it} + \beta_{Rh} treatment_i h_{it} + \beta_s survdays_{it} + \gamma_{-1}\, PD_1 + \gamma_1\, DY_1 + \gamma_2\, DY_2 + \gamma_3\, DY_3 + \gamma_4\, DY_4 + \gamma_5\, DY_5 + \theta_{-1} treatment_{it} \cdot PD_1 + \theta_1 treatment_{it} \cdot DY_1 + \theta_2 treatment_{it} \cdot DY_2 + \theta_3 treatment_{it} \cdot DY_3 + \theta_4 treatment_{it} \cdot DY_4 + \theta_5 treatment_{it} \cdot DY_5 + \omega_{it}$$

where Y_{it} is a binary variable for whether the beneficiary had an unplanned readmission (or long-term care entry); h_{it} is the predicted probability of dying in the demonstration year, derived from the estimated survival model in Equation (3) above; $survdays_{it}$ is the number of days from the beneficiary's eligibility date through the end of demonstration year or date of death, divided by 365 (or 366 for a leap year); and ω_{it} is a random error term. The remaining covariates are the same as those in Equation (1) above.

In Equation (4), the term h_{it} measures the predicted probability of beneficiaries' dying in the year, regardless of their actual survival or censoring status. Because h_{it} was derived from the difference-in-differences survival model, it accounted for any mortality difference between the IAH and comparison groups that was not captured in matching, as well as any mortality

[18] Ideally, the matching process would result in a comparison group with the same expected survival (as of the eligibility date) as the IAH beneficiaries. However, it is possible that factors not observable in claims data caused a differential expected survival between the two groups.

difference resulting from the demonstration. Coefficient β_h captured changes in the probability of readmission (long-term care entry) as the mortality rate increases, and coefficient β_{Rh} captured differential changes in this probability for those in the IAH group versus the comparison group. Last, coefficient β_s captured the effect of the length of time at risk of readmission (long-term care entry), conditional on predicted mortality.

We estimated Equation (4) using a logistic regression model.[19] As with mortality, we adjusted standard errors for hybrid clustering and used matching weights to ensure equal sizes of IAH and comparison groups. Because estimation of Equation (4) involves a generated regressor h_{it}, we bootstrapped our estimates and standard errors, employing a multiple-imputation approach (Deb 2016). After bootstrapping, we transformed $\theta_1 - \theta_5$ into difference-in-differences effect estimates, following steps similar to those we used in estimating impacts for other outcomes. For each outcome, we also estimated a separate difference-in-differences model that used a post-demonstration indicator and its interaction with the IAH status to obtain the five-year annual effect estimate.

4. Adjustment to standard errors for clustering

To obtain accurate estimates of standard errors for the impact estimates, it was important to account for possible clustering of observations within geographic areas. CMS selected certain practices to implement IAH, each of which serves beneficiaries in a specific area. We selected patients from the same geographic catchment area for the matched comparison group. The IAH group sample was clustered by practice in that geographic area—all beneficiaries who met the eligibility criteria and received home-based primary care from the same demonstration practice. However, we could not model practice-level clustering of the comparison group because we selected those beneficiaries without knowledge of the practice from which they received their primary care. We accounted for this asymmetric clustering structure of the two groups in our regression to avoid overstating the precision of the estimates.

In addition to the practice-level clustering, we had multiple observations for some beneficiaries in the sample. Because the observations on a given beneficiary in one period clearly were not independent of the observations on the same beneficiary in other periods, our estimator of the variance had to account for this time dependence of repeated observations.

[19] As noted previously, we calculated the long-term care measure through Year 4 because data were unavailable for Year 5.

To account for asymmetric practice-level clustering and multiple observations for some beneficiaries, we used what we refer to as a hybrid clustering approach. This approach accounted for clustering at the practice level for the IAH group only and took into account the time dependence of repeated observations for both IAH and comparison beneficiaries.[20] Implementing this approach meant that all IAH beneficiaries in a given site were from a single cluster. To correctly identify the clustering effect in the IAH group, we excluded the site fixed effects from the regression equation.[21]

Our approach to adjusting standard errors was consistent with the goal of evaluating only the practices that participated in the demonstration in this report. We could not generalize beyond the demonstration practices to home-based primary care provided across the nation as a whole because demonstration practices were not a random sample of all practices, and we did not know the extent to which IAH sites were similar to other practices and the types of patients they serve. Instead, we assumed that the IAH beneficiaries in a given practice were a random sample of all eligible beneficiaries of that practice; thus, our statistical tests accounted for the random variation among eligible beneficiaries who received care from the demonstration sites.

5. Weighting

For continuous and count outcomes, we estimated regressions with observations at the beneficiary level and weighted the observations to capture two factors: (1) the share of months a given beneficiary was eligible for the demonstration during each pre-demonstration or demonstration year and (2) the number of comparison beneficiaries matched to each treatment beneficiary. We referred to the former as the eligibility weight; it controlled for differences in the length of time that beneficiaries were observed during a given study year. We referred to the latter as the matching weight. Because we matched each treatment beneficiary to up to five comparison beneficiaries, applying matching weights ensured that the impact regression was not disproportionally weighted toward the comparison beneficiaries.

[20] Accounting for clustering at the practice level for the treatment group captures the correlation among observations in each IAH practice, whether for the same individual across time periods or different individuals in the same time period. We implemented the hybrid clustering approach in the statistical software used for the analysis (Stata) by defining a cluster variable that takes the value of the practice ID for the treatment group and the value of the beneficiary ID for the comparison group.

[21] Ideally, including site fixed effects would improve estimation by controlling for factors that varied across geographic locations and affected outcomes for IAH and comparison beneficiaries within a given area. However, because all IAH beneficiaries in a given site (stratum) were from a single practice (cluster), controlling for both stratification and clustering at the same level would lead to under-identification. That is, we could not identify the clustering effect with only one IAH group practice per site in a stratified design (Schochet 2008). Relative to the site fixed effects, clustering was by far the more important factor to account for when estimating the variance of the estimate. If we failed to account for clustering when estimating variance, the standard errors and statistical significance of the estimates would be misleading and could lead to incorrect conclusions about the impact of the demonstration. To avoid that problem, we could not take advantage of the gains we would have achieved by accounting for the stratified approach.

The construction of final beneficiary weights for continuous and count outcomes required three steps. First, we constructed the eligibility weight as the share of months eligible for the demonstration during each pre-demonstration or demonstration year. After we determined a beneficiary's eligibility for the demonstration in a given pre-demonstration or demonstration year, we included the beneficiary in the analysis sample beginning on the first day of the following month. That beneficiary remained in our analysis sample for the entire year unless he or she left Medicare FFS or died. For example, if a beneficiary entered the Year 5 sample on January 1, 2017 and died on June 20, 2017, that person was eligible for the demonstration for six months and thus had an eligibility weight of 0.5.

Second, we constructed matching weights to account for the size of the matched set. Each IAH beneficiary received a weight of 1, and each matched comparison beneficiary received a weight that was the inverse of the number of comparison beneficiaries within the matched set. For example, if an IAH beneficiary was matched to four comparison beneficiaries, each of the latter received a weight of 0.25. Comparison beneficiaries' matching weights ranged from 0.2 (if there were five matched comparisons for a particular IAH beneficiary) to 1 (one matched comparison). For all outcomes other than mortality, we obtained a composite weight by multiplying the eligibility weight by the matching weight.

In the third step, we created the final analytic weight for each beneficiary by rescaling the composite weight to ensure equality in the weighted number of IAH and comparison beneficiaries for each site and year. As described earlier, we implemented hybrid clustering adjustments but could not use site fixed effects (an indicator for each site). Because beneficiaries had different eligibility weights, the number of weighted IAH beneficiaries in a given site and year might differ from the number of weighted comparison beneficiaries in the same site and year if we used the composite weight without rescaling it. For this reason, we rescaled the weights for comparison beneficiaries by site and year so that for each year, the weighted number of IAH beneficiaries equaled the weighted number of comparison group beneficiaries for each site. This approach ensured that the estimated treatment-comparison differences and the difference-in-differences estimates for each year accounted for any differential weighting of the IAH and comparison groups. For more information on how we rescaled the composite weights, please see Kimmey et al. (2019).

For binary outcomes, we used matching weights only. We did not include an eligibility weight in the mortality regression because the survival model we employed takes into account differential observation lengths for the outcome via a hazard function.[22] For other binary outcomes (probability of unplanned readmission and entering institutional long-term care), we could not use the same eligibility weight used for other outcomes because the length of time observed did not matter in cases where the beneficiary experienced a particular outcome. Instead of relying on an eligibility weight, we used a survival-adjusted model for binary outcomes, which explicitly

[22] The weights used for the survival regression did not have to be rescaled because, without any eligibility weights, the matching weights ensured that the weighted number of IAH and comparison beneficiaries for each site and year were equal to each other.

takes into account the effects of mortality and time survived since eligibility. We describe the model specifications for these binary outcomes in subsections VI.A.2 and VI.A.3.

We refer to the above weighting scheme as beneficiary weighting. Under beneficiary weighting, large practices that served more beneficiaries had more influence on the estimated effect and smaller practices had less influence. In Appendix C, we report estimation results based on an alternative weighting scheme that allows all practices to have equal influence on the estimated effect, regardless of the size of their patient population.

6. Control variables

Although our matching process ensured that the comparison groups were very similar to the IAH groups along many characteristics, there might still be important differences in some of these characteristics that could affect the outcomes. Therefore, we included four types of control variables: (1) variables describing eligibility for the demonstration; (2) demographic characteristics; (3) ADL indicators; and (4) measures of health status, including HCC risk score, HCC indicators, and chronic condition indicators (Exhibit A.10). We included all specific HCC indicators and categories of HCCs used for matching (Exhibit A.5). Some of these control variables were at a more detailed level than the variables we used in matching; for example, we used three age categories in propensity-score matching, whereas we used five age categories in the outcome regressions.

As noted earlier, we included a dummy variable for each year and an indicator of whether the beneficiary was in the IAH or comparison group. Given the repeated cross-sections in our multiyear data set, we used contemporaneous control variables for all years of the demonstration; for example, in demonstration Year 5, we used the Year 5 values of all control variables, whether or not a beneficiary appeared in the sample in an earlier demonstration year.

Exhibit A.10. Control variables used in regressions

Variable
Eligibility for the demonstration
Number of months since most recent inpatient admission: 1, 2–3, 4 or more
Month of the demonstration year that beneficiary met eligibility criteria (1, 2–6, 7–12)[a]
Demographic characteristics
Age: younger than 65, 65–74, 75–79, 80–84, 85 or older
Gender
Race and ethnicity: white, black, Hispanic, Asian, American Indian/Alaska Native, other, or unknown
Dually eligible for Medicare and Medicaid
Original reason for Medicare entitlement: old age, ESRD or ESRD and disability, disability only
ADLs
Number of ADLs for which beneficiary requires human assistance: 2, 3 or 4, 5 or 6

Exhibit A.10 *(continued)*

Variable
Whether information about the feeding ADL was missing[b]
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
Chronic conditions measured by Chronic Conditions Data Warehouse
Alzheimer's or dementia
Acute myocardial infarction or ischemic heart disease
Asthma
Hip or pelvic fracture
Stroke or transient ischemic attack
Number of chronic conditions and the square of the number of conditions
Other measures of health status
Anemia[c]

Exhibit A.10 *(continued)*

Variable
Fluid and electrolyte disorders[c]
Chronically critically ill or medically complex diagnosis

Notes: This table lists HCCs used in all regressions.

[a]For all sites in Years 4 and 5, Month 1 is October. In Years 1–3, sites began the demonstration in June or September each year. For sites that began in June, Month 1 is June. For sites that began in September, Month 1 is September.

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

[c]Measured using claims from the most recent inpatient stay and observation stay in the year before the demonstration eligibility date. We drew diagnosis codes for these conditions from Gagne et al. (2011).

ADLs = activities of daily living; ESRD = end-stage renal disease; HCC = hierarchical condition category.

7. Testing for the validity of the difference-in-differences estimates

The validity of the difference-in-differences estimates for the demonstration years relied on the classic difference-in-differences assumption that there was no significant differential trend between the IAH and matched comparison groups during the pre-demonstration period. Therefore, the difference-in-differences estimate for two years before the demonstration, θ_{-1}, served two purposes: (1) it ruled out or identified significant treatment-comparison differences in trends during the pre-demonstration period; and (2) in so doing, it helped inform the more important difference-in-differences analysis for the demonstration period. Specifically, a statistically significant θ_{-1} indicated that the difference in a given outcome between the IAH and comparison groups changed significantly from two years before the demonstration to the year before the demonstration. This meant that the IAH and comparison groups could have been on nonparallel outcome trajectories during the pre-demonstration period. We referred to nonparallel outcome trajectories during the pre-demonstration period as a pre-existing difference in trend.

The possible presence of nonparallel pre-demonstration trends would have limited our confidence in the demonstration impact estimates for a given outcome. This was because the difference-in-differences estimates for the demonstration years could have reflected the continuation of a pattern—for example, narrowing or widening differences between the two groups—that began during the pre-demonstration period, rather than reflecting an impact of the demonstration payment incentive.

We examined the difference-in-differences estimate for two years before the demonstration for all outcomes reported. The estimate was not statistically significant for most outcomes, including expenditures and hospital care use, suggesting that the parallel-trend assumption held for those outcomes. We found statistically significant difference-in-differences estimates for two years before the demonstration for potentially avoidable hospital admissions and mortality, which violated the parallel-trend assumption. Because it was impossible to rule out the possibility of

truly nonparallel pre-existing trends for outcomes where the difference-in-differences estimate for two years before the demonstration was significant, we were cautious in interpreting the impact estimates for potentially avoidable hospital admissions and mortality.[23]

8. Assessing the relative influence of individual practices

As noted above, under beneficiary weighting, practices have different amounts of influence on the estimated effect depending on their sizes. To understand which practices drove the estimated effects of the demonstration, we re-estimated the beneficiary-weight regression, leaving out one practice at a time.[24] Specifically, we estimated 16 regressions (treating each member of the Richmond-based consortium separately), with each regression excluding the IAH beneficiaries from one practice and their matched comparisons in all years. If all 16 regressions showed similar estimates as the main regression, we would conclude that all practices equally influenced the full sample estimate. On the other hand, if excluding a given practice substantially changed the estimated effect, we would conclude that the site strongly influenced the full sample estimate. We performed this analysis on all outcomes reported in the Chapter II. In Appendix B, we report the year 5 estimated effects from each of the 16 regressions for selected outcomes.

B. Bayesian difference-in-differences models

1. Overview

In addition to the frequentist (traditional) analyses we describe in Section VI.A, we conducted a set of analyses using the Bayesian statistical paradigm. Assessing the effects of IAH probabilistically, as Bayesian techniques permit, maintains a rigorous statistical standard while providing a more flexible interpretation of the program's effects. The frequentist approach classifies the demonstration's impact as statistically significant or not statistically significant; in contrast, a Bayesian analysis allows probabilistic estimates about whether the demonstration achieved a certain outcome. For example, one could conclude that "there was an 84 percent chance that the IAH demonstration incentive produced savings of at least $50 PBPM in demonstration Year 5." Such conclusions offer the opportunity to tailor inferences to substantive questions of interest and apply subject matter expertise in deeming meaningful effects.

Overall, the Bayesian and frequentist analyses were similar, but they had some differences. As with the frequentist approach, the Bayesian analysis used a comparison group difference-in-differences design to identify effects attributable to the IAH demonstration. The outcome of interest was total Medicare expenditures PBPM. We used the same data sets for the frequentist and Bayesian analyses. Moreover, we used the same eligibility and matching weights, and the

[23] It is possible to control for pre-existing trends by including linear time trends in the regression. However, this approach would impose an overly restrictive assumption on our model—that the one-year pre-demonstration trends would continue throughout the demonstration.

[24] We could have estimated regressions separately for each site to obtain site-specific estimated effects. However, the statistical power for these regressions was too low due to the small sample sizes at the site level. Estimating the regressions while excluding one site at a time enabled us to assess the influence of each site by comparing those estimates to the estimate from the full sample.

same control variables. However, the Bayesian analysis diverged from the frequentist analysis in three ways, as described here. In this section, we describe the three factors that differentiated the Bayesian analyses from their frequentist counterparts: the prior distributions, the method used to account for clustering, and the computational approach used to fit the models.

a. Prior distribution

Assigning a prior distribution to each model parameter translated the model into the Bayesian framework and thus allowed for probabilistic inference. We placed a standard normal prior distribution—denoted $N(0,1)$—on the overall impact of IAH. By doing so, we incorporated a prior expectation that very large positive or negative impacts of IAH on expenditures were substantially less likely than small and moderate impacts. We based our prior expectation on the general result that other interventions of the impact of home-based primary care and other interventions for chronically ill, frail beneficiaries very rarely show effect sizes larger than two standard deviations. We centered the normal distribution at a mean of zero to remain agnostic about whether the IAH demonstration would be successful.

b. Method used to account for clustering

The full Bayesian model accounted for clustering by using random effects, whereas the frequentist analysis used cluster-robust standard errors (as described earlier in this appendix). Specifically, the two-stage full Bayesian model accounted for clustering using beneficiary- and site-specific random effects for both the IAH and comparison groups, where each site included IAH beneficiaries from a demonstration practice and their matched comparison beneficiaries. In contrast, the frequentist analysis estimated cluster-robust standard errors, which assumed that IAH beneficiaries were clustered by practices and comparison beneficiaries were clustered by individual beneficiaries rather than practices (a hybrid clustering approach). The Bayesian model could not apply the same approach because it accounted for clustering using random effects instead of cluster-robust standard errors.[25] This methodological difference in accounting for clustering could lead to differences in both point estimate and standard error of the estimate.

c. Two-stage model

We further modified the frequentist model to make Bayesian computationally feasible. We adopted these modifications purely as a computational convenience; they are not inherently Bayesian, and a traditional impact estimation framework could also adopt this approach. Ideally, we would have liked to fit a single, unified model at the beneficiary level, as in the frequentist analysis (see Equation [5] below), but such a model would have taken more than a month to converge on our analysis platform. Because of time constraints, we used a two-stage approximation of this ideal beneficiary-level model. In the first stage, we aggregated the

[25] A Bayesian model requires a fully model-based approach to account for clustering, whereas cluster-robust standard errors are an adjustment performed after the modeling process.

beneficiary-level data set to the site level. Using output from Stage 1, we estimated the impact of the IAH demonstration using a Bayesian difference-in-differences framework in Stage 2.

2. Full Bayesian model, pooled

To understand the full Bayesian model, we begin by presenting a single unified model at the beneficiary level. As we show in Equation (5), this procedure accomplishes impact estimation and risk adjustment simultaneously through a model of the following form:

$$(5)\ Y_{ijt} = \alpha + X_{it}\beta + \tau z_{it} + \gamma_t + \theta_t z_{it} + a_i + b_j + c_j z_{it} + d_{jzt} + \varepsilon_{it}$$

This model uses a slightly different notation than its frequentist counterpart, Equation (1), for clarity of presentation of the random effects.

- We use i to index beneficiaries; $j = 1, \ldots, 16$ to index geographic areas (or, loosely speaking, sites at which both IAH and comparison beneficiaries reside); and $t = -1, \ldots, 5$ to index years.
- Y_{ijt} is total Medicare expenditures PBPM measured for beneficiary i from site j in year t; X_{it} is a set of beneficiary characteristics measured in year t; z_{it} is the treatment status of beneficiary i in year t.
- Greek letters denote parameters to be estimated: α is a constant term; β contains the effects of the beneficiary characteristics; τ captures any differences between IAH and comparison beneficiaries in the year before the demonstration that persist despite matching; γ describes the secular time trend that applies to both IAH and comparison beneficiaries; and the θs are the difference-in-differences impacts of interest. As with the frequentist model, we estimated γ_{-1} and θ_{-1} for two years before the demonstration, and $\gamma_1 - \gamma_5$ and $\theta_1 - \theta_5$ for each of the five demonstration years. Note that $t = 0$ corresponds to the baseline year (the year before the demonstration), so γ_0 and θ_0 are both omitted from the model.
- Random effects are denoted by Roman letters: the a's and b's are beneficiary- and site-level random intercepts, respectively, which account for the correlation across repeated observations on a given beneficiary or site; the c's are site-specific baseline IAH/comparison differences; and the d's are site-treat-year random intercepts. We assume that the a's and d's each follow a univariate normal distribution, whereas the b's and c's jointly follow a bivariate normal distribution. The latter assumption allowed for correlation between a site's intercept and the IAH/comparison difference in that site.

Last, we weighted the regression using the same weighting schemes (beneficiary weighting and practice weighting) that we used in the frequentist analysis, as discussed in Section VI.A.5.

We estimated the adjusted total Medicare expenditures for the IAH and matched comparison groups in each year, the difference-in-differences estimates (θ_{-1}, θ_1-θ_5), and percentage impact relative to unadjusted IAH group mean expenditures in the year before the demonstration. In

addition, we estimated the probability of reducing expenditures by at least $50 or $100 PBPM. In all calculations, we adjusted the yearly average outcomes for both groups to reflect the covariate distribution of the IAH group in the latest (fifth) demonstration year—the same approach we used in the frequentist analysis.

Due to the number of observations in the data set, fitting Equation (5) as a single, unified model at the beneficiary level was computationally prohibitive. For this reason, we fitted the full Bayesian model using a two-stage approximation to decrease computational run times. The first-stage model was a beneficiary-year-level risk adjustment fit using hierarchical linear regression. The goals of the first-stage analysis were to aggregate beneficiaries to the site level and risk-adjust outcomes to enable comparisons across sites and years whose case mix differed (Equation [6]). In the first-stage model, we adjusted for the same beneficiary-level covariates as the frequentist model (see Exhibit A.10). The risk-adjusted site-year-level output from Stage 1 was used as data in Stage 2, which estimated the impact of IAH demonstration in a Bayesian difference-in-differences framework (Equation [7]).

(6) Stage 1: $Y_{ijt} = A_{jz_{it}t} + X_{it}\beta + a_i + \varepsilon_{it}$

As described above, the site-treatment-year effect A_{jzt} represents the estimated fixed effect for site j and treatment group z in year t. There were 240 such fixed effects from two groups (IAH and comparison) from each of the 16 sites in each year. The parameters β describe the effects of beneficiary-level control variables X_{it}, whereas beneficiary-level random effects a_i account for correlations across repeated observations on beneficiary i. We assumed that the beneficiary-level random effects a_i and the overall error term ε_{it} came from a normal distribution with mean zero and its own variance. Similar to the frequentist model, we used the rescaled composite weights for the Stage 1 model. Then, we used the aggregated site-treatment-year estimates ($\hat{A}_{jzt}$) and associated standard errors (s_{jzt}) from the Stage 1 model when we estimated the Stage 2 full Bayesian difference-in-differences regression (Equation [7]).

(7) Stage 2: $\hat{A}_{jzt} = \alpha + \tau z + \gamma_t + \theta_t z + b_j + c_j z + d_{jzt} + \varepsilon_{jzt}$

In the Stage 2 model, we included an overall intercept α and controls for the secular time trend γ_t and treatment τ. We accounted for clustering through random effects b_j, c_j, and d_j, as described earlier. The parameters of interest, θ_t, represent the overall difference-in-differences terms. To estimate the overall impact estimate of all five post-demonstration years, we re-estimated the Stage 2 model with one post-demonstration dummy instead of separate dummies for each demonstration year.

We assigned a standard normal distribution—Normal(0, 1)—as the prior for each model parameter: $\alpha \sim N(0,1), \tau \sim N(0,1), \gamma \sim N(0,1), \theta \sim N(0,1), \left(b_j, c_j\right) \sim MVN\left(0, \Sigma\right), d \sim N(0, \sigma^2)$ where σ^2 is the overall noise variance. The prior for Σ included two parts: one part to address correlations between b_j and c_j, and one to address the standard deviation of b_j and c_j.

The former part took on an LKJ correlation prior (Lewandowski et al. 2009); the latter took on a standard normal distribution. The multiplication of these two parts constituted the prior on $\Sigma : \Sigma = \begin{pmatrix} \sigma_c & 0 \\ 0 & \sigma_d \end{pmatrix} \Omega \begin{pmatrix} \sigma_c & 0 \\ 0 & \sigma_d \end{pmatrix}$ where $\sigma_c, \sigma_d \sim N(0,1)$ and $\Omega \sim \text{LKJ}(2)$.[26] Last, our prior on the error term is given by $\varepsilon_{jzt} \sim Normal\left(0, s_{jzt}^2\right)$. Therefore, both σ^2 and s_{jzt}^2 act as weights in Stage 2. We used the "lme4" package in R to fit the Stage 1 model. For Stage 2, we used a novel probabilistic programming language called Stan, which provides fast, full Bayesian inference, even for complex models.

[26] LKJ is a distribution on correlation matrices (usually called Ω). The distribution has one parameter, ν, so $\Omega \sim \text{LKJ}(\nu)$. When $\nu = 1$, the distribution is uniform over all possible correlation matrices. As ν increases, the distribution is more concentrated on the identity matrix, which corresponds to zero correlations. Thus, for $\nu = 2$, the distribution slightly favors less correlation, shrinking the correlations somewhat toward zero. This is a weakly informative prior to help stabilize the estimation.

VII. QUALITATIVE METHODS AND DATA

To understand why and how the incentive payments might (or might not) have affected outcomes, we needed to understand how IAH practices' provision of home-based primary care changed after the IAH demonstration began and throughout the demonstration. Identifying the potential effect of IAH practices' changes also required understanding how the IAH participating practices provided home-based primary care before the IAH demonstration. Understanding the care delivery model enabled us to assess whether changes made by the participating practices appeared to be designed to reduce Medicare expenditures without harming patients. When reporting information about the IAH practices in Chapter I and interpreting the results in Chapters II and III, we relied on qualitative data gathered from demonstration sites during demonstration Years 1 through 3 and Year 5.

- We conducted the most recent interviews in April 2017. During these interviews, we interviewed 25 clinical and administrative staff at 15 IAH practices and the Visiting Physicians Association (VPA) corporate office in Troy, Michigan. We asked respondents about changes their practices had made during the demonstration to reduce hospital admissions and readmissions, reduce avoidable ED use, coordinate care, ensure round-the-clock access to care, follow up with patients and reconcile medications within 48 hours after discharge from the hospital or ED, and document patients' preferences. We also asked about motivation for making changes, clinician and staff reactions to changes, and factors that affected implementation of those changes.
- During telephone interviews conducted in January and February 2017, we collected information about IAH practices' structural characteristics and how they deliver care.
- During visits to demonstration sites from April 2015 to October 2015, we interviewed the sites' IAH team members and administrative staff involved in implementing the IAH demonstration. During this round of site visits, we focused on documenting changes in how the practices delivered care, the barriers to and facilitators of meeting the requirements of the demonstration, and how sites planned to sustain the home-based primary care model.
- Finally, we provide information gathered during earlier rounds of site visits: February to May 2013 (visits during Year 1) and February to July 2014 (visits during Year 2). During these earlier site visits, we focused on documenting how the practices delivered care, including changes from the year before the demonstration to Year 1 and changes from Year 1 to Year 2. During this period, we also collected information on barriers to and facilitators of meeting the requirements of the demonstration, and how sites used information technologies such as electronic health records and health information exchange to support their work.

For all interviews, we coded the data using a template that reflected the various requirements of the IAH demonstration (for example, providing patients with 24-hour access to the care team, working to reduce ED visits). The coding template also captured aspects of the five domains identified by the Consolidated Framework for Implementation Research (Damschroder et al.

2009) as playing an important role in implementation success: (1) the inner setting (internal attributes) of the practice sites, including structural and cultural characteristics affecting capacity for change; (2) the external environment (such as the availability of clinicians in the IAH practice's local market); (3) characteristics of the IAH demonstration itself; (4) characteristics of the individuals involved in implementing the model; and (5) processes used to implement the model. We used ATLAS.ti software to sort data using this coding template. We analyzed the sorted data to identify key barriers to and facilitators of implementation of the IAH demonstration in each participating site and identified common themes across sites.

Our analysis of qualitative data entailed a description of what happened during the demonstration. We did not have a comparison group of primary care practices, so we could not be certain whether changes in practices' operations or structure occurred because of the demonstration. In addition, because we did not conduct site visits until after the demonstration began, data on practices' operations and structure before the demonstration was limited to what interviewees told us was different in Year 1 relative to before the demonstration.

VIII. LIMITATIONS

As with all analyses, our study of changes in how IAH practices provided care, patient and caregiver survey data, and the effect of the demonstration on outcomes such as Medicare expenditures and utilization has some limitations.

Generalizability of the practices. This examination was not designed to draw conclusions about how the IAH demonstration might affect outcomes for Medicare FFS beneficiaries who receive home-based primary care from practices other than those in the demonstration. Among the pool of home-based primary care practices that volunteered for the demonstration, CMS selected 18 sites to represent different types of practices and geographic areas. The IAH practices were not selected to represent the national population of practices providing home-based primary care to Medicare beneficiaries with multiple chronic conditions and substantial functional limitations. Thus, we could not generalize the results of this study to Medicare FFS beneficiaries who received home-based primary care from practices other than those in the demonstration.

In addition to the small number of demonstration sites, only a portion of any IAH site's patients qualified for the demonstration and, in many cases, the number who qualified was very small. These small numbers of observations made it difficult to compare demonstration sites and obtain robust information about what works for an individual site or across groups of sites. We did not have the ability (that is, statistical power) to identify small effects of the demonstration payment incentive across all demonstration sites.

Generalizability of the patient population to the target population. Congress identified a target population for the demonstration by establishing eligibility criteria in the IAH legislation. However, the criteria could be interpreted in different ways, and the IAH practices varied in how they interpreted and implemented them. Also, the approaches of Mathematica and the implementation contractor to identifying eligible beneficiaries yielded different counts of IAH practices' beneficiaries.

For the sake of scientific validity, our sample differed from the population of IAH enrollees in two ways. First, we excluded beneficiaries enrolled but not confirmed as eligible in the administrative data we used for the evaluation. If a beneficiary was eligible for and enrolled in the demonstration in one year and continued to be enrolled the next year, that beneficiary was in our sample in the next year only if he or she met all of the demonstration eligibility criteria again—thus, we excluded beneficiaries who avoided recent hospital stays or the use of rehabilitation services (two of the demonstration eligibility criteria). The value of the demonstration for beneficiaries with chronic conditions who avoided recent hospital stays or the use of rehabilitation services is not known and might differ from what we measured in the study. Second, we included in the evaluation those beneficiaries who received care from demonstration clinicians and were eligible for the demonstration based on administrative data but not enrolled in it. The fact that sites did not enroll all of the eligible beneficiaries we identified in the administrative data underscores the difficulties the demonstration faced in applying the eligibility criteria consistently.

Limitations regarding changes over time. There may have been unmeasured differences in how IAH and comparison beneficiaries changed over time. For example, in setting the beneficiary requirements for the IAH demonstration, Congress used four key health status and health care utilization factors to define eligibility: (1) two or more chronic conditions, (2) needing human assistance with two or more ADLs, (3) recent hospitalization, and (4) recent use of rehabilitation services. The last three of these measures can identify a patient who was temporarily acutely ill and disabled or one who was chronically ill and permanently disabled. Because we could not distinguish between these two underlying situations, we did not know whether the proportion of these types of beneficiaries had changed differently over time, which could have affected the measurement of the demonstration effect. More broadly, if the patient mix in the IAH and comparison groups changed over time in ways we could not observe, and the change was not due to the demonstration payment incentive, the results could be inaccurate.

Another factor that may have caused unmeasured changes in the IAH and comparison groups over time was the participation of several IAH practices in ACOs in Years 4 and 5. If ACO and non-ACO IAH beneficiaries had differences in health status that affected Medicare expenditures but that we could not measure in administrative data, and if the comparison group did not experience a similar change in health status, then participation in ACOs would cause bias in our effect results in Years 4 and 5 and the average annual effect of the demonstration. However, we have no strong evidence about whether such participation may have led to higher or lower expenditure reductions in Years 4 and 5 than would have occurred without participation in ACOs.

Possibility of differential outcome trends from the baseline. The validity of our estimated effects assumes that the outcomes of IAH and matched comparison groups followed the same trend before the demonstration. That is, we assumed that outcomes changed at the same rate for both groups in the two-year pre-demonstration period, so any difference in outcomes between the two groups would remain the same during that period. We examined this assumption by testing whether the outcomes changed differentially in the pre-demonstration period. Most outcomes did not have pre-demonstration differences in trends, but mortality and number of potentially avoidable hospitalizations did. In other words, for those two outcomes, the change for the IAH beneficiaries from two years before the demonstration to one year before it was statistically significantly different from the change for comparison beneficiaries over those two pre-demonstration years. It is possible that the differential trends might have contributed to the post-demonstration differences in these outcomes, thus masking the true effect of the demonstration (if any).

This page has been left blank for double-sided copying.

Appendix B

Supplementary tables

This page has been left blank for double-sided copying.

The tables in this section present results for the analyses we describe in Chapter II.

Exhibit B.1. Baseline unadjusted means of outcomes among all IAH beneficiaries

Outcome name	Baseline unadjusted mean
PBPM Medicare expenditures	
Total Medicare expenditures	$4,397
Inpatient hospital services	$1,741
Skilled nursing facilities	$605
Home health services (Parts A and B)	$781
Hospice services	$153
Outpatient services	$253
Physician/supplier	$715
Durable medical equipment	$150
Service utilization outcomes	
Number of hospital admissions per beneficiary per year[a]	1.78
Number of potentially avoidable hospital admissions per beneficiary per year[b]	0.46
Probability of having a qualifying hospital discharge and an unplanned readmission within 30 days of discharge (percentage)[c]	16.04
Number of ED visits per beneficiary per year	2.90
Number of outpatient ED visits per beneficiary per year[d]	1.46
Number of ED visits resulting in inpatient admission per beneficiary per year	1.44
Number of potentially avoidable outpatient ED visits per beneficiary per year[e]	0.19
Unintended consequences	
Death in the demonstration year (percentage)	14.85
Entry into institutional long-term care in the demonstration year (percentage)	7.00

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: We calculated the baseline means of outcomes using the sample of IAH-eligible beneficiaries in the year in the year before IAH started.

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

Exhibit B.1 (*continued*)

[b]The number of potentially avoidable hospital admissions includes observation stays. A potentially avoidable hospital admission is one in which appropriate primary and specialty care might prevent or reduce the need for a hospital admission.

[c]The probability of an unplanned readmission for a beneficiary is measured over the IAH-eligible months during each demonstration year. The probability equals zero for beneficiaries who did not have a qualifying hospital discharge or unplanned readmission within 30 days of a qualifying hospital discharge during the measurement period.

[d]The number of outpatient ED visits measures all those not resulting in a hospital admission, including those resulting in an observation stay.

[e]A potentially avoidable outpatient ED visit is one in which appropriate primary and specialty care might prevent or reduce the need for such visits. The measure excluded ED visits that led to an inpatient admission because there was no diagnosis from such a visit in a claim record when the ED visit led to an inpatient admission.

ED = emergency department; IAH = Independence at Home; PBPM = per beneficiary per month

Exhibit B.2. Estimated effect of the IAH payment incentive on total Medicare expenditures PBPM for IAH beneficiaries

	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Five-year average annual effect[b]	$4,362	$4,664	-$302 ($139)	-$200 ($151)	-4.6%
Year 5	$4,526	$4,958	-$432 ($158)	-$330* ($182)	-7.5%
Year 4	$4,301	$4,685	-$384 ($177)	-$282 ($205)	-6.4%
Year 3	$4,481	$4,762	-$280 ($129)	-$178 ($158)	-4.1%
Year 2	$4,639	$4,771	-$133 ($124)	-$31 ($139)	-0.7%
Year 1	$4,658	$4,879	-$221 ($146)	-$119 ($97)	-2.7%
One year pre-IAH[c]	$4,794	$4,896	-$102 ($186)	-	-
Two years pre-IAH	$4,972	$5,107	-$135 ($190)	-$33 ($57)	-0.8%

Total unweighted number of observations across all years: 295,292

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

[b]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[c]The difference-in-differences estimate for the period before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means

Exhibit B.2 (*continued*)

between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

IAH = Independence at Home; PBPM = per beneficiary per month.

Exhibit B.3. Estimated effect of the IAH payment incentive on total Medicare expenditures PBPM for IAH beneficiaries under Bayesian model

	Difference-in-differences estimated effect	Percentage effect[a]	Probability of savings ≥ $0	Probability of savings ≥ $100
Five-year average annual effect[b]	-$68 (-$220, $79)	-1.5%	77.1%	35.9%
Year 5	-$193 (-$413, $26)	-4.4%	95.6%	79.4%
Year 4	-$127 (-$350, $97)	-2.9%	86.8%	59.0%
Year 3	-$51 (-$277, $175)	-1.2%	67.4%	33.4%
Year 2	$128 (-$93, $353)	2.9%	13.0%	2.1%
Year 1	-$65 (-$290, $158)	-1.5%	71.7%	37.5%
One year pre-IAH[c]	-	-	-	-
Two years pre-IAH	-$39 (-$270, $194)	-0.9%	63.4%	30.1%

Total unweighted number of observations across all years: 295,292

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Parentheses report 90 percent credible intervals. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

[b]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[c]The difference-in-differences estimate for the period before the demonstration is zero (with no credible interval) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

IAH = Independence at Home; PBPM = per beneficiary per month.

Exhibit B.4. Estimated effect of the IAH payment incentive on Medicare expenditures PBPM for IAH beneficiaries, by service category

Service type	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Inpatient hospital services					
Five-year average annual effect[b]	$1,768	$2,076	-$308	-$118 ($95)	-6.8
Year 5	$1,834	$2,245	-$410	-$221* ($120)	-12.7
Year 4	$1,720	$2,065	-$346	-$156 ($134)	-9.0
Year 3	$1,838	$2,106	-$269	-$79 ($95)	-4.5
Year 2	$1,900	$2,130	-$231	-$41 ($77)	-2.4
Year 1	$1,923	$2,168	-$245	-$55 ($61)	-3.2
One year pre-IAH[c]	$1,943	$2,133	-$190	-	-
Two years pre-IAH	$2,032	$2,233	-$201	-$12 ($49)	-0.7
SNF					
Five-year average annual effect[b]	$653	$863	-$210	-$5 ($31)	-0.8
Year 5	$693	$900	-$207	-$2 ($46)	-0.3
Year 4	$676	$900	-$224	-$18 ($50)	-3.0
Year 3	$705	$911	-$205	$0 ($27)	0.0
Year 2	$678	$871	-$192	$13 ($25)	2.2
Year 1	$673	$892	-$219	-$14 ($19)	-2.2
One year pre-IAH[c]	$687	$893	-$206	-	-
Two years pre-IAH	$734	$949	-$215	-$9 ($18)	-1.5
Home health services (Parts A and B)					
Five-year average annual effect[b]	$697	$498	$200	-$5 ($28)	-0.7
Year 5	$703	$494	$208	$4 ($36)	0.5
Year 4	$660	$473	$187	-$17 ($36)	-2.2
Year 3	$659	$487	$172	-$33 ($38)	-4.2
Year 2	$788	$554	$234	$30 ($24)	3.8
Year 1	$748	$552	$196	-$8 ($17)	-1.1
One year pre-IAH[c]	$791	$586	$204	-	-
Two years pre-IAH	$849	$637	$212	$8 ($10)	1.0

Service type	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Hospice services					
Five-year average annual effect[b]	$160	$102	$58	$2 ($8)	1.3
Year 5	$167	$112	$55	-$1 ($11)	-0.8
Year 4	$169	$113	$55	-$1 ($10)	-0.8
Year 3	$162	$101	$61	$5 ($12)	3.0
Year 2	$151	$84	$67	$10 ($11)	6.7
Year 1	$162	$106	$56	$0 ($8)	-0.2
One year pre-IAH[c]	$164	$108	$56	-	-
Two years pre-IAH	$144	$100	$44	-$12 ($8)	-8.0
Outpatient services					
Five-year average annual effect[b]	$286	$360	-$73	-$12 ($10)	-4.9
Year 5	$319	$395	-$76	-$15 ($15)	-6.1
Year 4	$296	$375	-$79	-$19 ($14)	-7.3
Year 3	$294	$365	-$70	-$10 ($11)	-3.8
Year 2	$285	$346	-$61	-$1 ($9)	-0.3
Year 1	$263	$337	-$74	-$14* ($8)	-5.4
One year pre-IAH[c]	$273	$334	-$61	-	-
Two years pre-IAH	$268	$231	-$53	$8 ($8)	3.1
Physician/supplier services					
Five-year average annual effect[b]	$699	$677	$22	-$33 ($32)	-4.7
Year 5	$721	$724	-$2	-$57 ($35)	-8.0
Year 4	$693	$675	$17	-$38 ($39)	-5.3
Year 3	$723	$695	$28	-$27 ($32)	-3.8
Year 2	$727	$692	$35	-$20 ($32)	-2.8
Year 1	$744	$706	$38	-$18 ($24)	-2.5
One year pre-IAH[c]	$772	$717	$55	-	-
Two years pre-IAH	$775	$732	$42	-$13 ($10)	-1.8

Service type	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Durable medical equipment					
Five-year average annual effect[b]	$99	$89	$10	-$28*** ($7)	-18.8
Year 5	$89	$89	$0	-$38*** ($9)	-25.2
Year 4	$87	$82	$5	-$33*** ($11)	-22.2
Year 3	$100	$96	$4	-$35*** ($9)	-23.1
Year 2	$110	$94	$16	-$22*** ($8)	-14.9
Year 1	$145	$117	$28	-$10** ($5)	-6.7
One year pre-IAH[c]	$163	$125	$38	-	-
Two years pre-IAH	$170	$135	$35	-$3 ($5)	-2.1
Total unweighted number of observations across all years: 295,292					

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. We calculated percentage effects using the unadjusted IAH group mean in the year before the demonstration. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

[b]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[c]The difference-in-differences estimate for the year before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

IAH = Independence at Home; PBPM = per beneficiary per month; SNF = skilled nursing facility.

Exhibit B.5. Estimated effects of the IAH payment incentive on outcomes in Year 5 excluding one site at a time

	Medicare expenditures per beneficiary per month		Number of hospital admissions per beneficiary per year		Number of ED visits per beneficiary per year		Number of potentially avoidable hospital admissions per beneficiary per year	
Excluding site	Estimated effect	Percentage effect[a]	Estimated effect	Percentage effect[a]	Estimated effect	Percentage effect[a]	Estimated effect	Percentage effect[a]
A	-$410** ($189)	-9.2%	-0.16*** (0.06)	-8.6%	-0.20** (0.08)	-7.0%	-0.06*** (0.02)	-13.6%
B	-$403** ($198)	-9.2%	-0.17*** (0.06)	-9.9%	-0.24*** (0.09)	-8.4%	-0.07*** (0.02)	-15.3%
C	-$365* ($195)	-8.1%	-0.11* (0.06)	-6.2%	-0.22** (0.09)	-7.6%	-0.04* (0.03)	-9.5%
D	-$349** ($150)	-8.0%	-0.14** (0.07)	-8.1%	-0.20*** (0.07)	-6.8%	-0.06** (0.03)	-13.3%
E	-$341* ($183)	-7.7%	-0.13** (0.06)	-7.5%	-0.21** (0.08)	-7.1%	-0.06** (0.02)	-12.4%
F	-$340* ($189)	-7.7%	-0.15** (0.07)	-8.4%	-0.23*** (0.08)	-7.8%	-0.06** (0.03)	-13.3%
G	-$340* ($187)	-7.8%	-0.14** (0.07)	-8.0%	-0.22*** (0.08)	-7.5%	-0.06** (0.03)	-12.4%
H	-$338* ($183)	-7.7%	-0.13** (0.06)	-7.3%	-0.20** (0.08)	-7.1%	-0.06** (0.02)	-12.1%
I	-$335* ($185)	-7.6%	-0.12* (0.06)	-6.9%	-0.19** (0.08)	-6.6%	-0.05** (0.02)	-12.0%
J	-$333* ($184)	-7.5%	-0.13** (0.06)	-7.3%	-0.20** (0.08)	-6.8%	-0.06** (0.02)	-12.4%
K	-$332* ($187)	-7.6%	-0.13** (0.06)	-7.3%	-0.21*** (0.08)	-7.3%	-0.06** (0.02)	-12.3%
L	-$330* ($185)	-7.5%	-0.12* (0.06)	-7.0%	-0.20** (0.08)	-6.8%	-0.06** (0.02)	-12.4%
M	-$325* ($186)	-7.4%	-0.12* (0.06)	-6.7%	-0.19** (0.08)	-6.4%	-0.05** (0.02)	-11.7%
N	-$314 ($203)	-7.2%	-0.13* (0.07)	-7.6%	-0.23*** (0.08)	-8.1%	-0.05* (0.03)	-11.3%

Exhibit B.5 (*continued*)

	Medicare expenditures per beneficiary per month		Number of hospital admissions per beneficiary per year		Number of ED visits per beneficiary per year		Number of potentially avoidable hospital admissions per beneficiary per year	
Excluding site	**Estimated effect**	**Percentage effect[a]**	**Estimated effect**	**Percentage effect[a]**	**Estimated effect**	**Percentage effect[a]**	**Estimated effect**	**Percentage effect[a]**
O	-$298 ($191)	-6.8%	-0.12* (0.07)	-6.8%	-0.19** (0.08)	-6.6%	-0.05** (0.03)	-11.9%
P	-$111 ($151)	-2.6%	-0.02 (0.06)	-1.2%	-0.09 (0.07)	-3.0%	-0.03 (0.02)	-7.4%
Full sample	**-$330 ($182)**	**-7.5%**	**-0.13** (0.06)**	**-7.2%**	**-0.20** (0.08)**	**-6.9%**	**-0.06** (0.02)**	**-12.1%**

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: For each outcome in the table, we estimated 16 regressions (treating each member of the Richmond-based consortium separately), with each regression excluding the IAH beneficiaries from one practice and their matched comparisons in all years. Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights.

[a]We used the unadjusted treatment group mean in the year before the demonstration excluding the respective site to calculate the percentage effect for each demonstration year.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

ED = emergency department; IAH = Independence at Home.

Exhibit B.6. Estimated aggregate effects of the IAH payment incentive on total Medicare expenditures, full sample versus sample without site P

	Full sample		Sample without site P	
	Aggregate effect	90 percent CI	Aggregate effect	90 percent CI
Year 1	−$9,448,124	−$22,116,947; $3,220,699	$6,272,284	-$9,957,365; $16,890,005
Year 2	−$2,162,808	−$18,115,610; $13,789,994	$8,790,768	-$4,981,435; $22,562,971
Year 3	−$12,854,270	−$31,623,671; $5,915,131	$3,466,320	-$9,957,365; $16,890,005
Year 4	−$25,442,886	−$55,868,337; $4,982,565	-$3,699,143	-$29,672,089; $22,273,803
Year 5	-$31,350,990*	-$59,793,938; -$2,908,042	-$10,545,333	-$34,143,603; $13,052,937
Cumulative effect through Year 5	−$81,277,070	−$183,534,646; $20,980,506	$4,280,570	-$79,365,114; $87,926,254

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: These calculations are based on the beneficiary-level estimates shown in Exhibits B.2 and B.5, and the number of IAH beneficiaries and beneficiary months in each year.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

CI = confidence interval

Exhibit B.7. Estimated effect of the IAH payment incentive on hospital care use for IAH beneficiaries

Outcome	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Number of hospital admissions per beneficiary per year[b]					
Five-year average annual effect[c]	1.72	2.00	-0.28	-0.08 (0.05)	-4.6
Year 5	1.79	2.12	-0.33	-0.13** (0.06)	-7.2
Year 4	1.63	1.94	-0.30	-0.10 (0.07)	-5.7
Year 3	1.76	2.05	-0.29	-0.08 (0.06)	-4.7
Year 2	1.79	2.02	-0.23	-0.03 (0.05)	-1.7
Year 1	1.83	2.08	-0.25	-0.05 (0.04)	-2.8
One year pre-IAH[d]	1.90	2.10	-0.20	-	-
Two years pre-IAH	1.98	2.22	-0.24	-0.03 (0.03)	-1.9
Total number of ED visits per beneficiary per year[e]					
Five-year average annual effect[c]	2.98	3.14	-0.16	-0.14*** (0.06)	-4.9
Year 5	3.14	3.36	-0.22	-0.20** (0.08)	-6.9
Year 4	2.92	3.14	-0.22	-0.20** (0.08)	-7.1
Year 3	3.12	3.29	-0.17	-0.15* (0.09)	-5.3
Year 2	3.06	3.08	-0.02	-0.01 (0.06)	-0.3
Year 1	2.96	3.09	-0.13	-0.12 (0.07)	-4.0
One year pre-IAH[d]	3.08	3.10	-0.02	-	-
Two years pre-IAH	3.12	3.16	-0.04	-0.03 (0.05)	-0.9
Number of ED visits resulting in inpatient admission per beneficiary per year					
Five-year average annual effect[c]	1.38	1.53	-0.15	-0.11** (0.05)	-7.7
Year 5	1.45	1.64	-0.18	-0.15** (0.06)	-10.4
Year 4	1.30	1.47	-0.18	-0.14** (0.06)	-9.8
Year 3	1.41	1.57	-0.16	-0.13** (0.05)	-8.8
Year 2	1.43	1.53	-0.10	-0.07 (0.05)	-4.5
Year 1	1.50	1.60	-0.10	-0.07** (0.03)	-4.8
One year pre-IAH[d]	1.57	1.61	-0.03	-	-

Exhibit B.7 *(continued)*

Outcome	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Two years pre-IAH	1.66	1.71	-0.05	-0.01 (0.03)	-1.0
Number of outpatient ED visits per beneficiary per year[f]					
Five-year average annual effect[c]	1.60	1.62	-0.02	-0.02 (0.06)	-1.6
Year 5	1.68	1.73	-0.05	-0.06 (0.10)	-3.8
Year 4	1.61	1.66	-0.05	-0.06 (0.08)	-3.9
Year 3	1.70	1.71	-0.01	-0.02 (0.07)	-1.2
Year 2	1.64	1.56	0.08	0.07 (0.05)	4.9
Year 1	1.46	1.49	-0.03	-0.03 (0.05)	-2.4
One year pre-IAH[d]	1.50	1.50	0.01	-	-
Two years pre-IAH	1.48	1.47	0.00	0.00 (0.04)	-0.2
Total unweighted number of observations across all years: 295,292					

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

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

[c]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[d]The difference-in-differences estimate for the year before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

[e]The measure includes outpatient ED visits and visits resulting in inpatient admission.

[f]The number of outpatient ED visits measures all those not resulting in a hospital admission, including those resulting in an observation stay.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

ED = emergency department; IAH = Independence at Home.

Exhibit B.8. Estimated effect of the IAH payment incentive on potentially avoidable hospital admissions and outpatient ED visits, and probability of unplanned readmission

Outcome	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Number of potentially avoidable hospital admissions per beneficiary per year[b]					
Five-year average annual effect[c]	0.43	0.49	-0.07	-0.04** (0.02)	-8.1
Year 5	0.46	0.54	-0.09	-0.06** (0.02)	-12.1
Year 4	0.36	0.45	-0.09	-0.06** (0.02)	-12.8
Year 3	0.41	0.48	-0.07	-0.04* (0.02)	-7.9
Year 2	0.43	0.47	-0.04	0.00 (0.02)	-1.1
Year 1	0.46	0.51	-0.05	-0.01 (0.01)	-2.8
One year pre-IAH[d]	0.49	0.52	-0.03	-	-
Two years pre-IAH	0.51	0.57	-0.06	-0.03* (0.01)	-5.8
Number of potentially avoidable outpatient ED visits per beneficiary per year[e]					
Five-year average annual effect[c]	0.22	0.23	-0.01	-0.00 (0.01)	-0.7
Year 5	0.25	0.26	-0.02	-0.01 (0.01)	-3.6
Year 4	0.21	0.24	-0.03	-0.02 (0.01)	-9.3
Year 3	0.22	0.23	-0.01	0.01 (0.01)	2.6
Year 2	0.22	0.22	0.01	0.02 (0.01)	9.5
Year 1	0.20	0.21	-0.01	0.00 (0.01)	1.9
One year pre-IAH[d]	0.20	0.21	-0.01	-	-
Two years pre-IAH	0.20	0.21	0.00	0.01 (0.01)	3.5
Probability of having a qualifying hospital discharge and unplanned readmission within 30 days of discharge (percentage)[f]					
Five-year average annual effect[c]	15.39	17.32	-1.92	-1.13 (0.78)	-7.0
Year 5	15.86	18.14	-2.29	-1.47 (1.05)	-9.2
Year 4	13.65	15.97	-2.32	-1.50 (0.98)	-9.4
Year 3	16.18	18.70	-2.52	-1.70** (0.77)	-10.6
Year 2	16.32	18.14	-1.82	1.00 (0.98)	-6.3
Year 1	17.93	18.63	-0.70	0.12 (0.66)	0.7

Exhibit B.9 *(continued)*

Outcome	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
One year pre-IAH[d]	17.78	18.60	-0.82	-	-
Two years pre-IAH	19.88	20.89	-1.02	-0.20 (0.66)	-1.3
Total number of observations across the five years: 295,292					

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: A potentially avoidable hospital admission (or outpatient ED visit) is defined as one in which appropriate primary and specialty care may prevent or reduce the need for a hospital admission (or ED visit). Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

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

[c]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[d]The difference-in-differences estimate for the period before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

[e]The number of outpatient ED visits measures those not resulting in hospital admission, including those resulting in an observation stay. The measure excluded ED visits that led to an inpatient admission because there was no diagnosis from such a visit in a claim record when it led to an inpatient admission.

[f]We measured the probability of unplanned readmission for a beneficiary over the IAH-eligible months during each demonstration year. The probability equals zero for beneficiaries who did not have a qualifying hospital discharge or an unplanned readmission within 30 days of a qualifying hospital discharge during the measurement period.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

ED = emergency department; IAH = Independence at Home.

Exhibit B.9. Estimated effect of the IAH payment incentive on probability of dying within the demonstration year

	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Five-year average annual effect[b]	13.90	19.53	-5.63	-0.32 (0.55)	-2.2
Year 5	13.90	20.58	-6.68	-1.20* (0.70)	-8.1
Year 4	13.74	20.52	-6.77	-1.29** (0.63)	-8.7
Year 3	15.43	19.88	-4.44	1.04 (0.68)	7.0
Year 2	14.71	19.86	-5.15	0.33 (0.67)	2.2
Year 1	15.36	20.83	-5.47	0.02 (0.56)	0.1
One year pre-IAH[c]	15.54	21.03	-5.48	-	-
Two years pre-IAH	14.94	22.05	-7.11	-1.62*** (0.64)	-10.9

Total unweighted number of observations across all years: 295,292

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. Regression-adjusted means are obtained by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

[b]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[c]The difference-in-differences estimate for the year before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

IAH = Independence at Home.

Exhibit B.10. Estimated effect of the IAH payment incentive on probability of entering institutional long-term care within the demonstration year

	IAH	Comparison	Difference (IAH - comparison)	Difference-in-differences estimated effect	Percentage effect[a]
Four-year average annual effect[b]	6.77	12.79	-6.02	0.38 (0.37)	5.5
Year 4	6.51	12.78	-6.26	0.15 (0.51)	2.1
Year 3	7.44	13.26	-5.82	0.59 (0.49)	8.4
Year 2	7.10	12.76	-5.66	0.75 (0.47)	10.7
Year 1	7.46	13.49	-6.03	0.38 (0.47)	5.5
One year pre-IAH[c]	7.88	14.29	-6.41	-	-
Two years pre-IAH	8.03	15.14	-7.11	-0.70 (0.49)	-10.1

Total unweighted number of observations across all years: 243,947

Source: Mathematica's analysis of data from the IAH vz1implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

[b]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across four demonstration years.

[c]The difference-in-differences estimate for the year before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

IAH = Independence at Home.

This page has been left blank for double-sided copying.

Appendix C

Estimated effect of IAH on Medicare expenditures using an alternative weighting method

This page has been left blank for double-sided copying.

I. OVERVIEW

In the main analysis, we estimated the effects of the demonstration payment incentive by allowing each beneficiary to be counted equally, regardless of which Independence at Home (IAH) practice the beneficiary visited. This method, which we refer to as the beneficiary-weight method, produces an estimate that reflects the average effect of changes that practices made on outcomes for the average Medicare beneficiary in the demonstration (details of the analytic weight construction using the beneficiary-weight method are described in Appendix B, Section VI). In this appendix, we present results on total Medicare expenditures when we reweighted the beneficiary-level data so every practice in the demonstration was given equal weight. This method, which we refer to as the practice-weight method, yields an estimate that reflects the average effect of changes that practices made in response to the payment incentive. The beneficiary-weight method allows larger practices to have more influence on the result, and the practice-weight method requires each practice to have equal influence on the results.

An example may help explain the difference between the two methods. Let us assume that the demonstration had 4 large practices of 1,000 patients each that did not change care delivery in response to the demonstration, and 12 practices of 250 patients each that changed care delivery. The practice-weight method would give equal influence to all practices, and the estimate would reflect that most practices (12 out of 16) changed care delivery. The beneficiary-weight method, on the other hand, would give more influence to the 4 practices that served the majority of the beneficiaries (4,000 out of 7,000), and the estimate would reflect that the majority of the beneficiaries were treated by practices that did not change care delivery.

As the example above illustrates, when the effect of the demonstration differs across IAH practices, the beneficiary-weight method would lead to an estimated effect that tends to resemble those of the largest practices, thus masking the effect on smaller practices. The beneficiary-weight estimate reported in Chapter II can be informative for policymaking if the largest IAH practices are a representative sample of the largest home-based primary care providers in the nation. However, although the IAH practices were selected to include diverse approaches to providing home-based primary care, they do not represent all practices in the nation that provide home-based primary care. Therefore, the practice-weight estimate provides an important understanding of the average effect of the demonstration across a variety of delivery models for home-based primary care.

To implement the practice-weight method, we re-weighted the final analytic weights for all IAH and matched comparison beneficiaries using a ratio that varied by site and year so that the summed weights among all beneficiaries in each practice were equal across all 16 practices in each year (treating the three members of the Mid-Atlantic Consortium as separate sites). We then re-estimated the effect on total Medicare expenditures as specified in Equations (1) and (2) of Appendix B using this new weight.

II. COMPARISON OF RESULTS FROM BENEFICIARY-WEIGHT AND PRACTICE-WEIGHT METHODS

Overall, the estimated effect of the demonstration payment incentive on Medicare expenditures was smaller under the practice-weight method than the beneficiary-weight method. Using beneficiary weighting, the estimated effect of IAH on total Medicare expenditures was an annual reduction of $200 (4.6 percent) per beneficiary per month (PBPM) over the five years (Exhibit C.1). The practice-weight estimate was a much smaller annual reduction of $86 (2.0 percent) PBPM. Neither estimate was statistically significant. In Year 5, the beneficiary-weight estimate showed a statistically significant reduction of $330 PBPM for IAH beneficiaries (7.5 percent). However, this result differed from the result using the practice-weight method, which showed a much smaller, not statistically significant reduction of $131 PBPM (3.0 percent).

Because the practice-weight method requires each practice to have equal influence on the results, whereas the beneficiary-weight method allows larger practices to have more influence on the result, the difference in Medicare expenditures between methods suggests that large practices strongly influenced the estimated effect of the demonstration payment incentive, particularly in Year 5. As discussed in Chapter II, further analysis showed that one large site strongly influenced the estimated effect of the payment incentive on expenditures in Year 5.

Exhibit C.1. Estimated effect of the IAH payment incentive on total Medicare expenditures PBPM for IAH beneficiaries, using beneficiary and practice weighting

	Beneficiary weighting		Practice weighting	
	Estimated effect	Percentage effect[a]	Estimated effect	Percentage effect[a]
Five-year average annual effect[b]	-$200 ($151)	-4.6%	-$86 ($99)	-2.0%
Year 5	-$330* ($182)	-7.5%	-$131 ($131)	-3.0%
Year 4	-$282 ($205)	-6.4%	-$179 ($130)	-4.1%
Year 3	-$178 ($158)	-4.1%	-$69 ($102)	-1.6%
Year 2	-$31 ($139)	-0.7%	$98 ($110)	2.2%
Year 1	-$119 ($97)	-2.7%	-$102 ($125)	-2.3%
One year pre-IAH[c]	-	-	-	-
Two years pre-IAH	-$33 ($57)	-0.8%	-$16 ($109)	-0.4%

Total unweighted number of observations across all years: 295,292

Source: Mathematica's analysis of data from the IAH implementation contractor and 2009–2017 Medicare claims and enrollment data from the Chronic Conditions Warehouse for IAH and matched comparison group beneficiaries in all IAH practices that participated in Year 5.

Notes: Standard errors are given in parentheses. We computed coefficients and standard errors by using the weighted sample size, which considers both the matching and eligibility weights. The table reports the regression-adjusted means of the IAH and matched comparison groups in each year. We obtained regression-adjusted means by applying the estimated regression coefficients to the covariates of IAH beneficiaries in the latest demonstration year. Because of rounding, a difference-in-differences estimate displayed as zero might be shown alongside a percentage effect that exceeds zero.

[a]We used the unadjusted treatment group mean in the year before the demonstration to calculate the percentage effect for each demonstration year. Exhibit B.1 reports the baseline unadjusted treatment group mean for all outcomes.

[b]We estimated a separate model using a single demonstration indicator (instead of separate indicators for each demonstration year) and used its interaction with treatment status to obtain an average annual estimated effect across five demonstration years.

[c]The difference-in-differences estimate for the period before the demonstration is zero (with no standard error) in all regressions because we calculated that estimate for each year as the difference in means between IAH and comparison beneficiaries in that year minus the difference in the year before the demonstration.

*/**/*** The difference is statistically significant at the 0.10/0.05/0.01 level.

IAH = Independence at Home; PBPM = per beneficiary per month.

This page has been left blank for double-sided copying.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org