Evaluation of the Home Health Value-Based Purchasing (HHVBP) Model

First Annual Report: Quantitative Technical Appendix

Arbor Research Collaborative for Health
and L&M Policy Research

Prepared by:
Alyssa Pozniak, Marc Turenne, Purna Mukhopadhyay, Brant Morefield, Vladislav Slanchev, Kathryn Linehan, Lisa Green, Betsy Brigantti, Nicholas Byrd, Chad Cogan, Tony Cuttitta, Zach Dietrich, Julia Doherty, Diana Haggerty, Katherine Hanslits, Nan Ji, Yan Jin, Margaret Edder Johnson, Scott Liening, Kaden Milkovich, Danielle Norman, Jeffrey Pearson, Kaitlyn Repeck, Sarah Rosenblum, Jillian Schrager, Claudia Schur, Amanda Szymanski, and Lisa Tomai

Prepared for The Centers for Medicare & Medicaid Services (CMS)
Center for Medicare & Medicaid Innovation (CMMI)

July 2018

ARBOR RESEARCH
COLLABORATIVE
FOR HEALTH

NOTICE
The statements contained in this report are those of the authors and do not necessarily reflect the policies or views of the Centers for Medicare & Medicaid Services. Arbor Research Collaborative for Health assumes responsibility for the accuracy and completeness of the information contained in this report.

Table of Contents

Tables

Acronym List

Acronym	Term
ACH	Acute Care Hospitalization
AF	Analytic File
AHRQ	Agency for Healthcare Research and Quality
AHRF	Area Health Resource File
AT	Achievement Threshold
BM	Benchmark
CBSA	Core-Based Statistical Area
CCN	CMS Certification Number
CCS	Clinical Classifications Software
CCW	Chronic Conditions Data Warehouse
CME	Common Medicare Environment
CMS	Centers for Medicare and Medicaid Services
CY	Calendar Year
D-in-D	Difference-in-Differences
ED	Emergency Department
EDB	Medicare Enrollment Database
FFS	Fee-for-Service
FIPS	Federal Information Processing Standards
HH	Home Health
HHA	Home Health Agency
HHCAHPS	Home Health Consumer Assessment of Healthcare Providers and Systems
HHVBP	Home Health Value-Based Purchasing
HMO	Health Maintenance Organization
HUD	U.S. Department of Housing and Urban Development
ICD	International Classification of Diseases
IPF	Inpatient Psychiatric Facility
IRF	Inpatient Rehabilitation Facility
LTCH	Long-Term Care Hospital
LUPA	Low Utilization Payment Adjustment
MBSF	Master Beneficiary Summary File
OASIS	Outcome and Assessment Information Set
PEP	Partial Episode Payment
POS	Provider of Services
QIES	Quality Improvement and Evaluation System
RIF	Research Identifiable File
ROC	Resumption of Care
SNF	Skilled Nursing Facility
SOC	Start of Care
TPS	Total Performance Score

Introduction

This Technical Appendix provides details about the data sources and methods used to conduct the quantitative analyses for this Annual Report. Each of the following topics is covered below:

1. Detailed discussion of our quantitative analytic approach, including a discussion of how we created our comparison groups and our difference-in-differences framework (see **Section 1, "Analytic Approach"**)
2. Detailed explanations of the descriptive variables and impact measures that are presented in the Annual Report, covering Quarter 1, 2012 through Quarter 4, 2016 (see **Section 2, "Descriptive Variable and Impact Measure Definitions"**)
3. Information about data acquisition and processing to create the Analytic Files (AFs) that are necessary to define the impact measures of interest and conduct the analyses for this Annual Report (see **Section 3, "Data Sources"**)
4. Step-by-step discussion of how we created the AFs that we used to generate the results presented in this Annual Report (see **Section 4, "Analytic File Creation"**)
5. Summary of our falsification tests and parallel trend tests results that informed our comparison groups used in this Annual Report (see **Section 5, "Comparison Group Testing Results"**)
6. Presentation of a Glossary (see **Section 6, "Glossary"**)

1. Analytic Approach

We designed our quantitative analysis to address the question: What was the effect of the Home Health Value-Based Purchasing (HHVBP) Model on impact measures of interest such as health care utilization, quality of health care, health outcomes, and health care costs? To evaluate the impact of HHVBP by comparing the experience of beneficiaries and home health agencies (HHAs) in both intervention and non-intervention states, our empirical model must address differing characteristics of beneficiaries and HHAs between intervention and non-intervention groups. Our analyses use data from multiple sources (described in Section 3) to estimate impacts of HHVBP on the cumulative impact of HHVBP across the nine intervention states. Most of these data elements are available before and after the start of the HHVBP Model for both intervention HHAs and non-intervention HHAs. This allows for both comparing outcomes between the HHVBP intervention HHAs and non-intervention HHAs and assessing whether the relative outcomes for these two groups changed from before to after the start of the HHVBP Model.

Below, we describe the descriptive variables and impact measures used in this report. We then describe our overall analytic approach to construct comparison groups for the impact measures.

1.1 Descriptive Variables

We analyzed descriptive variables (described in Section 2.1) to assess patterns and trends among HHVBP states in the overall utilization of home health services among Medicare fee-for-service (FFS) beneficiaries and basic characteristics of HHAs and home health episodes. We reviewed annual comparisons of the descriptive measures between HHVBP and non-HHVBP states. For individual HHVBP states, comparisons are made with other states in the same regional grouping for monitoring basic patterns and trends over time in the overall use and delivery of home health care in HHVBP states relative to nearby home health markets where HHVBP was not implemented. We combined results for all nine HHVBP states and compared them to findings for the remaining 41 non-HHVBP states as an

indicator of whether any overall differences may be emerging nationally between HHVBP and non-HHVBP states.

Table 1 defines the regional groupings from the calendar year (CY) 2016 Final Rule that correspond to each of the nine HHVBP states and were used for comparison purposes for the descriptive measures.[1] For the regional groupings and the pooled HHVBP group, results for descriptive measures correspond to a weighted average across individual states, where results for each state are weighted to reflect the number of FFS beneficiaries, HHAs, or episodes in that state and year.

Table 1. Regional Groupings

HHVBP State	States in Regional Grouping
AZ	NM, CA, NV, UT, CO
FL	TX, OK, LA, MS
IA	ND, SD, MT, WI, MN
MA	VT, ME, CT, RI, NH
MD	DE, NJ, PA, NY
NC	AL, GA, SC, VA
NE	OH, WV, IN, MO, KS
TN	IL, KY, AR, MI
WA	OR, AK, HI, WY, ID

1.2 Impact Measures

We see two general reasons why outcomes may differ across intervention and non-intervention states: 1) differing observed characteristics of beneficiaries and HHAs studied; and 2) differing *un*observed characteristics of beneficiaries and HHAs. Our empirical strategies use information on observed characteristics to address differences between the treated populations and selected comparison. Specifically, these strategies include an approach to establishing comparison groups to address observed differences and the use of a difference-in-differences (D-in-D) framework to address unobserved differences.

1.2.1 D-in-D Approach for Impact Measures

We use a D-in-D framework to compare changes in impact measures observed over time in the HHVBP states relative to those in the comparison group as the basis for evaluating the effects of HHVBP. The D-in-D framework offers a quasi-experimental design that can address many threats to validity, and rests on the critical assumption that, in the absence of the HHVBP Model, the impact measures in the two groups would have changed in a parallel manner over time. Our D-in-D analysis compared changes in impact measures observed over time in the HHVBP states combined to corresponding changes in the comparison group. The basic D-in-D estimate is defined as the difference in an outcome of interest over time in the Model states, after subtracting the difference, over time, in the comparison group:

$$D\text{-}in\text{-}D = [Y_{INT,POST} - Y_{INT,PRE}] - [Y_{COMP,POST} - Y_{COMP,PRE}]$$

---

[1] HHS, CMS. (2015) 42 CFR 409, 424, 484. Medicare and Medicaid Programs; CY 2016 Home Health Prospective Payment System Rate Update; Home Health Value-Based Purchasing Model; and Home Health Quality Reporting Requirements; Final Rule. Federal Register 80 FR 68623. November 5, 2015. Accessed from here.

where $Y_{INT,POST}$ and $Y_{INT,PRE}$ are the post- and pre-intervention outcome levels, respectively, for the intervention group, and $Y_{COMP,POST}$ and $Y_{COMP,PRE}$ are the post- and pre-intervention outcome levels, respectively, for the comparison group.

With this model specification, the impact estimate is the differential change in an outcome for the HHVBP states between the baseline and follow-up period(s), relative to that same change for the comparison group. That is, the differential change in the outcome over time for the HHVBP states relative to its comparison group represents the estimated effect of HHVBP. The D-in-D design controls for unobserved, time-varying changes that are common to all beneficiaries (i.e., cyclical or seasonal trends or broader changes in the health system) as well as time-invariant, unmeasured differences between HHVBP and comparison states' markets and beneficiary populations. Moreover, through the use of a multivariate regression, we are able to adjust for observed characteristics of beneficiaries influencing the outcome.

For all of the impact measures of interest for the Annual Report, we used a D-in-D approach to estimate effects of the Model for all HHVBP states combined.[2] The exact D-in-D model specification that was used was adapted for the specific measures of interest and corresponding comparison group approach, and is described in further detail below (see Sections 1.4, 1.5, and 1.6).

1.3   Construction of Comparison Groups for Impact Measures

1.3.1   Motivation

We designed the quantitative analyses for this Annual Report to evaluate the effect of the HHVBP Model on a wide range of impact measures that include Medicare spending, utilization of services, quality of care, and patient experience. These analyses rely on comparisons for both beneficiaries and agencies between HHVBP and non-HHVBP states. As an important aspect of the Model design, the mandatory participation of all HHAs in the nine randomly selected HHVBP states helps to guard against selection bias.

However, unlike hospitals, HHAs reflect a particularly transient provider type.[3] Therefore, selection bias is still a potential issue, as entry and exit of HHAs into and out of a market after the start of HHVBP is necessarily endogenous to the presence of HHVBP in the intervention states; that is, entry/exit market changes are potentially outcomes of the intervention. Furthermore, mergers, acquisitions, and other agency-level changes are endogenous to the implementation of HHVBP in the intervention states. These endogenous agency-level changes need to be considered when evaluating the impact of HHVBP on the use of home health care and quality of home health being provided to beneficiaries. For instance, when new HHAs enter HHVBP states during the operation of the Model, the type of agency entering/leaving the market may differ in both observed and unobserved ways than would have been the case without the introduction of the Model. Given comparatively frequent turnover among HHAs prior to implementation of the Model and the expected lower barrier to entry/exit among HHAs relative to

---

[2] We are unable to use the D-in-D framework for the three new measures that are self-reported by HHAs via the Secure Web Portal since these data are only available for HHAs in the HHVBP states. As such, we instead focus on reporting rates among HHAs in the nine HHVBP states.

[3] To illustrate, in 2012 we identified 11,706 agencies with OASIS assessments in the year. By 2016, nationally, 1,906 agencies had closed (16.2%) and 1,489 new HHAs existed (12.7% relative to the original population).

other institutional providers, risks related to endogenous entry/exit are likely to be more pronounced in this evaluation.

Defining the treatment population, and an accompanying comparison population, as the population of beneficiaries receiving home health care alleviates the fundamental endogeneity concerns related to entry and exit of HHAs. That is, we examine the impact of care on beneficiaries within the treatment and comparison states, rather than finding comparisons for HHAs in the treatment state that may enter or exit the market due to the presence of HHVBP. Further, this definition is appropriate for examining changes in care that are provided to beneficiaries ultimately under the care of HHAs. However, defining the populations of interest as either HHAs, or beneficiaries who have received home health care, will not capture beneficiaries in the broader Medicare FFS population, to examine, for instance, the effect of HHVBP on which beneficiaries ultimately receive home health care. Since the evaluation is also interested in outcomes experienced in the overall Medicare FFS population, the evaluation must balance competing priorities: to examine effects of HHVBP on the broader FFS population as well as on users of home health care.

We investigated three ways to use geography to create one experimental sample (state-level matching, state-level reweighting, county-level reweighting). However, given the number of outcomes considered, the heterogeneity of HHVBP states, and the requirement that the evaluation produce both overall and state-level impact estimates, we were unable to identify an approach that would create a suitable comparison across even a majority of outcomes and across all treatment states. Due to the use of a D-in-D model as the primary estimator of HHVBP effects, and its parallel trend assumption, we used tests of parallel trends over the baseline period as a primary test of suitability.

1.3.2 State- or County-level Reweighting

We explored setting weights using entropy balancing. Entropy balancing is a procedure designed to find weights that allow the comparison population to exactly match the treatment population on the mean values of covariates entered into the procedure.[4] The evaluation must define the set of variables over which to match the treatment and comparison populations. Of note is that the variables that are relevant to reweighting in order to examine one impact measure may vary considerably from those appropriate for other impact measures. The use of different variables for reweighting different impact measures yields differing comparison populations and moves away from a unified framework, and also leads to a more complex approach.

To examine the feasibility of entropy balancing in generating comparison populations, we designed a program to explore which variables were related to the trends in nine HHVBP impact measures for the nine HHVBP states.[5] Specifically, we included varying combinations of characteristics, including episode characteristics, beneficiary characteristics, risk characteristics, provider characteristics, and market characteristics. Below, we provide a list of variables included in our final consideration. For the varying combinations, we created weighted comparison populations using entropy balancing.

---

[4] The procedure may also aim to match the treatment and comparison populations on other moments of the distribution of covariates (e.g., the first and second moments).

[5] The nine impact measures included: ED use, without hospitalization, per first FFS HH episodes; unplanned hospitalizations per first FFS HH episodes; improvement in ambulation-locomotion; improvement in bed transferring; improvement in bathing; improvement in dyspnea; improvement in management of oral medications; improvement in pain interfering with activity; and discharged to community.

Covariates for Consideration in Entropy Balanced Comparisons

Episode	Beneficiary	OASIS Risk Adjustment Factors	Claims Risk Adjustment Factors
HHRG Clin 1	White	Toilet Trans Min	No Prior Care
HHRG Clin 2	Black	Toilet Trans Med	ACH Prior Care (Any Days)
HHRG Clin 2	Hispanic	Toilet Trans High	Inpat 9+ Days
HHRG Func 1	Asian	Overall Status 1	SNF Prior Care (Any Days)
HHRG Func 2	Other Race	Overall Status 2	ED Prior Care (Any Days)
HHRG Func 3		Overall Status 3	IRF Prior Care (Any Days)
HHRG Serv 1	**Market**	Ambulation 1	IPF Prior Care (Any Days)
HHRG Serv 2	Rural	Ambulation 2	LTC Prior Care (Any Days)
HHRG Serv 3	Median HH Income, 2014	Ambulation 3	IP Medical 1-3 days
HHRG Serv 4	Unemp Rate, 2015	Ambulation 3, 4, 5	IP Medical 4-8 days
HHRG Serv 5	< HS Diploma 2010-2014		IP Medical 9+ days
HHRG Group 1	College Degree 2010-2014		IP Neur 1-3 days
HHRG Group 2	Work Travel Time 2010-2014		IP Neur 4-8 days
HHRG Group 3	Population Density 2010		IP Neur 9+ days
HHRG Group 4			IP Surgical 1-3 days
HHRG Group 5	**Agency**		IP Surgical 4-8 days
Outlier	Non-profit		IP Surgical 9+ days
PEP	For-profit		IP Cardio 1-3 days
LUPA	Gov't owned		IP Cardio 4-8 days
	Freestanding		IP Cardio 9+ days
	1-59 Episodes		IP Resp 1-3 days
	60-249 Episodes		IP Resp 4-8 days
	250-499 Episodes		IP Resp 9+ days
	500-999 Episodes		SNF Prior Care 1-13 Days
	1000+ Episodes		SNF Prior Care 14-41 Days
			SNF Prior Care 42+ days

Because the entropy balancing procedure may not converge on (that is, find), a solution for a complete set of variables, and narrowing the set to achieve convergence would necessitate that we choose the variables most important for reweighting, we included principal components of the variable lists in the balancing procedure. That is, for each potential set of variables, we balanced over the principal components obtained from a principal component analysis of these variables. If the procedure converges over all variables, then the procedure would also converge over all principal components. If

no convergence is found, the balancing procedure can capture the *most* variance among this set of variables. Thus, one has balanced over as much of the variance among the variables as is possible.

We iterated through the entropy balancing procedure for each state, including an additional principal component in each iteration, until the entropy balancing procedure failed to converge on a solution. The results proved that the number of principal components, or variables, for which the entropy balancing procedure could find a suitable solution varied by state. Movement toward a unified approach, such as using the same set of variables for all states or the same number of principal components, from this "best" position would worsen the performance of the comparison population. Furthermore, implementing entropy balancing is necessarily complicated and would need to extend to a broader array of outcome measures. Therefore, we concluded that a more direct approach to creating the comparisons was warranted.

1.3.3 Hybrid Approach

To accommodate the differing data sources, units of analysis, and denominator populations, we used a hybrid comparison group strategy that uses one of three approaches to selecting comparisons, depending on the underlying population (denominator) of interest for this report. The denominators for some impact measures require that we examine changes within a geographic area, such as those measured over the FFS population. The comparison population for measures impacting this population is defined geographically, through reweighting of comparison counties using entropy balancing.[6]

The denominators for some impact measures require that we examine the population of beneficiaries receiving home health care. This population may include Medicare or Medicaid beneficiaries enrolled in FFS or managed care plans and differ across measures of interest, as the requirements for inclusion in measure denominators differ. It is important, however, to appropriately risk-adjust outcomes for these populations. As such, we directly match comparison episodes to HHVBP episodes included in each measure on risk factors related to the outcome under consideration. The majority of impact measures in the first annual report are considered under this approach.

Finally, some impact measures can only be identified at the HHA level. Episode-level measures are aggregated to the agency level when necessary. The comparison population for HHVBP HHAs consists of reweighted HHAs from non-HHVBP states. The reweighting balances the distribution of the 2015 Total Performance Score (TPS) and a number of other agency characteristics in the HHVBP and comparison populations.

Table 2 maps each of the Annual Report metrics to the relevant comparison group approach. Additional details regarding the unit of analysis for the D-in-D models and the baseline period for analysis for each of the three approaches are included in the following subsections.

We acknowledge that the hybrid approach used in this annual report is complex. Because of its complexity, for future reports, we will be exploring strategies for unifying our comparison group methodologies in pursuit of a simpler approach. The results in this report should be considered preliminary pending implementation in future reports of a more unified comparison group methodology.

---

[6] Counties provide geographic units contained within states, which are boundaries within which HHAs typically operate.

Table 2. Comparison Group Approach Use for Each Impact Measure Analyzed

Measure	Comparison Group Approach	Unit of Analysis	Baseline Period
Utilization and Spending Measures:			
Average Number of FFS Home Health (HH) Episodes/1,000 FFS Beneficiaries	County-Level Reweighting (Comparison Group A)	County-Level	2012-2015
Average Medicare HH Spending/FFS Beneficiary	County-Level Reweighting (Comparison Group A)	County-Level	2012-2015
Average Medicare HH Spending/FFS HH Episode	County-Level Reweighting (Comparison Group A)	County-Level	2012-2015
ED Use (no Hospitalization)/First FFS HH Episodes (%)	Episode Matching (Comparison Group B)	FFS Episode	2012-2015
Unplanned Acute Care Hospitalization (ACH)/First FFS HH Episodes (%)	Episode Matching (Comparison Group B)	FFS Episode	2012-2015
Unplanned ACH/All FFS HH Episodes (%)	Episode Matching (Comparison Group B)	FFS Episode	2012-2015
Skilled Nursing Facility (SNF) Use/All FFS HH Episodes (%)	Episode Matching (Comparison Group B)	FFS Episode	2012-2015
Average Medicare Spending for SNF Stays/All FFS HH Episodes	Episode Matching (Comparison Group B)	FFS Episode	2012-2015
Average Medicare Spending for Unplanned ACH/FFS HH Episodes	Episode Matching (Comparison Group B)	FFS Episode	2012-2015
HHVBP OASIS-Based Outcome Measures:			
Improvement in Bathing	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
Improvement in Bed Transferring	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
Improvement in Ambulation-Locomotion	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
Improvement in Dyspnea	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
Improvement in Management of Oral Medications	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
Improvement in Pain Interfering with Activity	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
Discharged to Community	Episode Matching (Comparison Group B)	OASIS Episode	2012-2015
HHVBP Patient Experience and OASIS-Based Process Measures:			
Drug Education on Medications Provided to Patient/Caregiver during Episodes of Care	HHA Reweighting (Comparison Group C)	Agency-quarter	2015
Influenza Immunization Received for Current Flu Season	HHA Reweighting (Comparison Group C)	Agency-quarter	2015
Pneumococcal Polysaccharide Vaccine Ever Received	HHA Reweighting (Comparison Group C)	Agency-quarter	2015
How often the home health team gave care in a professional way	HHA Reweighting (Comparison Group C)	Agency-year	2015
How well did the home health team communicate with patients	HHA Reweighting (Comparison Group C)	Agency-year	2015
Did the home health team discuss medicines, pain, and home safety with patients	HHA Reweighting (Comparison Group C)	Agency-year	2015

Measure	Comparison Group Approach	Unit of Analysis	Baseline Period
How do patients rate the overall care from the home health agency	HHA Reweighting (Comparison Group C)	Agency-year	2015
Would patients recommend the home health agency to friends and family	HHA Reweighting (Comparison Group C)	Agency-year	2015
TPS	HHA Reweighting (Comparison Group C)	Agency-year	2015

Note that we do not include the three new HHA-reported measures (Influenza Vaccination Coverage for HHA Personnel; Herpes zoster [shingles] vaccination for patient; Advance Care Plan) since these data are only available for HHAs in the HHVBP states. Note that the duration of OASIS episodes of care may differ from that of Medicare FFS episodes.

1.4   Comparison Group Approach A: County-Level Reweighting

One possible effect of HHVBP is that it could lead to changes in the population of beneficiaries receiving home health or substitute types of care. To explore this possibility, we evaluated changes induced by HHVBP on measures without conditioning on the receipt of home health. These effects may be reflected over the population of FFS beneficiaries since they may receive home health care at a later time.

1.4.1   D-in-D Model

The estimated treatment effects are generated from a D-in-D model, even though any pre-existing differences between the intervention and comparison states are virtually nonexistent, since comparison states were reweighted to target baseline levels of the impact measures in HHVBP states. We defined the baseline period to be 2012-2015 and estimated cumulative treatment effects from a D-in-D model for each of the four quarters in 2016. To obtain quarterly cumulative effects in the post-intervention period, we restricted the estimation sample to include observations up to the quarter of interest (i.e., quarter 1, 2, 3, or 4 in 2016). Defining each county, C, in quarter q and year t, identifying the intervention state counties with an indicator variable $Treat_C$, and identifying the post-intervention time periods with an indicator variable $I(t \geq 2016)$, which takes a value of 1 for post-intervention time periods and 0 otherwise, the D-in-D estimator for outcome Y is implemented as:

$$Y_{C,q,t} = \beta_0 + \beta_1 * Treat_C + \beta_2 * I(t \geq 2016) + \beta_3 * Treat_C * I(t \geq 2016) + \sum_{j=1}^{j=3} \delta_j I(q = j) + \epsilon_{C,q,t}$$

where β_0 is an intercept, β_1 is the average difference between the treatment and comparison populations over the baseline period, β_2 is the average change from baseline to post-intervention for the comparison population, β_3 is the difference in change from baseline to post-intervention for the treatment population, relative to the comparison population—the estimated treatment effect of HHVBP, and δ_j coefficients capture seasonal effects in each quarter of the year. In the regression equation, we have only three estimates capturing quarterly effects (δ_1, δ_2 and δ_3) since we include a constant in the equation. As described above, we obtained four estimated treatment effects for the four quarterly cumulative effects in the post-intervention period. Each county in the intervention state is weighted to reflect the number of FFS beneficiaries residing in the county in the quarter or the number of home health episodes occurring in county in the quarter. The comparison county weights are normalized such that the sum of comparison weights equals the sum of treatment weights. Standard errors are clustered at the county level.

1.5   Comparison Group Approach B: Home Health Episode Matching for Measures Corresponding to Home Health Users

While HHVBP may lead to changes in the population of beneficiaries receiving home health care and accompanying risk profiles, an estimator of the effect of HHVBP on the quality of care provided to beneficiaries should accurately control for altered risk profiles for these impact measures. Approach B matches treatment and comparison beneficiaries' episodes directly, over observed characteristics that were used in the risk adjustment process for the two HHVBP claims-based measures themselves (unplanned ACH and Emergency Department [ED] measures).[7] This matching process ensured appropriate comparisons for the impact measures. For the other four claims-based utilization and spending measures not included as performance measures in the HHVBP Model, comparisons were matched on the same set of variables used in risk adjustment for the closely related unplanned ACH and ED impact measures (e.g., the SNF impact measures use the same risk adjustment profiles as the unplanned ACH and ED impact measures).

Both the HHVBP and comparison populations vary across the impact measures studied under this approach. By matching the changing characteristics of the treatment population across impact measures, the comparison group reflects those treated, and we measure an average effect on the treated population observed for each impact measure.

A subset of impact measures can be identified only for Medicare FFS beneficiaries because claims data are necessary, while other impact measures are identified for a broader population – Medicare and Medicaid, FFS, or managed care – and are limited to Outcome and Assessment Information Set (OASIS) assessment information.

Furthermore, because some impact measures are defined using the first episode only, and other impact measures are defined over the entire sequence of home health episodes, we allowed the comparison population to differ for these two types of measures. As discussed in Section 4.2 below, the impact measures that were defined using only the first episode had the same matched comparison, and the impact measures that were defined using all episodes had the same matched comparison.

Varying impact measures also carry varying requirements for inclusion. For instance, OASIS improvement measures have different denominators that reflect measure-specific inclusion and exclusion criteria (e.g., home health episodes of care for which the patient was able to ambulate independently at start/resumption of care [SOC/ROC] were excluded from the improvement in ambulation-locomotion measure, whereas episodes of care for which the patient was not short of breath at any time at SOC/ROC were excluded from the improvement in dyspnea measure). To achieve one comparison match for treatment beneficiaries across outcomes, one would need identical eligibility for each of the outcomes considered for the treatment beneficiary (e.g., improvement in dyspnea and improvement in ambulation/locomotion) and similar risk profiles for each of the impact measures, which were not feasible given varying inclusion and exclusion criteria. Henceforth, the evaluation matches impact measures separately.

---

[7] See CMS Specifications for Home Health Claims-Based Utilization Measures.

1.5.1   D-in-D Model

We defined the baseline period to be 2012-2015 and estimated average treatment effects from a D-in-D model through each of the four quarters of 2016, with the fourth quarter estimate representing the average effect for the full year (2016). To obtain average effects in the post-intervention period, we restricted the estimation sample to include observations through the quarter of interest (i.e., quarter 1, 2, 3, or 4 in 2016). Defining each episode *i* in quarter *q* and year *t*, identifying the treatment episodes with an indicator variable $Treat_i$, identifying the post-intervention time periods with an indicator variable I (t≥2016), which takes a value of 1 for post-intervention time periods and 0 otherwise, and identifying a vector of covariates as $\mathbf{P}_{Cov}$ (defined below for claims and OASIS measures separately), the D-in-D estimator for outcome Y is implemented as:

$$Y_{i,q,t} = \beta_0 + \beta_1 * Treat_i + \beta_2 * I(t \geq 2016) + \beta_3 * Treat_i * I(t \geq 2016) + \sum_{j=1}^{j=3} \delta_j I(q = j) + \omega \mathrm{P}_{Cov} + \epsilon_{i,q,t}$$

where β_0 is an intercept, β_1 is the average difference between the treatment and comparison populations over the baseline period, β_2 is the average change from baseline to post-intervention for the comparison population, β_3 is the difference in the change from baseline to post-intervention for the treatment population relative to the comparison population (i.e., the estimated treatment effect of HHVBP), δ_j coefficients capture seasonal effects associated with the four quarters of the year, and ω is a vector of coefficients associated with vector of covariates $\mathbf{P}_{Cov}$. In the regression equation, we have only three estimates (δ1, δ2, and δ3) capturing quarterly effects since we included a constant in the equation. As described above, we obtained four estimated treatment effects measuring the average treatment effect from the beginning of 2016 until the end of each respective quarter in the post-intervention period. Each episode was given an equal weight, and standard errors were clustered at the agency-level.

We followed this D-in-D specification for the nine intervention states separately, and we then constructed the overall treatment effect to account for the differences in the size of the population using HHA services across the states. The overall treatment effect was calculated as a weighted average of the nine state-specific treatment effects, weighted by the number of episodes (explained below). We used one regression equation with differing weights corresponding to each state and also interacted covariates in the regression to obtain this overall estimated treatment effect along with its standard error. Specifically, for each measure, the HHVBP state-level weights equal the number of post-intervention episodes in the denominator of the respective measure for the given state, divided by the total number of pre-intervention episodes in the denominator for all HHVBP states. Pre- and post-intervention weights were calculated to represent each intervention state proportionally, in each time period. The weights for the comparison groups for each HHVBP state were set equal to the weights of the respective HHVBP states. Additionally, we fully interacted the state indicator variables with all of the variables from the state-level models, except for the treatment effect. This weighted pooled regression provided the overall treatment effect, equal to the weighted average of the state estimates, along with the pooled standard error.

Model Specification for Claims-Based Impact Measures
The vector of covariates $\mathbf{P}_{Cov}$ that we included in the D-in-D equation for the claims-based measures included both patient and HHA characteristics. We controlled for differences in HHA characteristics by

including ownership type, freestanding status (see Section 2.1.3, "HHA Characteristics"), and size of provider (see Section 2.2, "Additional Variables Used in Comparison Group Construction and D-in-D") as covariates in the D-in-D specification for all six claims measures. In addition, we adjusted for selected patient characteristics, which were considered to be highly prevalent and related to the impact measures, but were not used in the matching process. The patient characteristics that are used in the D-in-D specification are beneficiary chronic condition indicators for: Alzheimer's disease and related disorders or senile dementia, depression, anxiety, psychotic and stress disorders, disabilities (intellectual, learning or developmental delays), cancer (breast, colorectal, prostrate, lung, endometrial or leukemia), liver disease, arthritis, chronic obstructive pulmonary disease, ischemic heart disease, atrial fibrillation, and anemia. As with the chronic conditions used in the episode matching process, these chronic condition indicators are derived from the Master Beneficiary Summary File (MBSF).

Model Specification for OASIS-Based Impact Measures

As part of our D-in-D model specification for the seven OASIS-based impact measures, we controlled for characteristics of HHAs. The vector of covariates P_{Cov} included ownership type, freestanding status, (see Section 2.1.3, "HHA Characteristics") and size of provider (see Section 2.2, "Additional Variables Used in Comparison Group Construction and D-in-D").

1.6 Comparison Group Approach C: HHA Reweighting for Measures Evaluated at the HHA Level

Our final comparison group approach is designed to support analyses of impact measures at the HHA level. This approach balances several key agency characteristics between HHAs in the HHVBP states and those included in the resulting comparison group. With this approach, we are able to control for changes in agency characteristics that are associated with HHA-level impact measures of interest. This approach is used to evaluate relative changes in quality of care among HHAs in the intervention and comparison groups while accounting for changes over time in the types of HHAs comprising each group (e.g., as a result of market entry/exit). That is, we ensure that the intervention and comparison groups remain comparable through the end of the first year of HHVBP (2016) based on several characteristics that may affect the impact measures of interest. To the extent that any such changes in the composition of HHAs in operation in the intervention states during 2016 are a direct result of the HHVBP Model (i.e., reflecting endogeneity), there is a risk that this approach may lead us to adjust away certain effects of HHVBP that are also of interest. Our analyses focuses on the early stage of the operation of the HHVBP Model, which reflects the experience of HHAs at least one year prior to the application of the initial HHVBP payment adjustments in 2018. We view our approach as potentially accounting for changes over time in HHA characteristics that were likely already underway prior to HHVBP and are more likely to represent a source of potential confounding for early estimates of the impact on quality of care at the agency level. In future Annual Reports that will also incorporate additional data, we plan to expand our analyses to separately examine both the effect of HHVBP on HHA entry/exit decisions and include an approach that allows any corresponding impact on quality of care to be captured in overall estimates of the impact of the Model.

The comparison group of HHAs was drawn from all 41 non-HHVBP states. The weighting was done such that the reweighted comparison population exactly matches the HHVBP HHA population under examination on the following agency characteristics: ownership type; setting (i.e., freestanding vs. hospital-based); agency age; chain status; agency size; and 2015 TPS quintiles (see Section 2.1.3, "HHA

Characteristics" and Section 2.7, "TPS"). We created strata based on the combinations of agency characteristics and derived weights such that the reweighted proportion of comparisons agencies matches the proportion of treatment agencies in each stratum. These calculated weights were subject to change across years or quarters, as the number of active HHAs within each stratum could change over time.

1.6.1 D-in-D for the OASIS Process Measures

We defined the baseline period to be 2015 and estimated cumulative treatment effects from a D-in-D model for each of the four quarters in 2016. To obtain quarterly cumulative effects in the post-intervention period, we restricted the estimation sample to include observations up to the quarter of interest (i.e., quarter 1, 2, 3, or 4 in 2016).

Defining each episode i in quarter q and year t, identifying the treatment episodes with an indicator variable $Treat_i$, identifying post-intervention time periods with an indicator variable $I(t \geq 2016)$, which takes a value of 1 for post-intervention time periods and 0 otherwise, the D-in-D estimator for outcome Y is implemented as:

$$Y_{i,q,t} = \beta_0 + \beta_1 * Treat_i + \beta_2 * I(t \geq 2016) + \beta_3 * Treat_i * I(t \geq 2016) + \sum_{j=1}^{j=3} \delta_j I(q = j) + \epsilon_{i,q,t}$$

where β_0 is an intercept, β_1 is the average difference between the treatment and comparison populations over the baseline period, β_2 is the average change from baseline to post-intervention for the comparison population, β_3 is the difference in the change from baseline to post-intervention for the treatment population relative to the comparison population (i.e., the estimated treatment effect of HHVBP), and δ_j coefficients capture seasonal effects in each quarter of the year. Unlike OASIS outcome measures, the OASIS process measures represent quality measures that were defined without beneficiary risk adjustment. Therefore, we also do not adjust for beneficiary characteristics in the regression model. Additionally, HHA reweighting balances several key agency characteristics between HHAs in the HHVBP states and those included in the resulting comparison group. As described above, we obtained four estimated treatment effects for the four quarterly cumulative effects in the post-intervention period. Standard errors are clustered at the agency-level.

1.6.2 D-in-D for the HHCAHPS Measures and TPS

Unlike previous impact measures, the relevant unit of analysis for both TPS and Home Health Consumer Assessment of Healthcare Providers and Systems (HHCAHPS) measures is the agency year. Defining each agency i in year t, identifying the treatment agencies with an indicator variable $Treat_i$, identifying post-intervention time periods with an indicator variable $I(t \geq 2016)$, which takes a value of 1 for post-intervention time periods and 0 otherwise, the D-in-D estimator for outcome Y is implemented as:

$$Y_{i,t} = \beta_0 + \beta_1 * Treat_i + \beta_2 * I(t \geq 2016) + \beta_3 * Treat_i * I(t \geq 2016) + \epsilon_{i,t}$$

where β_0 is an intercept, β_1 is the average difference between the treatment and comparison populations over the baseline period, β_2 is the average change from baseline to post-intervention for the comparison population, β_3 is the difference in the change from baseline to post-intervention for the treatment population relative to the comparison population (i.e., the estimated treatment effect of

HHVBP). We do not adjust for covariates in the regression model. Since models are estimated using annual (versus quarterly) data for each provider, this report includes only one D-in-D estimate (i.e., for 2016). Standard errors are clustered at the agency-level.

2. Descriptive Variable and Impact Measure Definitions

Below, we describe how we specified and defined the descriptive variables and impact measures that were used in this Annual Report.

2.1 Descriptive Variables

2.1.1 Beneficiary Characteristics

Total Number of Beneficiaries Receiving HH Care. This measure used HH Claims to identify and count the number of unique home health beneficiaries with at least one home health episode ending in a specified time period.

Age. This measure used HH Claims and MBSF data file to identify all beneficiaries with at least one home health episode ending in a specified period, and then used the *AGE_AT_END_REF_YR* variable to identify beneficiary ages. The mean age of home health beneficiaries was calculated such that each beneficiary had an equal weight, ignoring the number of episodes s/he may have received.

Gender. This measure used HH Claims and MBSF data file to identify all beneficiaries with at least one home health episode ending in a specified period, and then used the *SEX_INDENT_CD* variable to identify the proportion of home health beneficiaries of each gender. The proportion of home health beneficiaries was calculated such that each beneficiary had an equal weight, ignoring the number of episodes she may have received.

Race/Ethnicity. This measure used HH Claims and MBSF data file to identify all beneficiaries with at least one home health episode ending in a specified period, and then used the *BENE_RACE_CD* variable to calculate the proportion of each race category for home health beneficiaries. The proportion of home health beneficiaries was calculated such that each beneficiary had an equal weight, ignoring the number of episodes she may have received.

Dual Eligible. This measure used HH Claims and MBSF data file to identify all beneficiaries with at least one home health episode ending in a specified CY, and then used the *DUAL_ELGBL_MONS* variable to identify beneficiaries who qualify as dual eligible. If a beneficiary was dual eligible for any month within the specific time period, they were considered dual eligible for the purposes of this metric.

Rural/Urban. This measure used HH Claims data to identify all beneficiaries with at least one home health episode ending in a specified period. County Core-Based Statistical Area (CBSA) codes from the Area Health Resource File (AHRF) data were used to identify rural counties (i.e., those that lacked a CBSA code) and urban counties (i.e., those with a CBSA code). Rural/urban indicators were then matched to beneficiaries' county information from HH Claims files to calculate the proportion of beneficiaries from rural versus urban counties.

2.1.2 Episode Characteristics

Episodes by Type. This measure used HH Claims to generate a list of all home health episodes that ended within a specified time period, and MBSF data to exclude home health episodes with beneficiaries that were not continuously enrolled in Medicare FFS throughout the duration of the time

period. Episodes with outlier payment adjustments were identified using the variable claim value code = 17. Low Utilization Payment Adjustment (LUPA) payment adjustments were identified using the variable *CLM_HHA_LUPA_IND_CD*, and Partial Episode Payment (PEP) payment adjustments were identified using *PTNT_DSCHRG_STUS_CD = '06.'* Otherwise, all episodes without outlier, LUPA, or PEP adjustments were categorized as "Normal".

Episodes within a Sequence. This measure used HH Claims to generate a list of all home health episodes that ended within a specified time period, and MBSF data to exclude those home health episodes with beneficiaries that were not continuously enrolled in Medicare FFS throughout the duration of the time period. Episode sequences were put together by combining episodes whose end date and start dates were within 60 days of one another; a gap of at least 60 days indicated the start of a new sequence. Three different categories of episode were created depending on the order in which an episode appears relative to the larger sequence of episodes in which it occurs: first in a sequence, second in a sequence, and third or higher in a sequence. All episodes, regardless of episode type (i.e., normal, outlier, LUPA, PEP), were included. The logic we employed is in alignment with the methodology for determining HH stays for the two HHVBP claims-based measures.[7]

Visits in an Episode. This measure used home health revenue center codes and HH Claims to count the individual visits associated with each home health episode that ended within a specified time period. Using the revenue center codes associated with each home health claim, visits per episode were counted for each type of home health service: physical therapy visit (revenue center code 042x), occupational therapy visit (revenue center code 043x), speech language pathology visit (revenue center code 044x), skilled nursing visit (revenue center code 055x), medical social services visit (revenue center code 056x), and home health aide visit (revenue center code 057x). These visits were then summed to calculate the total visits per episode.

Visits in an Episode by Type of Visit. This measure used home health revenue center codes and HH Claims to count the individual visits associated with each home health episode that ended within a specified time period. Using the revenue center codes associated with each home health claim, visits per episode were counted for each type of home health service: physical therapy visit (revenue center code 042x), occupational therapy visit (revenue center code 043x), speech language pathology visit (revenue center code 044x), skilled nursing visit (revenue center code 055x), medical social services visit (revenue center code 056x), and home health aide visit (revenue center 057x).

2.1.3 HHA Characteristics

Total Number of HHAs. This measure used HH Claims to generate a list of all unique HHAs with at least one home health episode ending in a specified time period.

Ownership. This measure used HH Claims to generate a list of all unique HHAs with at least one home health episode ending in a specified time period. Using publicly available Centers for Medicare and Medicaid Services (CMS) Provider of Services (POS) data, HHAs were categorized as for-profit, non-profit, or government-owned, conditional on control type. From the provider data, control type is recoded as "non-profit" (control type codes 1,2,3), "for-profit" (control type code 4), and "government-owned" (control type codes 5,6,7). Percentages were calculated by dividing these category counts by the total number of HHAs (defined above).

Setting: Hospital-Based vs. Freestanding. This measure used HH Claims to generate a list of all unique HHAs with at least one home health episode ending in a specified time period. Using publicly available CMS POS data, HHAs were categorized as freestanding or hospital-based conditional on facility type. From the provider data, facility type is recoded as "freestanding" (facility type codes 1,2,3,7) or "hospital-based" (facility type codes 4,5,6). Percentages were calculated by dividing these category counts by the total number of HHAs (defined above).

HHA Age. This measure used HH Claims to generate a list of all unique HHAs with at least one home health episode ending in a specified time period, as well as HHA information from the POS file. Age was then calculated for each HHA by subtracting the specified year from the year the HHA opened, as derived from the *ORGNL_PRTCPTN_DT* variable.

Entry of New/Exit of Existing Providers. This measure was calculated using OASIS data. A list of unique HHAs with an OASIS assessment was generated at the year level. OASIS assessments were used rather than HH claims to more fully capture all HHAs (e.g., including HHAs that may not have had a FFS claim). Each year's HHA list was then compared to each other year's HHA list, to identify HHAs that were present in one year but not the other. These HHAs were counted to identify how many HHAs were new or no longer active between two given years.

Chain Membership. Freestanding HHAs were categorized as chain-affiliated explicitly using the respective Cost Report field. Hospital-based agencies were categorized as chain-affiliated if any of the Cost Report fields corresponding to a facility being part of a chain organization, such as home office name, location, and contractor number, were populated. For HHAs that had a missing Cost Report at some point during the 2012 –-2016 period, chain status was imputed using the Cost Report submitted in the last year prior to the missing observation. However, if a Cost Report was missing in 2012, the chain status from the 2013 Cost Report was used. Agencies with no Cost Reports between 2012 and 2016 were assumed to be unaffiliated with a chain.

HHAs per 10,000 Medicare Beneficiaries. This measure used HH claims to count the total number of unique HHAs that had a FFS episode ending in a specified time period, and the MBSF to generate a count of total FFS beneficiaries. The count of HHAs was then divided by the total number of Medicare-eligible FFS beneficiaries alive at the beginning of the time period. That ratio was further scaled by multiplying by 10,000.

2.2 Additional Variables Used in Comparison Group Construction and D-in-D Analyses

Prior Care Setting. SNF, Inpatient, and Outpatient claims data were utilized for determining the patient's care setting prior to each home health episode sequence. Prior care setting was used to construct comparison groups for the claims-based impact measures, as it is included in the measure developers' risk adjustment process for the two HHVBP claims-based measures.[8] For this purpose, all final action claims with a claim through date from Quarter 3 of 2011 through Quarter 4 of 2016 were extracted from each care setting (inpatient, SNF, and outpatient), restricted to claims for beneficiaries with a sequence of HH episodes starting in Quarter 3 of 2011 or later. Inpatient and SNF claims are combined into one inpatient stay in the case where a patient has multiple claims for different facilities of the same type, in which the initial claim ends on the same day or one day prior to when the ensuing claim starts. When these "transfer" claims are combined, relevant information such as diagnosis, revenue and procedure

---

[8] See CMS Specifications for Home Health Claims-Based Utilization Measures, "Prior Care Setting" in Section 1.5.

codes, CMS Certification Number (CCN), etc., are pulled from the claim with the later "claim through" date. Inpatient claims are subdivided into categories (as specified in the claims-based measure specifications) based on the following criteria:

- Inpatient acute
    - Positions 3-6 of CCN = 0001-0879 or 0800-0899 or 1300-1399[1]
        - Further subcategories defined using Agency for Healthcare Research and Quality (AHRQ) International Classification of Diseases (ICD)9/10 procedure Clinical Classifications Software (CCS)
            - Surgery/Gynecology
            - Cardiorespiratory
            - Cardiovascular
            - Neurology
            - Medicine
- Inpatient rehabilitation facility (IRF)
    - (Positions 3-6 of CCN = 3025-3099) or (Position 3 = R or T)
    - NCH_CLM_TYPE_CD = 60 or 61
- Inpatient psychiatric facility (IPF)
    - (Positions 3-6 of CCN = 4000-4499) or (Position 3 = M or S)
- Long-term care hospital (LTCH)
    - Positions 3-6 of CCN = 2000-2299

Outpatient claims are restricted to those indicating ED admissions, based on the same criteria as used to generate the outpatient AF (revenue center codes 0450-0459 or 0981). The final prior care setting file used for comparison group creation includes episode sequence-level data with indicator variables for the presence of inpatient hospital, inpatient SNF, and/or outpatient ED claims within 30 days prior to the episode sequence start date. The final prior care setting for each sequence is determined according to a hierarchy of events occurring within the prior care window, should multiple events occur. The order of hierarchy is SNF stay, inpatient admission (including IRF, IPF, and LTCH), and finally ED event, such that if a patient experienced an SNF stay followed by a hospitalization (i.e., an inpatient admission), and then began home health care all within 30 days, their final prior care setting would reflect their SNF stay; this hierarchy is consistent with the risk adjustment methodology of the claims-based HHVBP measures.[8] Within this hierarchy, inpatient admissions are further classified by procedure code-based flags and diagnosis-based flags, in which procedure code-based flags take higher priority.

Using these sequence-level indicator variables, the percentage of all home health episodes that include a prior care setting was calculated for each prior care category: ED use without admission, inpatient hospitalization, and SNFs. The count of each indicator type was calculated and divided by the total number of home health episodes for a given geographic area and time period of interest.

Health Conditions. The following chronic health conditions were included in the episode-level matching criteria based on a given beneficiary's chronic conditions data for the year in which the episode began: chronic kidney disease, congestive heart failure, diabetes, and ulcers.

Using the chronic condition indicators listed, the percentage of all home health episodes that include a chronic condition was calculated for each condition category: Alzheimer's disease and related disorders or senile dementia, depression, anxiety, psychotic and stress disorders, disabilities (intellectual, learning

or developmental delays), cancer (breast, colorectal, prostrate, lung, endometrial, or leukemia), liver disease, arthritis, chronic obstructive pulmonary disease, ischemic heart disease, atrial fibrillation, and anemia. The count of each indicator type was calculated and divided by the total number of home health episodes for a given geographic and time period of interest.

HHA Size. In order to estimate the operating size of a particular HHA in 2015, we categorized HHAs based on their total count of beneficiaries with at least one OASIS-based quality episode ending in 2015, as provided by the HHVBP Implementation contractor (and explained in Section 3.10). HHAs were divided into the following groups based on count of beneficiaries: those with 1 – 59 beneficiaries, those with 60 to the national median number of beneficiaries, and those with greater than the national median number of beneficiaries.

2.3   Claims-Based Impact Measures

This section presents how the claims-based measures were created. Of note, the bottom row of each table notes the data source(s) used to create the impact measure. Each of the impact measures, episode-level numerator and denominator indicators were merged with other variables to create the AFs (Sections 4.1 and 4.2) that are used to conduct the analyses and produce the results presented in the main report.

Table 3. Average Number of FFS HH Episodes per 1,000 FFS Beneficiaries

Measure Concept	Definition
Measure Category	Utilization
Measure Description	Number of HH episodes ending in a given quarter per 1,000 FFS beneficiaries alive at the beginning of the quarter.
Measure Numerator	Total number of HH episodes ending in the quarter.
Numerator Details	Numerator includes all HH episodes of all types (LUPAs, outliers, PEPs, etc.) irrespective of whether they are first, second, or higher in the sequence, ending in a given quarter.
Measure Denominator	Total number of Medicare-eligible FFS beneficiaries alive at the beginning of the given quarter divided by 1000.
Denominator Details	Total number of Medicare-eligible FFS beneficiaries alive at the beginning of the quarter (e.g., if a beneficiary is no longer FFS next month, they are still included) is obtained and then the number is divided by 1000.
Data Sources	MBSF, HHA Claims

Table 4. Average Medicare HH Spending per FFS Beneficiary

Measure Concept	Definition
Measure Category	Spending
Measure Description	Average Medicare payment amount for episodes ending in a given quarter per FFS beneficiary alive at the beginning of the quarter.
Measure Numerator	Numerator includes total HH spending in state.
Numerator Details	Sum of claim payment amount (total episode payment), including all payments (e.g., outlier payments) for episodes ending in the quarter.
Measure Denominator	Total number of Medicare-eligible FFS beneficiaries alive at the beginning of the given quarter.
Denominator Details	Total number of Medicare-eligible FFS beneficiaries alive at the beginning of the quarter (e.g., if a beneficiary is no longer FFS next month, they are still included).
Data Sources	MBSF, HHA Claims

Table 5. Average Medicare HH Spending per FFS HH Episode

Measure Concept	Definition
Measure Category	Spending
Measure Description	Average Medicare payment amount per HH episode among episodes ending within a given quarter.
Measure Numerator	Numerator includes total HH spending in state.
Numerator Details	Sum of claim payment amount (total episode payment including all payments [e.g., outlier payments]) for episodes ending in quarter.
Measure Denominator	Total number of HH episodes ending in a given quarter.
Denominator Details	Denominator includes all HH episodes of all types (LUPAs, outliers, PEPs, etc.) irrespective of whether they are first, second, or higher in the sequence, ending in a given quarter. Denominator excludes HH episodes for patients who are not continuously enrolled in FFS Medicare during the numerator window (total episode window) or until death.
Data Sources	MBSF, HHA Claims

Table 6. ED Use (no Hospitalization) per First FFS HH Episodes (%)

This impact measure is similar to the HHVBP measure, "Emergency Department Use without Hospitalization". [9] However, unlike the HHVBP measure, it is not risk-adjusted.

Measure Concept	Definition
Measure Category	Utilization
Measure Description	Percentage of first HH episodes in a sequence with at least one outpatient ED visit without ACH within 60 days of the start of the episode.
Measure Numerator	Number of HH episodes for patients who have at least one Medicare claim for outpatient ED use and no claims for ACH in the 60 days following the start of the HH episode.
Numerator Details	▪ The 60-day time window is calculated by adding 59 days to the "from" date of the first HH episode (in the sequence). ▪ If the patient has any Medicare outpatient claims with any emergency room visit revenue center codes (0450-0459, 0981) during the 60-day window *and* if the patient has no Medicare inpatient claims for admission to an acute care hospital as identified by the CCN on the inpatient claim during the 60-day window, then the stay is included in the measure numerator.
Measure Denominator	Total number of eligible HH episodes, starting in a given quarter that are first in the sequence of episodes.
Denominator Details	All HH episodes that start in the quarter are included with the following exclusions: ▪ HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode. ▪ HH episodes for patients who were not FFS eligible in the six months prior. ▪ HH episodes that are NOT the first in the sequence of HH stay. ▪ HH episodes that begin with a LUPA claim. ▪ HH episodes in which the patient receives service from multiple agencies during the 60-day window (see "Transfer HHAs within 60 days" in Glossary [Section 6]).
Data Sources	MBSF, HHA Claims, Inpatient Claims, Outpatient Claims. Enrollment status and beneficiary death date are obtained from MBSF.

---

[9] See CMS Specifications for Home Health Claims-Based Utilization Measures, "Emergency Department Use without Hospitalization".

Table 7. Unplanned Acute Care Hospitalizations per First FFS HH Episodes (%)

This impact measure is similar to the HHVBP measure, "Acute Care Hospitalization". [10] However, unlike the HHVBP measure, it is not risk-adjusted.

Measure Concept	Definition
Measure Category	Utilization
Measure Description	Percentage of first HH episodes in a sequence with at least one unplanned admission to an acute care hospital within 60 days of the start of the episode.
Measure Numerator	Number of HH episodes for patients who have at least one Medicare claim for an unplanned admission to an acute care hospital in the 60 days following the start of the episode.
Numerator Details	▪ The 60-day time window is calculated by adding 59 days to the "from" date of the HH episode. ▪ ACH occurs (and the HH episode is included in the numerator) if the patient has at least one Medicare inpatient claim from short-term or critical access hospitals (identified by CCN ending in 0001-0879, 0800-0899, or 1300-1399) during the 60-day window. ▪ Inpatient claims for planned hospitalizations are excluded from the ACH measure numerator. The HHVBP ACH measure specifications provide a list of AHRQ Procedure and Condition CCS and additional ICD-9-CM procedure codes that indicate planned hospitalizations. The measure specifications, including the AHRQ codes, were pulled from the CMS Home Health Claims-Based Utilization Measures Specifications.[10]
Measure Denominator	Total number of eligible HH episodes starting in a given quarter that are first in the sequence of episodes.
Denominator Details	All HH episodes that start in the quarter are included with the following exclusions: ▪ HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode. ▪ HH episodes for patients who were not FFS eligible in the six months prior. ▪ HH episodes that are NOT the first in the sequence of HH stay. ▪ HH episodes that begin with a LUPA claim. ▪ HH episodes in which the patient receives service from multiple agencies during the 60-day window (see "Transfer HHAs within 60 days" in Glossary [Section 6]).
Data Sources	MBSF, HHA Claims, Inpatient Claims. Enrollment status and beneficiary death date are obtained from MBSF.

---

[10] See CMS Specifications for Home Health Claims-Based Utilization Measures, "Acute Care Hospitalization."

Table 8. Unplanned Acute Care Hospitalizations per All FFS HH Episodes (%)

Measure Concept	Definition
Measure Category	Utilization
Measure Description	Percentage of HH episodes with at least one unplanned admission to an acute care hospital within 60 days of the start of the episode or until the start of the next HH episode that begins on or before the 60th day.
Measure Numerator	Number of HH episodes with at least one unplanned admission to an acute care hospital within 60 days of the start of the episode or until the start of the next HH episode that begins on or before the 60th day.
Numerator Details	▪ The 60-day time window is calculated by adding 59 days to the "from" date of the HH episode. If a subsequent HH episode starts on or before the 60th day, the time window is ended early on the day prior to the start of the next episode. ▪ ACH occurs (and the HH episode is included in the numerator) if the patient has at least one Medicare inpatient claim from short-stay or critical access hospitals during the 60-day window. ▪ Planned hospitalizations (defined by a list of AHRQ Procedure and Condition CCS and additional ICD-9-CM procedure codes) are excluded from the measure numerator. The measure specifications, including the AHRQ codes, were pulled from the CMS Home Health Claims-Based Utilization Measures Specifications.[7]
Measure Denominator	Total number of eligible HH episodes starting in a given quarter.
Denominator Details	All HH episodes that start in the quarter are included with the following exclusions: ▪ HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode. ▪ HH episodes for patients who were not FFS eligible in the six months prior. ▪ HH episodes that begin with a LUPA claim. ▪ HH episodes in which the patient receives service from multiple agencies during the 60-day window (see "Transfer HHAs within 60 days" in Glossary [Section 6]).
Data Sources	MBSF, HHA Claims, Inpatient Claims. Enrollment status and beneficiary death date are obtained from MBSF.

Table 9. Medicare Spending for Unplanned Acute Care Hospitalizations per All FFS HH Episodes (%)

Measure Concept	Definition
Measure Category	Spending
Measure Description	Medicare payments all unplanned ACHs within 60 days of the start of the HH episode or until the start of the next HH episode that begins on or before the 60th day.
Measure Numerator	Total Medicare payments associated with all unplanned ACHs within 60 days of the start of the HH episode and prior to the start of the next HH episode that begins on or before the 60th day (if applicable).
Numerator Details	▪ Total Medicare payments associated with the entire unplanned ACH are included if the ACH occurs within 60 days of the start of the HH episode and prior to the start of the next HH episode that begins on or before the 60th day (if applicable). ▪ Inpatient claims that completely overlap (i.e., claim through date of first claim is greater than claim from date of ensuing claim) are combined in terms of expense and duration. Further, in the case of two consecutive acute care hospital claims for which the later claim begins on the same or next day of the prior claim's end date, following logic is applied: if the provider on each claim is different, then combine the claims into one hospital stay in which the patient transferred hospitals; if the provider on each claim is the same, then maintain the two separate stays and consider the second claim to be a re-admission due to a potentially different diagnosis. ▪ If there are multiple distinct eligible claims associated with a single HH episode then the costs associated with all of them are included. ▪ Planned hospitalizations (defined by a list of AHRQ Procedure and Condition CCS and additional ICD-9-CM procedure codes) are excluded from the measure numerator. ▪ The measure specifications, including the AHRQ codes, were pulled from the CMS Home Health Claims-Based Utilization Measures Specifications.[7]
Measure Denominator	Total number of eligible HH episodes starting in a given quarter.
Denominator Details	All HH episodes that start in the quarter are included with the following exclusions: ▪ HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode. ▪ HH episodes for patients who were not FFS eligible in the six months prior. ▪ HH episodes that begin with a LUPA claim. ▪ HH episodes in which the patient receives service from multiple agencies during the 60-day window (see "Transfer HHAs within 60 days" in Glossary [Section 6]).
Data Sources	MBSF, HHA Claims, Inpatient Claims. Enrollment status and beneficiary death date are obtained from MBSF.

Table 10. SNF Use per All FFS HH Episodes (%)

Measure Concept	Definition
Measure Category	Utilization
Measure Description	Percentage of HH episodes with at least one admission to a SNF within 60 days of the start of the HH episode or until the start of the next HH episode that begins on or before the 60th day.
Measure Numerator	Number of HH episodes with at least one admission to a SNF within 60 days of the start of the HH episode or until the start of the next HH episode that begins on or before the 60th day.

Numerator Details	▪ The 60-day time window is calculated by adding 59 days to the "from" date of the HH episode. If a subsequent HH episode starts on or before the 60th day, the time window is ended early on the day prior to the start of the next episode.
	▪ The SNF admission is counted if the patient has at least one SNF claim during the 60-day window.
	▪ SNF admissions following planned ACH (defined by a list of AHRQ Procedure and Condition CCS and additional ICD-9-CM procedure codes) are excluded from the measure numerator under the following conditions:
	  1. The planned hospitalization starts within the HH episode 60-day window.
	  2. The planned hospitalization ends within the HH episode 60-day window.
	  3. The SNF stay starts within the HH episode 60-day window.
	  4. The SNF stay starts on or after the planned hospitalization end date.

Measure Denominator	Total number of eligible HH episodes starting in a given quarter.
Denominator Details	All HH episodes that start in the quarter are included with the following exclusions: ▪ HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode. ▪ HH episodes for patients who were not FFS eligible in the six months prior. HH episodes that begin with a LUPA claim. HH episodes in which the patient receives service from multiple agencies during the 60-day window (see "Transfer HHAs within 60 days" in Glossary [Section 6]).
Data Sources	HHA Claims, SNF Claims, MBSF, Inpatient RIF. Enrollment status is identified using the Medicare Enrollment Database (EDB).

Table 11. Medicare Spending for SNF Stays per All FFS HH Episodes

Measure Concept	Definition
Measure Category	Spending
Measure Description	Medicare payments for all SNF stays within 60 days of the start of the HH episode or until the start of the next HH episode that begins on or before the 60th day.
Measure Numerator	Total Medicare payments associated with all SNF admissions that occur within 60 days of the start of the HH episode and prior to the start of the next HH episode that begins on or before the 60th day (if applicable).
Numerator Details	- Total Medicare payments associated with the entire SNF stay are included if the SNF admission occurs within 60 days of the start of the HH episode and prior to the start of the next HH episode that begins on or before the 60th day (if applicable). - SNF claims that completely overlap (i.e., claim through date of first claim is greater than claim from date of ensuing claim) are combined in terms of expense and duration. Further, in the case of two consecutive SNF claims for which the later claim begins on the same or next day of the prior claim's end date, following logic is applied: if the provider on each claim is different, then combine the claims into one SNF stay in which the patient transferred facilities; if the provider on each claim is the same, then maintain the two separate admissions and consider the second claim to be a re-admission due to a potentially different diagnosis. - If there are multiple distinct eligible SNF claims associated with a single HH episode, then the costs associated with all of them are included. - SNF admissions following planned ACHs (defined by a list of AHRQ Procedure and Condition CCS and additional ICD-9-CM procedure codes) are excluded from the measure numerator under the following conditions: - The planned hospitalization starts between the HH episode start date and 60 days past discharge. - The planned hospitalization ends between the HH episode start date and 60 days past discharge. - The SNF stay starts between the HH episode start date and 60 days past discharge. - The SNF stay starts on or after the planned hospitalization end date.
Measure Denominator	Total number of eligible HH episodes starting in a given quarter.
Denominator Details	All HH episodes that start in the quarter are included with the following exclusions: - HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode. - HH episodes for patients who were not FFS eligible in the six months prior. - HH episodes that begin with a LUPA claim. - HH episodes in which the patient receives service from multiple agencies during the 60-day window (see "Transfer HHAs within 60 days" in Glossary [Section 6]).
Data Sources	HHA Claims, SNF Claims, MBSF, Inpatient RIF. Enrollment status is identified using the Medicare EDB.

2.4   OASIS-Based Outcome Impact Measures

This section presents information on the seven HHVBP OASIS-based outcome impact measures. Of note, the measure values were included as part of our Quality Improvement and Evaluation System (QIES) extract (see Section 3.6.2); no additional measure calculations were necessary. The tables below summarize the measure definitions provided by CMS.[11] These impact measures were part of the AFs (Section 4.3) that are used to conduct the analyses and produce the results presented in the main report.

Table 12. Improvement in Bathing

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH episodes of care during which the patient got better at bathing self.
Measure Numerator	Number of HH episodes of care where the value recorded on the discharge assessment indicates less impairment in bathing at discharge than at (SOC/ROC).
Measure Denominator	Number of HH episodes of care ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH episodes of care for which the patient, at SOC/ROC, was able to bath self independently, episodes that end with inpatient facility transfer or death, or patient is nonresponsive.

Source: *CMS OASIS-C2 Home Health Outcome Measures*

Table 13. Improvement in Bed Transferring

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH quality episodes during which the patient improved in ability to get in and out of bed.
Measure Numerator	Number of HH quality episodes where the value recorded on the discharge assessment indicates less impairment in bed transferring at discharge than at SOC/ROC.
Measure Denominator	Number of HH quality episodes ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which the patient, at SOC/ROC, was able to transfer independently, episodes that end with inpatient facility transfer or death, or patient is nonresponsive.

Source: *CMS OASIS-C2 Home Health Outcome Measures*

---

[11] See *CMS OASIS-C2 Home Health Outcome Measures*.

Table 14. Improvement in Ambulation-Locomotion

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH quality episodes during which the patient improved in ability to ambulate.
Measure Numerator	Number of HH quality episodes where the value recorded on the discharge assessment indicates less impairment in ambulation/locomotion at discharge than at SOC/ROC.
Measure Denominator	Number of HH quality episodes ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which the patient, at SOC/ROC, was able to ambulate independently, episodes that end with inpatient facility transfer or death, or patient is nonresponsive.

Source: CMS OASIS-C2 Home Health Outcome Measures

Table 15. Improvement in Dyspnea

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH quality episodes during which the patient became less short of breath or dyspneic.
Measure Numerator	Number of HH quality episodes where the discharge assessment indicates less dyspnea at discharge than at SOC/ROC.
Measure Denominator	Number of HH quality episodes ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which the patient, at SOC/ROC, was not short of breath at any time, or episodes that end with inpatient facility transfer or death.

Source: CMS OASIS-C2 Home Health Outcome Measures

Table 16. Improvement in Management of Oral Medications

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH quality episodes during which the patient improved in ability to take their medicines correctly (by mouth).
Measure Numerator	Number of HH quality episodes where the value recorded on the discharge assessment indicates less impairment in taking oral medications correctly at discharge than at SOC/ROC.
Measure Denominator	Number of HH quality episodes ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which the patient, at SOC/ROC, was able to take oral medications correctly without assistance or supervision, episodes that end with inpatient facility transfer or death, patient is nonresponsive, or patient has no oral medications prescribed.

Source: CMS OASIS-C2 Home Health Outcome Measures

Table 17. Improvement in Pain Interfering with Activity

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH quality episodes during which the patient's frequency of pain when moving around improved.
Measure Numerator	Number of HH quality episodes where the value recorded on the discharge assessment indicates less frequent pain at discharge than at SOC/ROC.
Measure Denominator	Number of HH quality episodes ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which the patient, at SOC/ROC, had no pain reported, episodes that end with inpatient facility transfer or death, or patient is nonresponsive.

Source: CMS OASIS-C2 Home Health Outcome Measures

Table 18. Discharged to Community

Measure Concept	Definition
Measure Category	OASIS Outcome
Measure Description	Percentage of HH episodes after which patients remained at home.
Measure Numerator	Number of HH episodes where the assessment completed at the discharge indicates the patient remained in the community after discharge.
Measure Denominator	Number of HH quality episodes ending with a discharge or transfer to inpatient facility during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes that end in patient death.

Source: CMS OASIS-C2 Home Health Outcome Measures

2.5   OASIS-Based Process Impact Measures

This section presents information on the three HHVBP OASIS-based process impact measures. Of note, the measure values were included as part of our QIES extract (see Section 3.6.2); no additional measure calculations were necessary. The tables below summarize the measure definitions provided by CMS.[12] These impact measures were part of the AFs (Section 4.4) are used to conduct the analyses and produce the results presented in the main report.

---

[12] See CMS OASIS-C2 Home Health Process Measures.

Table 19. Drug Education on Medications Provided to Patient/Caregiver during Episodes of Care

Measure Concept	Definition
Measure Category	OASIS Process
Measure Description	Percentage of HH quality episodes during which patient/caregiver was instructed on how to monitor the effectiveness of drug therapy, how to recognize potential adverse effects, and how and when to report problems (at the time of or at any time since the most recent SOC/ROC assessment).
Measure Numerator	Number of HH quality episodes during which patient/caregiver was instructed on how to monitor the effectiveness of drug therapy, how to recognize potential adverse effects, and how and when to report problems (at the time of or at any time since the most recent SOC/ROC assessment).
Measure Denominator	Number of HH quality episodes ending with a discharge or transfer to inpatient facility during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which the patient was not taking any drugs since the last OASIS assessment prior to transfer/discharge or the patient died.

Source: *CMS OASIS-C2 Home Health Process Measures*.

Table 20. Influenza Immunization Received for Current Flu Season

Measure Concept	Definition
Measure Category	OASIS Process
Measure Description	Percentage of HH quality episodes during which patients received influenza immunization for the current flu season.
Measure Numerator	Number of HH quality episodes during which the patient: a) received vaccination from the HHA; b) had received vaccination from HHA during earlier episode of care; or c) was determined to have received vaccination from another provider.
Measure Denominator	Number of HH quality episodes ending with a discharge or transfer to inpatient facility during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes for which no care was provided during October 1–March 31, the patient died, or the patient did not meet age/condition guidelines for influenza vaccine.

Source: *CMS OASIS-C2 Home Health Process Measures*.

Table 21. Pneumococcal Polysaccharide Vaccine Ever Received

Measure Concept	Definition
Measure Category	OASIS Process
Measure Description	Percentage of HH quality episodes during which patients were determined to have ever received Pneumococcal Polysaccharide Vaccine.
Measure Numerator	Number of HH quality episodes during which patients were determined to have ever received Pneumococcal Polysaccharide Vaccine.
Measure Denominator	Number of HH quality episodes ending with a discharge or transfer to inpatient facility during the reporting period, other than those covered by generic or measure-specific exclusions.
Measure-Specific Exclusions	HH quality episodes during which patient died, or patient did not meet age/condition guidelines for Pneumococcal Polysaccharide Vaccine.

Source: CMS OASIS-C2 Home Health Process Measures.

2.6   HHCAHPS-Based Impact Measures

For the five HHVBP performance measures that address beneficiary experience, we used the publicly available, HHA-level HHCAHPS data for CYs 2012 – 2016. To receive the annual Home Health Prospective Payment System payment update, HHAs that do not qualify for an exemption from participating in the HHCAHPS Survey must contract with an approved HHCAHPS Survey vendor, administer the survey on an ongoing (monthly) basis, and submit HHCAHPS Survey data to the HHCAHPS Data Center on a quarterly basis. Agencies are exempted if they serve 59 or fewer survey-eligible patients a year. Survey-eligible patients are those who are at least 18 years old and have their skilled care covered by Medicare or Medicaid.[13]

The five measures—constructed from 19 HHCAHPS questions—are summarized below.[14]

1. **Care of patients** reflects "patients who reported that their home health team gave care in a professional way."[14] This composite measure is comprised of four HHCAHPS questions that address how frequently the HHA treated the patient gently, with courtesy and respect, how frequently the HHA seemed informed and up-to-date, and if the patient had any problems with the care received.

2. **Communication between providers and patients** reflects "patients who reported that their home health team communicated well with them."[14] This composite measure is comprised of six HHCAHPS questions related to different aspects of communication, including how frequently the HHA explained things in an easy to understand manner, listened carefully, and kept the patient informed about when staff would arrive.

3. **Specific care issues** reflect "patients who reported that their home health team discussed medicines, pain and home safety with them."[14] This composite measure is comprised of seven HHCAHPS questions related to these three areas of care (that is, medicines, pain, and home safety).

4. **Rating of care provided by the agency** is a global rating measure that reflects the percentage of respondents who gave a rating of 9 or 10 to the question, "Using any number from 0–10, where 0 is the worst home health care possible, and 10 is the best home health care possible, what number would you use to rate your care from this agency's home health providers?"[14]

5. **Willingness to recommend the agency to friends and family** is a global rating measure that reflects the percentage of respondents who answered "Definitely Yes" to the question, "Would you recommend this agency to your family and friends if they needed home health care?"[14]

---

[13] Additional criteria are available here.
[14] Additional information on measure construction is available here.

2.7   TPS

Guided by parameters established by CMS for CY 2016,[15] the TPS was calculated as an aggregate performance metric based on 17 HHVBP measures, including: seven OASIS-based outcomes, three OASIS-based processes, two claims-based measures, and the five HHCAHPS measures.[16]

HHA measure scores were created as 12-month weighted averages, weighted by the episode counts, and rolled up from the agency-month to the agency-year level. An HHA's Performance Year measure rates (ranging from 0 – 100 points) are compared to its baseline year measure scores, as well as state-level performance standards: the achievement thresholds (ATs) and benchmarks (BMs). In the HHVBP Model, the baseline year is defined as CY 2015. For our computation of the TPS for years prior to the implementation of HHVBP, we defined the baseline year as the year prior to the designated Performance Year (e.g., for our calculation of TPS values for CY 2014, we used CY 2013 as the baseline year). ATs and BMs for each measure are calculated based on the distribution of baseline year measure rates for all eligible participating HHAs within a given state (see below for eligibility criteria). For each eligible measure, HHAs receive the higher of either an Achievement Score or an Improvement Score, between 0 and 10 points. Achievement/Improvement Scores are summed across all eligible measures to form an unadjusted performance measure score. For each HHA, this score was then weighted based on the number of eligible measures reported.

HHA eligibility criteria for the calculation of the ATs and BMs are as follows:

- For OASIS and claim-based measures, an HHA must have at least 20 episodes of care in the baseline year.
    - If an HHA did not have 20 or more episodes of care for a particular measure, the reported measure rates were recoded as missing.
- For HHCAHPS-based measures, an HHA must have at least 40 completed patient surveys during the baseline year for the five measure rates.
    - If an HHA did not have 40 or more completed patient surveys, the five HHCAHPS measure rates were recoded as missing.
- An HHA must have non-missing data for at least five of 17 eligible measures.
    - An "eligible measure count" (0 – 17) was created to tally the number of non-missing measures for each HHA to determine their inclusion/exclusion from AT/BM calculations.

HHA eligibility criteria for the calculation of TPS are as follows:

- HHAs must have data from the full 12 months of baseline year.
    - HHAs were flagged based on their participation date, extracted from the POS file. HHAs with a participation year greater than or equal to the baseline year were excluded from the data set of eligible agencies.
- HHAs must be in operation as of the release of the latest available POS file.

---

[15] See 2015 HHVBP Final Rule.

[16] Scores for the three new measures were not factored into our calculation of the TPS since these data are unavailable for non-HHVBP HHAs. Thus, the adjusted composite score for the 17 performance-based measures will be given full weight, as compared to the 90% weight that has been stipulated by CMS.

- HHAs were flagged based on their termination status, extracted from the POS file; HHAs that were flagged as terminated as of the POS data extraction were excluded from the data set of eligible agencies.
- HHAs must have at least five eligible measures for both the baseline year and the performance year.
  - For measure-level eligibility, see prior list above ("HHA eligibility criteria for calculations of AT/BMs").
- Of an HHA's eligible measures, at least five measures must be the same for both the baseline year and the performance year.
  - HHAs that did not have at least five shared measures between the baseline and performance years were excluded from the data set of eligible agencies.

Using the above methodology, we calculated the TPS for 2013 – 2016. Our 2016 TPS calculations for the HHVBP HHAs were consistent with those reported in the Preliminary Annual TPS and Payment Adjustment Report (released August 2017).

2.8   New Performance Measures

Three new measures were reported by HHAs through the CMS Secure Portal:

1. The proportion of eligible HHA personnel vaccinated for influenza within the year.
2. The proportion of patients aged 60 or older who either received or reported having a herpes zoster vaccination.
3. The proportion of patients who have an advance care plan or surrogate decision-maker documented in their medical record, or who had a documented discussion about advance care plans or surrogate but refused or were unable to provide the information.

A binary variable indicating the reporting status of each variable (i.e., reported/not reported) at the HHA-level was available from the CMS Secure Portal. We calculated the proportion of HHAs that reported each new measure. The numerator included all HHAs that reported the measure of interest, and the denominator included all HHAs operating under the HHVBP Model.

3.   Data Sources

For this Annual Report, we accessed CMS administrative data from several sources, including the Chronic Conditions Data Warehouse (CCW) via the Virtual Research Data Center, publicly available data sources, and other CMS HHVBP contractors. We also received varying analytic levels of measure-specific Outcome and Assessment Information Set (OASIS) data, extracted from CMS' QIES. We used these data sources to create the AFs necessary to conduct the analyses included in this Annual Report. Claims-based impact measures are calculated and analyzed using several data sources, including:

- Common Medicare Environment (CME) enrollment data
- HHA claims
- SNF claims
- Inpatient Hospitalization claims
- Outpatient ED claims
- POS files
- Area Health Resource File (AHRF)

The data sources discussed below are combined to create impact measures and descriptive variables, inform and construct comparison groups, and contribute to the AFs that are used to conduct the analyses and produce the results presented in the main report. Below, we describe the process for obtaining data from these sources in more detail.

3.1   Home Health Agency Claims

Purpose. HHA claims served as the foundation for determining home health episodes of care and defining the beneficiary population receiving home health care in a given quarter.

Data Acquisition. HHA claims data were pulled from the CCW's Research Identifiable Files (RIFs) in July 2017, which included all final action claims with claim type code 10 and a service end date (claim "through" date) ranging from January 1999 through June 2017. Although the measurement period for this evaluation begins in January 2012, prior years of HHA claims data are needed to establish accurate episode sequence information.

Data Processing. In order to establish the complete set of home health episodes of care, all final-action HHA claims that met the following conditions were included:

- Claim frequency code not equal to each of the following: missing; '0' (non-payment/zero claim); or '2' (Request for Anticipated Payment)
- Includes at least one covered visit
- Received a Medicare payment amount greater than $0

Furthermore, if a beneficiary had multiple claims with the same "Statement Covers From" date (i.e., "claim from" date), only the claim with the latest Fiscal Intermediary claim process date is included. In the event that multiple claims for the same beneficiary overlap in statement period "from" and "through" dates, the "Statement Covers Through" date (i.e., "claim through" date) on the claim starting earlier is adjusted to be the date before the ensuing claim from date. These data steps ensure that a given beneficiary could not be attributed to multiple HHAs on a given day when calculating episode-based impact measures (discussed below).

Each of the resulting HHA claims is considered a final home health episode with episode start date corresponding to the "claim from" date, and episode end date corresponding to the "claim through" date. In concordance with the measure specifications for the two HHVBP claims-based measures,[7] a sequence of episodes (or "home health stay") is defined as a series of consecutive home health episodes for a given beneficiary, in which the maximum time between consecutive episodes, end date to start date, is 60 days or less. If the time between the prior episode end date and ensuing episode start is greater than 60 days, the ensuing episode start date begins a separate home health stay.

An important by-product of HHA claims processing is a beneficiary finder file that includes a unique list of all beneficiaries with a home health episode ending on or after January 1, 2012, which includes the full measurement period associated with this report (2012 – 2016). For the remainder of this report, we will refer to this data set as the "HH Beneficiary Finder File."

3.2   Master Beneficiary Summary File

Purpose. Master Beneficiary Summary File (MBSF) data are the source for determining: beneficiary eligibility in impact measures based on FFS enrollment status, beneficiary demographics used in risk

adjustment of impact measures, and chronic condition status for comparison group construction and D-in-D model adjustment related to the claims-based measures.

Data Acquisition. MBSF data, sourced from the CME, are included in the CCW as annual snapshots that are divided into multiple segments: Base (Parts A/B/C/D), Chronic Conditions, Other Chronic or Potentially Disabling Conditions, Cost and Use, and National Death Index. For this annual report, we utilized the Base, Chronic Conditions, and Other Chronic Conditions segments.

The MBSF Base segment data provide monthly indicators of enrollment status, in addition to beneficiary demographic information (e.g., state and county of residence, date of birth, gender, race, etc.), for all Medicare enrollees. For this report, beneficiary year-level MBSF Base data were compiled from 2011 – 2016. For beneficiary gender, race, date of birth and date of death, only information from the most recent year of available MBSF for a given beneficiary are included in analyses.

As of the time of this report, the MBSF Chronic Conditions and Other Chronic or Potentially Disabling Conditions segments contain beneficiary-level 60 condition flags that are "developed from algorithms that search the CMS administrative claims data for specific diagnosis codes, MS-DRG codes, or procedure codes."[17] The condition flags are provided at both the middle and end of each year for each Medicare beneficiary; for this report, only end-of-year condition flags were used.

Data Processing. In a given month, a beneficiary is determined to be enrolled in "full" Medicare FFS if they are enrolled in both Parts A and B (including beneficiaries with dual enrollment in Medicare and Medicaid) and concurrently not enrolled in a Health Maintenance Organization (HMO). Based on this definition, monthly indicator variables are created to determine a beneficiary's full FFS enrollment status, which is later used as one of the factors to determine eligibility in claims-based impact measure denominator populations (See Section 2.3).

End-of-year condition indicator variables from both chronic condition MBSF segments provide indication of whether the beneficiary met the CCW claims criteria and/or whether the beneficiary meets the coverage criteria (enrolled in Medicare Parts A and B for the entire specified period). From these indicator variables, we further derived condition flags that indicate whether a beneficiary met the claims criteria portion of the CCW condition algorithm, regardless of whether the beneficiary met the FFS coverage criteria (FFS coverage is separately accounted for in the MBSF Base segment).

3.3   Non-Home Health Agency Claims

Purpose. We analyzed non-home health claims in order to create impact measures for spending and utilization of services outside of home health care. These claims were also used to determine a beneficiary's care setting immediately prior to a sequence of home health episodes, which was used in comparison group construction (See Section 1.5). In this section, we discuss preliminary data processing to support impact measure calculation; refer to Section 2.2 for details on determining prior care setting based on inpatient, outpatient, and SNF claims.

Data Acquisition. For this Annual Report, final action SNF (claim type codes 20 and 30), Inpatient (claim type code 60), and Outpatient (claim type code 40) claims were pulled from the CCW RIFs in August 2017, including claims with a claim through date from April 2010 through June 2017. Claims occurring in

---

[17] See CCW Condition Categories.

this date range will potentially contribute to impact measure calculation and determining a home health beneficiary's prior care setting.

Data Processing. For impact measure calculation, SNF stays were constructed based on SNF claims with an admission date starting on or after October 1, 2011; furthermore, this set of claims was subset to include only claims corresponding to beneficiaries in the HH Beneficiary Finder File (see Section 3.1). SNFs submit monthly claims throughout a beneficiary's duration of stay, which spans from admission date to discharge date; therefore, a beneficiary's SNF stay is constructed by combining each of the individual SNF claims with the same associated admission date. The SNF stay start date corresponds to the first claim's "claim from" date, while the SNF stay end date corresponds to the last claim's "claim through" date. Medicare payment amounts for each claim within a given stay are summed up to a final stay-level payment amount.

Inpatient and outpatient claims were used to support impact measures related to unplanned ACH and ED use. Inpatient claims were included if they indicated a planned ACH. To identify a planned ACH, we scanned all diagnoses (ICD 9 or 10 codes) reported on the inpatient claim and cross-referenced the list of AHRQ CCS that defined planned hospitalization, as is done in the measure developer's documentation for the two HHVBP claims-based measures.[7] Outpatient claims were included if they indicated ED visits, as identified by the presence of revenue center codes 0450-0459 or 0981; therefore, only outpatient claims containing these revenue center codes are included. This approach is in alignment with the measure developer's documentation for the HHVBP claims-based ED use without hospitalization measure.[7]

3.4   Provider Data

Purpose. We utilized publicly available data on HHAs to control for a variety of agency characteristics (i.e., ownership status, hospital-based vs. freestanding) in construction of comparison groups and D-in-D modeling.

Data Acquisition. POS data was downloaded from the CMS "Provider of Services" site.

Data Processing. The final annual POS data sets from each year 2012 – 2016 were subset to HHAs based on provider category code "5". Control types provided in the POS data were re-coded into larger groups of "non-profit" (control type codes 1,2,3; church, private not-for-profit, and other, respectively), "for-profit" (control type code 4; private for-profit), and "government-owned" (control type codes 5,6,7; federal, state, and local, respectively). Additionally, facility type codes were re-coded into groups of "hospital-based" (facility type codes 4,5,6; rehabilitation facility, SNF, and hospital, respectively), and "freestanding" (facility type codes 1,2,3,7; visiting nurse association, combination government voluntary, official health agency, and other, respectively).

3.5   County-Level AHRF Data

Purpose. Utilize county-level data from the AHRF to inform comparison group construction based on key county-level demographic information.

Data Acquisition. AHRF data are publicly available from the Health Resources and Services Administration data warehouse, from which we downloaded the 2016 county-level data set.

Data Processing. Only one data element from the AHRF data set was used in final analyses: indication of whether the county was in a rural or urban area (based on CBSA indicator). This variable is used to define rurality of a county, including beneficiaries receiving care in that county, across all analyses.

3.6   OASIS Data

3.6.1   Predicted Probabilities for the Risk-Adjusted OASIS-Based Outcome Measures

Purpose. We obtained predicted probabilities for the risk-adjusted OASIS-based outcome impact measures to support OASIS-based outcomes impact measure calculation, comparison group construction, and analysis of the seven OASIS-based outcome impact measures.

Data Acquisition. We received OASIS-based episode-level data (extracted from QIES) for each of the HHVBP OASIS-based outcome measures, in which episodes of care are determined from a series of OASIS assessments and have an episode end date ranging from 2012 through 2016.

Data Processing. The data set contains episode-level measure-specific observed and predicted probability values for each of the seven HHVBP OASIS-based outcome impact measures as well as a state and facility identifier (unique only within a given state) in which the episode of care occurred; the combination of state and facility identifier are used to look up the HHA's CCN using the CCW's HHA facility file.

3.6.2   QIES Roll-Up Measure Data for the OASIS-Based Process Measures

Purpose. We obtained QIES roll-up measure data for OASIS-based process measures in order to calculate and analyze the three OASIS-based process impact measures.

Data Acquisition. We received HHA-month-level data sets (extracted from QIES) that contain observed measure values and episode counts for each of the process measures, spanning 2012 through 2016.

Data Processing. Similar to the episode-level QIES data set described above, the data set also contained state and facility identifier (unique only within a given state) in which the episode of care occurred; the combination of state and facility identifier are used to look up the HHA's CCN using the CCW's HHA facility file.

3.7   HHCAHPS Data

Purpose. We utilize HHCAHPS data to analyze the five patient experience impact measures.

Data Acquisition. We downloaded publicly available data from the Home Health Compare (HHC) website in September 2017 for CYs 2012 – 2016.[18]

Data Processing. These data include a score value for each of the five HHCAHPS-based impact measures (see Section 2.6, "HHCAHPS-Based Impact Measures"), rounded to the nearest whole number, HHA's CCN, the number of completed surveys by respondents that received care from the given HHA, and the response rate. For this report, we used data from the January through December report from each year of our analyses, 2012 – 2016 (i.e., measurement period is the CY).

3.8   New Measures Data

Purpose. To provide data on the three new agency-reported measures.

---

[18] These data are available here.

Data Acquisition. We downloaded the Preliminary Annual TPS and Payment Adjustment Report for CY 2016, made available on the CMS Enterprise Portal on October 3, 2017.

Data Processing. This report provides HHA-level indicators for whether the agency reported on each of three measures, in addition to providing the number of points earned for each new measure and how these points are weighted to contribute to the final TPS.

3.9   Cost Reports

Purpose. Public use Cost Report files (CMS Form 1728-94 and Form 2552-10) for fiscal years 2012 – 2016 were used to obtain chain information.

Data Acquisition. Publicly available HHA-level Cost Report data sets are publicly available via CMS' Healthcare Cost Report Information System.[19]

Data Processing. The fiscal year 2016 Cost Report file was not finalized at time of reporting, and all records for 2016 represent the most current data available. For any provider number with more than one Cost Report record in a given year, the Cost Report representing the longest period of time between fiscal year beginning and end date was preferred.

3.10 Data from the HHVBP Implementation Contractor

Purpose. To support TPS calculations for non-intervention HHAs, as a means to select an appropriate comparison group.

Data Acquisition. We requested QIES measure roll-up extracts spanning 2012 – 2016 and HHA size data.

Data Processing. With guidance from the HHVBP Implementation contractor, we received the QIES roll-up data sets that align in content with those used to produce the Annual TPS and Payment Adjustment Report; however, the variable timing of QIES data extracts, as compared to those used by the Implementation contractor, may result in small discrepancies in measure values. We also received an HHA-level file containing metrics of HHA size, including a count of episodes and unique beneficiaries during 2015; these values were used by the HHVBP Implementation contractor to determine the size cohort for each of the HHAs that are subject to payment adjustment. These files, in conjunction with the publicly reported HHCAHPS measure data, allow us to calculate a TPS for both intervention and non-intervention HHAs, which serves as both a useful impact measure for comparative analyses and also as a metric on which to construct a comparison group of HHAs with similar performance rates to those in the intervention group.

4.   Analytic File Creation

Below, we describe how we created the analytic files (AFs) for each of the three Comparison Groups that were used to generate the results presented in this report.

4.1   Comparison Group Approach A

1. To construct AFs for variables that were used in Comparison Group A, we began by identifying the county in which each HH claims-based episode occurred (i.e., beneficiary residence). The CBSA where services were provided is a claim line item, indicated by value code 61; in our

---

[19] These data are available [here](#).

analyses, this is considered the most reliable source of service location. However, in the case where the beneficiary lives in a rural area (outside of a CBSA), the line item will only indicate the state of service; further, many CBSAs include multiple counties. To address these situations while still leveraging the accuracy of the CBSA for determining location of service, we followed a multi-step approach for determining the county in which services were provided during an HH episode:

a. Using the National Bureau of Economic Research's CBSA to Federal Information Processing Standards (FIPS) county crosswalk file, we mapped any valid CBSA code to one or more of its associated counties.

   i. If a non-rural CBSA maps to one valid county, then the episode was associated with that county in analyses.

   ii. If the line item indicates a rural area (i.e., no valid CBSA), or if the CBSA does not map to a single valid FIPS county code, then we proceeded to the next step.

b. We next used patient ZIP code on the OASIS assessment nearest to the claims-based episode start date to determine the county where home health services were provided. We included any OASIS assessment within 90 days of the claims-based episode start date.

   i. For the selected OASIS assessment, we used the U.S. Department of Housing and Urban Development (HUD) ZIP to county crosswalk to map the county of service. For instances when the CBSA from Step 1a mapped to multiple counties, we ensured that the county derived from the OASIS assessment ZIP code matched one of the counties within the CBSA.

   ii. For instances when the episode of care was provided in a rural area, we ensured that the county derived from the OASIS assessment ZIP code existed in the same state that was indicated by the claim line item value.

   iii. If there were no OASIS assessments for the beneficiary within 90 days of the claims-based episode start date, *or* the county derived from the OASIS ZIP code did not align with the CBSA or rural area's state, then we proceeded to the next step.

c. Next, we examined the monthly beneficiary county of residence data from the MBSF for the month in which the claims-based episode ended.

   i. If the CBSA from Step 1a mapped to multiple counties, we maintained the county from the MBSF if it matched one of the counties within the CBSA.

   ii. If services were provided in a rural area, we maintained the county from the MBSF provided that the corresponding state matched the state of the rural area.

   iii. Otherwise, we proceeded to the next step.

d. We next examined the mailing address county associated with the HH claim. We applied the same logic as described in Step 1c, but maintained the county from the claim (vs. MBSF).

e. Next, we examined the MBSF monthly beneficiary county of residence data, but expanded our search from Step 1c to include the 3 months preceding and following the month in which the HH episode ended. We applied the same logic as described in Step 1c, giving higher priority to counties from months that were closer to and preceding the

month in which the HH episode ended (e.g., 1 month before takes precedence over 2 months before, and also takes precedence over 1 month after). If none of the 6 months evaluated yielded a county that aligns with the CBSA or rural area, we proceeded to the next step.

f. Next, we examined the ZIP in which the HHA is located, sourced from the CCW's HHA facility files. Using the HUD ZIP code to county crosswalk file (as in Step 1b), we mapped the associated county in which the HHA is located.

    i. If the CBSA from Step 1a mapped to multiple counties, then we assigned the HHA's county to the episode only if it matched one of the counties within the CBSA.

    ii. If services were provided in a rural area, then we assigned the HHA's county to the episode only if the corresponding state matched the state of the rural area.

    iii. Otherwise, we proceeded to the final step to determine the county in which services were provided.

g. Finally, to account for any remaining episodes that had not yet been assigned a county through this multi-step process (<10% of total episodes), we repeated Steps 1b through Step 1e, but *without* enforcing that the county align with the CBSA or rural area state found on the claim.

2. We excluded episodes served by HHAs located in areas outside of the 50 US states, which were not included in the regional groupings as specified in the CY 2016 Final Rule (e.g., the District of Columbia, Puerto Rico, and other US territories).[1]

3. Using the county identified in Step 1, we then calculated the three claims-based impact measures associated with this comparison group at the county level (See Section 2.3).

a. Medicare FFS beneficiary counts were obtained at the county level from the MBSF, in which full Medicare FFS eligibility is defined, as described in Section 3.2.

b. Any counties that did not have home health episodes in each quarter of the time period under consideration (i.e., Q1 2012 through Q4 2016) were excluded.

4. County-level measure values from Step 3 were merged with the county-level file described in Section 3.5, to obtain rurality indication from the AHRF data. If the CBSA indicator from the AHRF data indicated that the county was not part of a CBSA, then the county was considered rural.

5. All remaining counties were divided into two groups: those that occur within an HHVBP state ("intervention episode") and those that occur within one of the 41 non-HHVBP states ("comparison episode"). This AF contained 2,960 counties, contributing 59,200 county-quarter observations through 2016.

This county-level file was then used to construct comparison groups for the three impact measures (Table 2).

6. For each intervention state, we used entropy balancing to select weights for counties in all 41 comparison states.[20] A weight for each comparison county was chosen such that the weighted mean of impact measures over all comparison counties equaled the mean of the impact measures in the intervention state in each of the 16 baseline quarters, 2012 through 2015 (i.e., 16 means, one for each quarter, over which we balance).

   a. One set of weights was chosen such that the Medicare home health payment per FFS beneficiary *and* the number of home health episodes per FFS beneficiary were equal between each treatment state and the reweighted comparison.

   b. An additional weighted comparison was created for each intervention state to compare changes in Medicare home health payment for home health episodes. Again, entropy balancing was used to identify weights such that the average of the impact measure was equal between the intervention state and the weighted comparisons. However, in addition to baseline values of the impact measure, we used the entropy balancing procedure to also equate the proportion of home health beneficiaries residing in a rural area in one of the baseline quarters, Q1 of 2014. Because home health beneficiaries residing in a rural area (non-CBSA county) have triggered a rural add-on payment since 2010, and the rural add-on will not apply to episodes ending on or after January 1, 2018, we sought to balance treatment and comparison populations to incur similar time-varying effects from this change.[21]

4.2 Comparison Group Approach B, Claims-Based Impact Measures

1. To construct an AF for the claims-based impact measure analyses, we began by calculating the impact measures. To construct the impact measures, we began with all HH episodes that start in a quarter.

2. We then applied the following exclusions to each HH episode:

   a. HH episodes for patients who are alive and are not continuously enrolled in FFS Medicare for the 60 days following the start of the HH episode.

      i. Starting with HH claims-based episodes, as defined in Section 3.1, we determined the beneficiary's FFS enrollment status relative to the episode start and end dates. First, using MBSF monthly FFS enrollment flags for a given beneficiary, we constructed time periods of consecutive months of full FFS enrollment status. Next, the resulting full FFS enrollment periods were joined to the HH episodes by beneficiary identifier and whether the HH episode start date was contained within the enrollment period. If the beneficiary was not continuously enrolled in full FFS for the entire time period starting with episode start date and ending with the earlier of beneficiary death date (if beneficiary is

---

[20] Hainmueller J. (2012) Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis,* 20(01):25-46.

[21] HHS, CMS. (2017) 42 CFR 484. Medicare and Medicaid Programs; CY 2018 Home Health Prospective Payment System Rate Update and CY 2019 Case-Mix Adjustment Methodology Refinements; Home Health Value-Based Purchasing Model; and Home Health Quality Reporting Requirements; Final Rule. *Federal Register* 82 FR 51676.

deceased) or episode start date plus 59 days, the episode was excluded from each of the measure denominators.

b. HH episodes for patients who were not FFS eligible in the six months prior.
   i. Furthermore, if the beneficiary was not enrolled in full FFS for the full 6 months prior to episode start date, then the episode was excluded from each of the measure denominators.

c. HH episodes that begin with a LUPA claim.
   i. Episodes subject to LUPA were identified by the LUPA indicator code available in the HH claims data, in which a value "L" indicates the episode payment was adjusted due to low utilization and is subject to exclusion from each of the measure denominators.

d. HH episodes in which the patient receives service from multiple agencies during the 60-day window.
   i. If a beneficiary has multiple HH episodes for different HHAs in which one episode starts within 60 days of the previous episode start date, the first episode was excluded from each of the measure denominators.

e. For the two impact measures that were derived from the first HH episode only (i.e., ED Use (no Hospitalization) per First FFS HH Episodes and Unplanned Hospitalization per First FFS HH Episodes), we also restricted HH episodes that were not the first episode in the sequence of HH stay.
   i. As defined in Section 3.1, a new sequence of HH episodes begins when a beneficiary begins a new HH episode after more than 60 days since the end date of their most recent prior HH episode. This new HH episode (first in sequence) was included in the denominator of each of these two impact measures, while any ensuing HH episode within the same sequence was excluded from the denominators of these two impact measures.

3. Based on these exclusions, we then constructed the claims-based impact measures. See Section 2.3 for specific information on each of the impact measure's numerator and denominator definitions.

4. In preparation for the episode-level matching, we merged in data files that contained the variables over which episodes would be matched on.

a. From the MBSF, we merged the following beneficiary data associated with each HH episode based on the year in which the episode begins: age, gender, Medicare entitlement, and chronic condition indicators: chronic kidney disease, congestive heart failure, diabetes, and ulcers.

b. We added prior care setting indicators associated with each sequence of HH episodes, which were derived from the non-HHA claims files (See Section 2.2). For episodes that are 3rd or later within a sequence, prior care setting was defined as "Home Health", because these episodes began a significant amount of time after the sequence start date (reference point for prior care setting determination). This determination will impact one-to-one matching in subsequent steps, as episodes that are third or later

within a sequence could only potentially match to other episodes that are third or later within a sequence.

5. After the impact measures were created, all episodes beginning in the quarter and included in the impact measure denominator were divided into two groups: those that occur within an HHVBP state ("intervention episode") and those that occur within one of the 41 non-HHVBP states ("comparison episode").

6. We then used one-to-one matching with replacement to match episodes of care from HHVBP ("intervention") states to episodes of care in the non-HHVBP ("comparison") states. Episode-level matching was done quarter-by-quarter, in which an intervention episode beginning in one quarter can only be matched to a comparison episode beginning in the same quarter. For intervention episodes that match with multiple comparison episodes, only one comparison episode was randomly selected from the pool of matches.

7. Within each quarter, an intervention episode was matched to one comparison episode belonging to the same measure denominator based on the following characteristics:
    ▪ Prior care setting indicators
    ▪ Age category
    ▪ Gender
    ▪ Medicare entitlement reason
    ▪ Presence of four chronic conditions: chronic kidney disease, congestive heart failure, diabetes, and ulcers

Aside from the beneficiary chronic condition component, indicator variables for each of these characteristics were constructed as defined in the risk adjustment methodology for the two HHVBP claims-based measures.[7] The chronic conditions were included since they were found to be both highly prevalent among beneficiaries and significant predictors of the impact measures.

Matching was done separately within each of the two measure denominators: impact measures that included all episodes in a sequence; and impact measures that included only the first episode in a sequence. This ensures that intervention episodes included in the measure denominator can only be matched to comparison episodes included in the measure denominator (i.e., only non-first comparison episodes were matched to non-first treatment episodes). Furthermore, among impact measures that included all episodes in a sequence, only third or later intervention episodes could be matched to third or later comparison episodes. The two impact measures for first episode in a sequence had the same matched comparison. Similarly, the impact measures for all episodes in a sequence had the same matched comparison.

8. Intervention episodes with no exact matches for comparison episodes were excluded from analyses. Across all quarters during the measurement period, this led to an exclusion of about 0.5% of intervention episodes in the impact measures for first episode in sequence and about 0.3% of intervention episodes in the impact measures for all episodes in sequence.

9. After deriving our intervention episodes and their associated matched comparison episodes, we then merged the resulting data set with additional data files to obtain HHA characteristics that were used as covariates in our D-in-D model specification for the claims-based impact measures (see Section 1.5.1 for details). Merging onto the HHA-year level POS file by the HHA's CCN and the year in which the episode of care began, we derived indicator variables for the different levels of the following HHA characteristics from the POS file: ownership type (non-profit, for-profit, government-owned) and setting (freestanding vs. hospital-based). In order to capture HHA size, we calculated the total number of claims-based episodes of care provided by each HHA per year; these HHA-year level counts were then merged onto the matched episode data set by CCN and year. Furthermore, we categorized HHA size as follows: 0-59 episodes; 60 to 249 episodes; 250 to 499 episodes; 500 to 999 episodes; 1000+ episodes. We used the resulting AFs to conduct the D-in-D analyses (discussed below) for the claims-based impact measures.

10. As explained in Section 1.5.1, the vector of covariates included in the D-in-D equation for claims-based measures reflected both beneficiary and HHA characteristics. HHA characteristics included ownership type, freestanding status, and agency size. Patient characteristics included beneficiary chronic condition indicators for: Alzheimer's disease and related disorders or senile dementia, depression, anxiety, psychotic and stress disorders, disabilities (intellectual, learning or developmental delays), cancer (breast, colorectal, prostrate, lung, endometrial or leukemia), liver disease, arthritis, chronic obstructive pulmonary disease, ischemic heart disease, atrial fibrillation, and anemia.

4.3 Comparison Group Approach B, OASIS-Based Outcome Impact Measures

1. To create an AF for the OASIS-based outcome impact measures, we used the episode-level predicted probabilities for the risk-adjusted OASIS-based outcome measures from CMS-CCSQ (sourced from QIES; see Section 3.6.1). As described above, the data set contains episode-level measure-specific observed and predicted probability values for all home health episodes of care ending from 2012 through 2016, for each of the seven HHVBP OASIS-based outcome measures. Observed values indicate whether the episode of care is included in the measure denominator and further if the episode is included in the measure numerator. For home health episodes of care that are eligible for inclusion in the measure denominator, the measure-specific predicted probability value (between 0 and 1) captures the likelihood (based on patient risk factors) of the episode being included in the measure numerator. The combination of state and facility identifier were used to look up the HHA's CCN using the CCW's HHA facility file.

2. All episodes ending in the quarter and included in the measure denominator were divided into two groups: those that occurred within an HHVBP state ("intervention episode") and those that occurred within one of the 41 non-HHVBP states ("comparison episode").

3. We then matched each intervention episode to one or more comparison episodes with the exact same predicted probability value (that is, we utilized episode-level predicted probabilities as a proxy for matching on all covariates included in the risk adjustment model). If an intervention episode matched to multiple comparison episodes, we randomly selected one matching comparison episode. This process was applied to each of the seven OASIS-based outcome measures and within each of the calendar quarters from 2012 through 2016.

Because each of the seven OASIS outcome measures have different denominators (that is, episodes of care included in the denominator differ across measures), we conducted the separate matches for each measure. Through our episode matching process, we dropped between 0.03% and 0.09% of intervention episodes of care across the seven OASIS outcome measures.

4. After deriving our intervention and comparison groups of episodes, we then merged the resulting data set with additional data files to obtain HHA characteristics that were used as covariates in our D-in-D model specification for the OASIS-based outcome impact measures. Merging onto the HHA-year level POS file by the HHA's CCN and the year in which the episode of care ended, we derived indicator variables for the different levels of the following HHA characteristics from the POS file: ownership type (non-profit, for-profit, government-owned) and setting (freestanding vs. hospital-based). In order to capture HHA size, we calculated the total number of OASIS episodes of care provided by each HHA per year based on the data from CMS-CCSQ; these HHA-year level counts were then merged onto the matched episode data set by CCN and year. Furthermore, we categorized HHA size as follows: 0-59 episodes; 60 to 249 episodes; 250 to 499 episodes; 500 to 999 episodes; 1000+ episodes. We used the resulting AF to conduct the D-in-D analyses (discussed below) for the OASIS-based outcome impact measures.

4.4   Comparison Group Approach C

1. To construct AFs for variables that were used in Comparison Group Approach C, we started by constructing an HHA level file from several different data sources. We began with HHAs that had non-missing POS data and were active at some point in 2015, which included 12,595 agencies. To determine whether an HHA is active at a given point in time, we referenced the March 2017 quarterly release of the POS file and use original participation and termination dates.

2. We then merged the HHA-level data set with additional data files to obtain other HHA characteristics that were used to define our reweighting strata. Merging by HHA's CCN:
   a. We obtained three variables from the POS:
      - **Ownership type** (non-profit, for-profit, government-owned);
      - **Setting** (freestanding vs. hospital-based); and
      - **Agency Age** (based on original participation date). HHAs were categorized into one of three agency age categories: Original participation date prior to 2005; 2005 – 2011; and 2012 – 2015.
   b. We obtained **Chain Status** in 2015 from Cost Reports. For HHAs that had a missing Cost Report in 2015, chain status was imputed using the Cost Report submitted in the last year prior to the missing observation. Agencies with no Cost Reports or missing chain status information were assumed to be unaffiliated with a chain.
   c. We obtained **HHA Size** from a data set provided by the HHVBP Implementation contractor. We defined size based on the unique beneficiaries reported during 2015, and created three categories: 1 to 59 beneficiaries; 60 to national median value; and greater than national median value.

    d.   We obtained the 2015 TPS values (as an overall measure of quality of care, fixed at baseline) for each HHA that we computed from Medicare claims, OASIS QIES extracts, and HHCAHPS data (see Section 2.7). National 2015 TPS values were used to place values into **2015 TPS Quintiles**.

The bolded HHA characteristics were used to construct the strata for our reweighting mechanism for all impact measures in Comparison Group C. Although this allows for potentially 540 strata, after applying the exclusions described in the next step, there were 168 strata for the HHA cohort to support HHCAHPS analysis, and 172 strata for the HHA cohort to support the OASIS process and 2016 TPS impact measures analyses.

3.   We then restricted the data set to those HHAs that: 1) had a non-missing 2015 TPS value (i.e., had data in both the baseline year for capturing any improvement in performance (i.e., 2014) and in the performance year (i.e., 2015); and 2) were active in 2016. Among the remaining 8,588 HHAs, we assigned each into one of two groups: those in an HHVBP state ("intervention") and those in one of the 41 non-HHVBP states ("comparison").

4.   Because the distribution of HHAs across the above strata is not entirely comparable between the intervention group and the non-intervention group, we applied a weighting mechanism to ensure that the relative contribution of agencies in each stratum of the non-intervention group to the analysis corresponds to the relative contribution of agencies in the same stratum of the intervention group.

    a.   To accomplish this, we first determined the proportion of all HHAs in the intervention group that reside within a given stratum.

    b.   To achieve similar proportions for each stratum among the non-intervention HHAs, we also determined the proportion of all non-intervention HHAs within each stratum, and then defined the non-intervention weight for a given stratum as the ratio of the corresponding proportion for intervention HHAs within a given stratum to the corresponding proportion of non-intervention HHAs in the same stratum.

    c.   For all analyses of HHA-level impact measures, the resulting weight for each provider stratum was applied to all measure values for non-intervention HHAs residing within that stratum, with all intervention HHAs receiving a weight of 1. This reweighting approach yielded a comparison group of non-intervention HHAs that is similar to the intervention group with regard to both the individual agency characteristics (described above) and an overall measure of quality of care fixed at baseline (i.e., the 2015 TPS quintiles[22]) used to define the strata and the various combinations of these factors.

    d.   The measurement period for calculating weights was quarterly (for the years 2012-2016) for the OASIS-based process measures, and yearly (for the years 2012-2016) for HHCAHPS measures and 2016 TPS impact measure.

        i.   For the OASIS-based process impact measures, we reweighted on a quarterly basis for the years 2012-2016, reflecting data from the set of HHAs

---

[22] For the OASIS process measure and TPS analyses, the national TPS quintiles are: 17.21, 24.67, 32.25, 43.16, and 97.63. For the HHCAHPS measure analyses, the national TPS quintiles are: 17.36, 24.65, 32.16, 42.56, and 93.99.

     from the original cohort of 8,588 HHAs that were also active at the start of each quarter.

  ii. For the HHCAHPS impact measures, we reweighted on an annual basis, reflecting the set of HHAs from the original cohort that were active at the start of the year and also had at least one completed survey in that year.

  iii. For the 2016 TPS impact measure, we reweighted on an annual basis, reflecting the set of HHAs from the original cohort that were active for both the entire prior year (TPS baseline) and current year (TPS performance year).

5. The HHA reweighting approach was applied to all HHVBP states combined. In addition, given the exit of HHAs from the industry over time, separate reweighting was used for each measurement period across all baseline and intervention years to account for HHAs that were no longer certified or in operation as of the beginning of the measurement period, and which might otherwise have resulted in an imbalance in the above characteristics of the remaining HHVBP HHAs in operation and their respective comparison groups.

  a. As discussed earlier (in Section 1.6), there is a risk that this approach may lead us to adjust away certain effects of HHVBP, which are also of interest if there are changes in the composition of HHAs in operation in the intervention states during 2016 that are a direct result of the HHVBP Model. However, at this early stage of the operation of HHVBP, which reflects the experience of HHAs at least one year prior to the application of the initial HHVBP payment adjustments in 2018, we view this approach as potentially accounting for changes over time in HHA characteristics that were likely already underway prior to HHVBP and are more likely to represent a source of potential confounding for early estimates of the impact of the Model on quality of care.

  b. For the weights applied across each measure and measurement period, HHAs contributed to the strata weights if they met the criteria described in Step 4d above, regardless of whether they contributed a value to the measure calculation in a given measurement period. The criteria described in Step 4d were designed to capture all HHAs that would be eligible to contribute a measure value.

6. We then merged data files (by HHA CCN and year) that contained the relevant impact measures:

  a. We obtained 2016 TPS values for each HHA that we computed from Medicare claims and OASIS QIES extracts and HHCAHPS data (see Section 2.7 for how the 2016 TPS impact measure was computed).

  b. We obtained OASIS process measure data from CMS (sourced from QIES). For each of the three HHVBP OASIS-based process measures, the data set contained agency-month measure values, monthly episode counts, state, and facility identifier (unique only within a given state) for all HHAs from 2012 through 2016. We rolled up the agency month-level data to agency quarter-level, weighting the monthly measure values by the corresponding monthly episode counts. We used the combination of state and facility identifiers to look up the HHA's CCN using the CCW's HHA facility file.

    c.   We obtained HHCAHPS-based impact measures for 2012-2016 from the HHC website (see Section 3.7). From the agency-level data sets, we retained HHA's CCN, year, and HHCAHPS impact measure values. We appended the five annual data sets into one file.

7.   HHAs that were balanced (i.e., had non-missing weights) were included in the D-in-D analysis for the 2016 TPS impact measure. To be included in the D-in-D analysis for the OASIS process measures, we also required HHAs to have non-missing values for at least one of the three OASIS process measures. The calculated weights were applied in our D-in-D analysis (discussed below).

8.   We calculated separate weights for each year or quarter depending on the time period over which each measure was calculated. In particular, we calculated annual weights for the TPS and HHCAHPS measures and quarterly weights for the OASIS process measures. The strata weights were calculated separately for each of the three categories of HHA-level impact measures (HHCAHPS, OASIS process measures and TPS), but do not differ across impact measures within each category (e.g., the same strata weights are used for the three OASIS process measures). Within each of the three categories, agencies must have at least one non-missing impact measure to contribute to the stratum weight calculation for that time period. The number of HHAs with available HHCAHPS data for the D-in-D analysis did not differ across the five HHCAHPS measures. However, we note that, for the OASIS quarterly process measures, some HHAs had data available for some but not all of the process measures (as shown in Table 8n of the Appendix of Supplemental Tables and Results). As a result, there are exceptions among the OASIS process measures where the strata weights that are being applied may not be exactly equal for the intervention and comparison groups, based on those strata where one or more HHAs have missing data for a given OASIS process measure in a given quarter. This was not a common occurrence, as the number of agency quarters with available measure data in 2015 and 2016 differed by less than 0.1% between the OASIS drug education and pneumococcal vaccine measures and by approximately 2%-3% between the influenza measure and the other two process measures (see Table 8n of the Appendix of Supplemental Tables and Results).

The final analytic data set used for the OASIS process measure analyses contained 1,622 intervention HHAs in 2016, corresponding to 6,305 agency quarters. The final analytic data set used for the 2016 TPS impact measures analyses contained 1,595 intervention HHAs in 2016. The final analytic data set used for the HHCAHPS analyses contained 1,337 intervention HHAs in 2016.[23]

5. Comparison Group Testing Results

Below, we summarize the results from our Comparison Group construction and testing.

---

[23] Approximately 34% of HHAs in 2015 having missing HHCAHPS data (i.e., five of the impact measures). By definition, these HHAs are not included in the HHCAHPS analyses, but *are* included in the TPS and OASIS process measures analyses.

5.1   Falsification Tests

We conducted falsification tests for all the impact measures analyzed in the Annual Report. This testing allowed us to examine the performance of the estimator and comparison group construction, knowing that the effect of HHVBP should be zero for the measures examined.

To conduct the falsification tests, we estimated the effects of HHVBP for each of the four quarters in 2016 for the 41 non-intervention states, where the effect should be null since HHVBP was not implemented in these states. The falsification tests were conducted exactly as outlined for the treatment states above (see Section 1.4 for Comparison Group Approach A, Section 1.5 for Comparison Group Approach B, and Section 1.6 for Comparison Group Approach C), falsely changing the treatment indicator to one for a non-intervention state and repeating the process for each of the 41 non-intervention states. Since each iteration of the falsification test uses one non-intervention state as a "treated" unit, the 40 remaining non-intervention states are used to comprise the comparison groups (as opposed to 41 states that are used to comprise comparisons for the actual intervention states). The tests generate 164 estimated effects of HHVBP—41 times four quarters—in states where HHVBP was not implemented and should not show a significant HHVBP effect.

We examined the results of these tests and determined if the average estimated effect was different from zero and how frequently the computed standard errors rejected a true null hypothesis (that is, the Type I error rate). The rate was calculated by dividing the number of statistically significant estimated effects divided by 164, the total number of estimates in the falsification tests.

Table 22 presents the results of the falsification tests (i.e., the mean estimated effect of HHVBP) and Type I error rate for all of the impact measures analyzed in the Annual Report, organized by Comparison Group Approach. We also include the 2015 mean values for the impact measures to facilitate interpretation of the results. For most impact measures, the average estimated effect of HHVBP on the 41 comparison states suggest that there is no bias in our D-in-D estimators. This can be seen in the second column of Table 22, where many of the estimates are near zero. For nearly all of the impact measures, the average estimated effect of HHVBP corresponded to less than 1% of the average 2015 value for all HHVBP states combined, with the exception of the SNF spending measure and TPS where the average estimated effect corresponded to less than 3% of the average 2015 value (e.g., $16.15/$654.50=2.5% for spending on SNF stays).

Type I error for all the impact measures ranged from 2.4% to 31.1%. Assuming a 0.05 level of significance, a Type I error rate greater than 5% is more than what is expected to occur by chance. In particular, the Type I error rates for many of the claims-based measures and most of the OASIS-based measures were higher than would be expected by chance.[24] Statistical inference of any impact measures with Type I error rates over 5% should be considered cautiously, as the findings suggest that our clustering may not appropriately adjust the standard errors for estimating the effect of HHVBP. In the second year of the evaluation, we will update the data used for matching claims-based measures in comparison group B. We will, again, perform falsification tests using the updated comparison populations once any changes are made.

---

[24] We note this also affects the Type I error rate for TPS, given that together, the claims and OASIS measures account for 12 of the 17 measures that comprise the TPS.

Table 22. Falsification Test Results for Impact Measures

Measure	Average Estimated Effect of HHVBP^	Average 2015 Value for All HHVBP States	Type I Error Rate (# of Significant Estimates, α ≤ 0.05, Divided by 164)
Comparison Group Approach A (Claims-Based Measures)			
Average Number of FFS HH Episodes/1,000 FFS Beneficiaries	0.24	47.13	5.5%
Average Medicare HH Payment/FFS Beneficiary	$0.72	$138.84	4.5%
Average Medicare HH Payment/FFS HH Episode	$12.07	$2,945.75	23%
Comparison Group Approach B (Claims-Based Measures)			
Unplanned Hospitalization/First FFS HH Episodes (%)	-0.0051	16.7%	7.3%
ED Use (no Hospitalization)/First FFS HH Episodes (%)	0.0918	11.9%	14.6%
Unplanned Hospitalizations/All HH FFS Episodes (%)	-0.0555	17.5%	13.4%
SNF Use/All HH FFS Episodes (%)	-0.1024	5.2%	12.8%
Medicare Spending for Unplanned Hospitalizations/All HH FFS Episodes	$8.42	$2,177.72	7.9%
Medicare Spending for SNF Stays/All HH FFS Episodes	-$16.15	$654.50	10.9%
Comparison Group Approach B (OASIS-Based Outcome Measures)			
Improvement in Ambulation-Locomotion (%)	-0.0572	68.3%	14.6%
Improvement in Bathing (%)	0.0403	72.2%	23.8%
Improvement in Bed Transferring (%)	-0.0682	64.7%	14.0%
Discharged to Community (%)	-0.0078	72.4%	28.7%
Improvement in Dyspnea (%)	-0.4586	70.1%	31.1%
Improvement in Management of Oral Medications (%)	-0.1912	55.0%	5.5%
Improvement in Pain Interfering with Activity (%)	-0.0340	71.9%	31.1%
Comparison Group Approach C (OASIS-Based Process Measures)			
Drug Education on Medications Provided to Patient/Caregiver during Episodes of Care (%)	0.3562	94.5%	9.8%
Influenza Immunization Received for Current Flu Season (%)	0.2276	58.7%	20.7%
Pneumococcal Polysaccharide Vaccine Ever Received (%)	0.1291	67.0%	25.6%
Comparison Group Approach C (HHCAHPS-Based Measures)			
How often the home health team gave care in a professional way (%)	0.1395	88.8%	4.9%
How well did the home health team communicate with patients (%)	-0.0044	85.8%	2.4%
Did the home health team discuss medicines, pain, and home safety with patients (%)	-0.1278	82.9%	9.8%
Would patients recommend the home health agency to friends and family (%)	0.1426	79.4%	5%
How do patients rate the overall care from the home health agency (%)	0.1165	84.4%	7.3%
Comparison Group Approach C (TPS)			
TPS	0.9074	30.9	21.9%

^For measures that represent percentages (e.g., unplanned hospitalization per first FFS HH episodes), the estimated effects are reported in percentage points.

Note: Type I Error rate indicates the percentage of the 164 estimated effects—41 states times 4 quarters—that were statistically significant at the 5% level of significance.

5.2   Parallel Trend Tests

The D-in-D estimator requires the assumption that, in the absence of treatment, adjusted impact measures for the intervention and comparison populations would move in a parallel manner. To help inform the validity of the parallel trend assumption, we tested for differential changes in impact measures between the intervention and comparison between the first two years of the baseline period (i.e., 2012-2013) and the latter two years of the baseline period (i.e., 2014-2015). That is, we applied a D-in-D specification falsely assigning 2012-2013 as the baseline years and 2014-2015 as post-intervention time period, and computed D-in-D estimates for the eight quarters during 2014-2015. These estimated effects of HHVBP for each of the eight quarters during 2014-2015 should be null since HHVBP was not implemented in 2014. D-in-D estimates that are not statistically different from zero suggest that the impact measures moved in a parallel manner over the baseline period. We followed the same D-in-D specification that was used for each of the Comparison Group Approaches (see Section 1.5 for Comparison Group Approach B and Section 1.6 for Comparison Group Approach C).[25]

We tested parallel trends for nine intervention states only. The tests generated 72 estimated effects of HHVBP—nine intervention states times eight quarters—for a time period where HHVBP was not implemented and should not affect outcomes.

Table 23 presents the results of the parallel trend testing (i.e., the mean estimated effect of HHVBP during 2014-2015, where a null effect is expected) and Type I error rate for the impact measures in Comparison Group Approaches B and C.[25] We also included the 2012 mean values for the impact measures to facilitate interpretation of the results. For all of the impact measures, the average estimated effect of HHVBP corresponded to less than or equal to 1% of the average 2012 value for all HHVBP states combined, with the exception of the TPS where the overall mean D-in-D estimate corresponded to 2% of the average 2013 TPS value. Overall, the results provide some indication of null effects – that is, that intervention and comparison populations moved in a parallel manner pre-HHVBP implementation.

Across the parallel trends tests performed on the impact measures, the Type I error rates ranged from 0% to 38.9%. Assuming a 0.05 level of significance, a Type I error rate greater than 5% is more than what is expected to occur by chance. Among the claims measures, the higher than expected error rates were largely driven by Florida's estimates,[26] which suggests refinements to the model may be needed in the future. As specified in the D-in-D estimation equations above (see Section 1.5 for Comparison Group Approach B and Section 1.6 for Comparison Group Approach C), standard errors are clustered at the agency-level for both episode-matching and agency-reweighting measures. For now, statistical inference of any impact measures with Type I error rates over 5% should be considered cautiously, as the findings suggest that model specifications may not have appropriately adjusted standard errors for estimating the effect of HHVBP.

---

[25] By definition, Comparison Group Approach A – which employed entropy balancing – ensured that the intervention and comparison groups are matched exactly on the impact measures in the pre-intervention period. As such, parallel trend testing for these three measures was unnecessary.

[26] Excluding Florida, the Type I error rate for the other eight HHVBP states ranged from 0% to 4.7% for all the claims-based measures, except for the SNF Use impact measure, which still had a higher than expected error rate (7.8%).

Table 23. Parallel Trend Test Results for Impact Measures

Measure	Average Estimated Effect of HHVBP^	Average 2012 Value for All HHVBP States	Type I Error Rate (# of Significant Estimates, α ≤ 0.05, Divided by 72)
Comparison Group Approach A (Claims-Based Measures)			
Average Number of FFS HH Episodes/1,000 FFS Beneficiaries	N/A*	49.89	N/A*
Average Medicare HH Payment/FFS Beneficiary	N/A*	$144.81	N/A*
Average Medicare HH Payment/FFS HH Episode	N/A*	$2,902.57	N/A*
Comparison Group Approach B (Claims-Based Measures)			
Unplanned Hospitalization/First FFS HH Episodes (%)	-0.0203	16.4%	11.1%
ED Use (no Hospitalization)/First FFS HH Episodes (%)	-0.0459	10.6%	6.9%
Unplanned Hospitalizations/All HH FFS Episodes (%)	0.1111	17.2%	13.9%
SNF Use/All HH FFS Episodes (%)	0.0425	4.7%	15.3%
Medicare Spending for Unplanned Hospitalizations/All HH FFS Episodes	$5.67	$2,001.02	13.9%
Medicare Spending for SNF Stays/All HH FFS Episodes	$1.04	$584.07	2.8%
Comparison Group Approach B (OASIS-Based Outcome Measures)			
Improvement in Ambulation-Locomotion (%)	0.1277	59.8%	0%
Improvement in Bathing (%)	0.3810	67.7%	2.8%
Improvement in Bed Transferring (%)	0.1314	55.7%	2.8%
Discharged to Community (%)	-0.0810	72.5%	20.8%
Improvement in Dyspnea (%)	-0.0082	63.1%	0%
Improvement in Management of Oral Medications (%)	0.3340	47.0%	2.8%
Improvement in Pain Interfering with Activity (%)	-0.1212	69.7%	8.3%
Comparison Group Approach C (OASIS-Based Process Measures)			
Drug Education on Medications Provided to Patient/Caregiver during Episodes of Care (%)	-0.0552	90.0%	38.9%
Influenza Immunization Received for Current Flu Season (%)	-0.2094	63.2%	9.7%
Pneumococcal Polysaccharide Vaccine Ever Received (%)	0.6682	65.7%	15.3%
Comparison Group Approach C (HHCAHPS-Based Measures)			
How often the home health team gave care in a professional way (%)	-0.2796	88.6%	11.1%
How well did the home health team communicate with patients (%)	-0.3026	85.9%	22.2%
Did the home health team discuss medicines, pain, and home safety with patients (%)	0.0456	82.6%	5.6%
Would patients recommend the home health agency to friends and family (%)	-0.1157	79.9%	11.1%
How do patients rate the overall care from the home health agency (%)	-0.2398	84.5%	0%
Comparison Group Approach C (TPS)			
TPS	-0.5852	29.6**	10%

^ For measures that represent percentages (e.g., unplanned hospitalization per first FFS HH episodes), the estimated effects are reported in percentage points.

* By definition, Comparison Group Approach A ensured that the intervention and comparison groups are matched exactly on the impact measures in the pre-intervention period. As such, parallel trend testing was unnecessary.

** 2012 TPS values were unavailable, so 2013 TPS values are reported instead.

NOTE: Type I Error rate indicates the percentage of the 72 estimated effects—nine intervention states times 8 quarters—that were statistically significant at the 5% level of significance

5.3   Descriptive Results for HHVBP and Comparison Group across the Three Comparison Group Approaches

We generated descriptive statistics on patient demographic variables in the baseline period for all HHVBP states combined and its respective comparison group for each of the impact measures (See Tables 24-30). Consistent with our analytic approach, we use 2012-2015 as the baseline period for impact measures that are evaluated using Comparison Group Approaches A and B. For Comparison Group Approach C, the baseline period is 2015. Across all comparison group approaches, variables that are used for matching or reweighting (see Table footnotes) are exactly balanced between the intervention group and its respective comparison group by design. Other patient characteristics not used in the matching or reweighting process are often similar between the HHVBP states and its comparison group.

Table 24. Comparison Group A: Beneficiary Characteristics for HHVBP and Comparison Group

Characteristic	HHVBP	Comparison Group
Age		
% age <65	12.2	13.1
% age 65-74	23.8	24.3
% age 75-84	31.5	31.6
% age 85+	32.6	31.0
% Female	62.0	61.6
Race/Ethnicity		
% White	82.3	77.9
% Black	10.7	9.8
% Hispanic	4.9	8.2
% Other	2.1	4.1
% Rural	6.9	19.0
% Dual Eligible	28.5	32.7
Chronic Conditions		
% Chronic Kidney Disease	46.4	45.2
% Congestive Heart Failure	42.9	44.9
% Diabetes	44.4	45.9
% Pressure Ulcers and Chronic Ulcers	25.2	24.5
% Alzheimer's Disease and Related Disorders or Senile Dementia	34.5	31.3
% Ischemic Heart Disease	55.1	54.6
% Anemia	60.8	57.9
Reason for Medicare Entitlement		
% Original End-Stage Renal Disease (ESRD)	1.2	1.5
% Original Disabled	25.3	26.6
Current ESRD and Disabled status		
% Current ESRD	0.8	1.0
% Current Disabled	12.0	12.8

For each characteristic, weighted averages are calculated over the baseline period (2012-2015) using county level data from the 2015/2016 Area Health Resource File. HH claim-level and beneficiary-level data, taken from HH claims, the MBSF, and CCW HCC risk score data, were also aggregated to the county level.

Weights used represent those for the two outcomes over FFS beneficiaries (average number of HH episodes per 1,000 FFS beneficiaries and average Medicare HH payment per FFS beneficiary).

Table 25. Comparison Group B: Beneficiary Characteristics for HHVBP and Comparison Group for Claims-Based Measures

Characteristics	All Episodes*		First Episodes**	
	HHVBP	Comparison Group	HHVBP	Comparison Group
Number of Episodes	4,990,923	4,990,923	2,921,416	2,921,416
Average Age***	78.2	78.2	78.1	78.1
% Female***	63.1	63.1	62.6	62.6
Race/Ethnicity				
% White	78.1	73.8	78.7	78.3
% Black	9.9	15.0	9.2	11.9
% Hispanic	10.3	7.8	10.3	6.2
% Other	1.7	3.4	1.8	3.6
% Rural	5.2	9.1	4.3	7.8
% Dual Eligible	30.8	30.9	26.7	25.9
Chronic Conditions				
% Chronic Kidney Disease***	45.4	45.4	41.7	41.7
% Congestive Heart Failure***	43.6	43.6	38.3	38.3
% Diabetes***	46.7	46.7	42.8	42.8
% Pressure Ulcers and Chronic Ulcers***	24.7	24.7	17.2	17.2
% Alzheimer's Disease and Related Disorders or Senile Dementia	37.2	34.9	32.7	31.7
% Ischemic Heart Disease	58.9	57.8	55.9	54.6
% Anemia	62.4	58.7	59.9	57.1
Reason for Medicare Entitlement				
% Original ESRD***	1.0	1.0	0.8	0.8
% Original Disabled***	24.5	24.5	21.4	21.4
Current ESRD and Disabled Status				
% Current ESRD***	0.7	0.7	0.5	0.5
% Current Disabled***	11.2	11.2	9.9	9.9

Footnote: For each characteristic, percentages are calculated for the baseline period (2012-2015) from non-missing values (e.g., denominators differ across beneficiary characteristics).

*The denominator includes the sample of matched home health claim episodes eligible for the four claims-based outcomes that include all FFS HH episodes: unplanned hospitalization per all HH FFS episodes, SNF use per all HH FFS episodes, average Medicare spending for SNF per HH FFS episodes, average Medicare spending for unplanned ACH per HH FFS episodes.

**The denominator includes the sample of matched home health claim episodes that are 1st in the sequence and eligible for the two claims-based outcomes that include only first HH FFS episodes: ED Use (no Hospitalization) per first HH FFS episodes, unplanned hospitalization per first FFS HH episodes.

*** Characteristic was used in the matching procedure to define the comparison group, such that the exact balance between the intervention and comparison groups occurs by design. Note that we matched on age categories (vs. average age reported above).

Table 26. Comparison Group B: Beneficiary Characteristics for HHVBP and Comparison Group for OASIS-Based Outcome Measures

Characteristics	Ambulation		Bathing		Bed Transfer	
	HHVBP	Comparison Group	HHVBP	Comparison Group	HHVBP	Comparison Group
Number of Episodes	4,143,944	4,143,944	4,237,444	4,237,444	3,902,160	3,902,160
Average Age	77.9	77.2	77.8	77.0	78.0	77.3
% Female	62.9	63.2	62.7	63.1	63.0	63.4
Race/Ethnicity						
% White	77.4	75.5	77.4	75.5	77.1	75.3
% Black	9.4	13.9	9.4	13.8	9.4	14.0
% Hispanic	11.4	6.8	11.3	6.9	11.7	6.9
% Other	1.8	3.9	1.8	3.8	1.8	3.9
% Rural	3.9	6.3	3.9	6.4	3.8	6.3
% Dual Eligible	28.0	27.3	28.0	27.4	28.0	27.4
Chronic Conditions						
% Chronic Kidney Disease	35.1	33.1	34.9	33.0	35.2	33.3
% Congestive Heart Failure	32.8	33.7	32.6	33.6	32.9	34.0
% Diabetes	37.5	36.4	37.4	36.3	37.6	36.6
% Pressure Ulcers and Chronic Ulcers	15.2	14.4	15.0	14.2	15.4	14.7
% Alzheimer's Disease and Related Disorders or Senile Dementia	28.8	24.7	28.6	24.5	29.3	25.2
% Ischemic Heart Disease	48.0	44.8	47.9	44.7	48.3	45.0
% Anemia	49.7	44.5	49.5	44.4	50.0	44.8
Reason for Medicare Entitlement						
% Original ESRD	1.0	1.3	1.0	1.3	1.0	1.3
% Original Disabled	22.4	25.1	22.5	25.3	22.2	25.0
Current ESRD and Disabled Status						
% Current ESRD	0.6	0.8	0.6	0.9	0.6	0.8
% Current Disabled	10.4	12.3	10.6	12.5	10.2	12.1

Footnote: Denominator includes the sample of matched home health OASIS episodes of care eligible for each measure. For each characteristic, percentages are calculated for the baseline period (2012-2015) from non-missing values (e.g., denominators differ across beneficiary characteristics).

Table 27. Comparison Group B: Beneficiary Characteristics for HHVBP and Comparison Group for OASIS-Based Outcome Measures, Continued

Characteristics	Dyspnea HHVBP	Dyspnea Comparison Group	Oral Medications HHVBP	Oral Medications Comparison Group	Pain Management HHVBP	Pain Management Comparison Group
Number of Episodes	3,077,598	3,077,598	3,149,263	3,149,263	3,471,834	3,471,834
Average Age	78.0	77.1	78.9	78.4	77.0	76.3
% Female	62.3	62.5	61.7	61.9	64.1	64.3
Race/Ethnicity						
% White	76.7	74.1	76.6	74.8	76.2	74.7
% Black	9.0	14.7	9.4	13.8	9.0	14.2
% Hispanic	12.7	7.2	12.2	7.3	13.0	7.2
% Other	1.7	3.9	1.9	4.1	1.8	3.9
% Rural	4.1	6.9	3.9	6.1	4.2	6.7
% Dual Eligible	30.0	29.8	29.6	28.3	29.8	28.6
Chronic Conditions						
% Chronic Kidney Disease	37.4	35.5	36.9	35.1	33.8	32.0
% Congestive Heart Failure	37.2	38.3	35.1	36.1	31.3	32.4
% Diabetes	39.8	38.9	38.3	37.1	37.7	36.5
% Pressure Ulcers and Chronic Ulcers	15.3	14.4	15.7	15.4	14.8	13.9
% Alzheimer's Disease and Related Disorders or Senile Dementia	28.9	24.7	35.2	31.7	26.0	22.3
% Ischemic Heart Disease	52.0	48.6	49.7	46.8	47.6	44.3
% Anemia	51.1	45.8	50.5	45.3	49.7	44.6
Reason for Medicare Entitlement						
% Original ESRD	1.0	1.3	0.9	1.2	1.0	1.3
% Original Disabled	23.0	26.4	21.3	23.4	24.1	26.8
Current ESRD and Disabled Status						
% Current ESRD	0.6	0.9	0.6	0.8	0.7	0.9
% Current Disabled	10.4	12.7	9.4	10.8	11.8	13.6

Footnote: Denominator includes the sample of matched home health OASIS episodes of care eligible for each measure. For each characteristic, percentages are calculated for the baseline period (2012-2015) from non-missing values (e.g., denominators differ across beneficiary characteristics).

Table 28. Comparison Group B: Beneficiary Characteristics for HHVBP and Comparison Group for OASIS-Based Outcome Measures, Continued

Characteristics	Community Discharge	
	HHVBP	Comparison Group
Number of Episodes	5,871,337	5,871,337
Average Age	77.3	76.5
% Female	61.5	61.9
Race/Ethnicity		
% White	78.0	74.9
% Black	10.2	14.5
% Hispanic	10.0	7.1
% Other	1.8	3.6
% Rural	4.4	6.9
% Dual Eligible	30.0	29.5
Chronic Conditions		
% Chronic Kidney Disease	40.6	38.6
% Congestive Heart Failure	38.1	38.9
% Diabetes	40.1	39.3
% Pressure Ulcers and Chronic Ulcers	19.0	18.0
% Alzheimer's Disease and Related Disorders or Senile Dementia	30.7	26.7
% Ischemic Heart Disease	51.2	48.3
% Anemia	54.0	49.4
Reason for Medicare Entitlement		
% Original ESRD	1.4	1.8
% Original Disabled	24.9	27.5
Current ESRD and Disabled Status		
% Current ESRD	0.9	1.2
% Current Disabled	12.0	13.8

Footnote: Denominator includes the sample of matched home health OASIS episodes of care eligible for each measure. For each characteristic, percentages are calculated for the baseline period (2012-2015) from non-missing values (e.g., denominators differ across beneficiary characteristics).

Table 29. Comparison Group C: Beneficiary and HHA Characteristics for HHVBP and Comparison Group for OASIS-Based Process Measures and TPS

	HHVBP	Comparison Group
Number of HHAs	1,666	7,028
Beneficiary Characteristics*		
Average Age	77.6	76.2
% Female	61.7	61.9
Race/Ethnicity		
% White	71.7	69.6
% Black	9.6	16.1
% Hispanic	16.4	9.6
% Other	2.2	4.7
% Rural	8.1	11.3
% Dual Eligible	36.9	37.5
Chronic Conditions		
% Chronic Kidney Disease	46.4	46.5
% Congestive Heart Failure	39.4	43.0
% Diabetes	46.9	47.5
% Pressure Ulcers and Chronic Ulcers	20.4	19.2
% Alzheimer's Disease and Related Disorders or Senile Dementia	36.4	32.4
% Ischemic Heart Disease	56.1	55.5
% Anemia	59.5	56.1
Reason for Medicare Entitlement		
% Original ESRD	1.2	1.6
% Original Disabled	24.2	28.4
Current ESRD and Disabled Status		
% Current ESRD	0.7	0.9
% Current Disabled	11.4	14.6
HHA Characteristics		
Setting**		
Hospital-Based	9.8	9.8
Freestanding	90.2	90.2
Ownership**		
For-Profit	75.8	75.8
Non-Profit	17.3	17.3
Government-Owned	7.0	7.0
Agency Size**		
0-59 Beneficiaries	7.5	7.5
60-229 Beneficiaries	30.7	30.7
230 + Beneficiaries	61.8	61.8
Agency Age**		
Opened 2012-2015	7.6	7.6
Opened 2005-2011	35.7	35.8
Opened before 2005	56.7	56.6
% Chain**	34.0	34.0
2015 TPS/Quality Index-Quintiles**		
20th	17.1	17.1
40th	21.3	21.4
60th	20.5	20.5
80th	20.7	20.6
100th	20.5	20.5

* Beneficiary-level characteristics are weighted averages at the agency-level (weighted by the respective agency weights). For each characteristic, percentages are calculated for the baseline period (2015) from non-missing values (e.g., denominators differ across beneficiary characteristics).

** Agency characteristic was used in the reweighting procedure to define the comparison group such that the exact balance between the intervention and comparison groups occurs by design.

Table 30. Comparison Group C: Beneficiary and HHA Characteristics for HHVBP and Comparison Group for HHCAHPS Measures

	HHVBP	Comparison Group
Number of HHAs	1,443	5,989
Beneficiary Characteristics*		
Age		
% age <65	13.2	14.9
% age 65-74	25.2	26.4
% age 75-84	32.3	32.2
% age 85+	29.3	26.5
% Female	62.1	62.6
Race/Ethnicity (not mutually exclusive)		
% White	83.7	76.5
% Black	7.9	12.0
% Hispanic	12.0	7.9
% Other	2.3	4.4
% Dual Eligible	2.6	3.0
Overall Health		
% Excellent	5.5	4.9
% Very Good	16.7	15.7
% Good	29.6	29.9
% Fair	11.5	11.4
% Poor	11.1	11.8
% Current ESRD	1.2	1.6
HHA Characteristics		
Setting**		
Hospital-Based	11.2	11.2
Freestanding	88.8	88.8
Ownership**		
For-Profit	72.7	72.7
Non-Profit	19.9	19.9
Government Owned	7.4	7.4
Agency Size**		
0-59 Beneficiaries	4.2	4.2
60-286 Beneficiaries	33.2	33.2
287 + Beneficiaries	62.6	62.7
Agency Age**		
Opened 2012-2015	5.1	5.1
Opened 2005-2011	31.8	31.9
Opened before 2005	63.1	63.0
% Chain**	38.7	38.8
2015 TPS/Quality Index-Quintiles**		
20th	17.5	17.5
40th	21.0	21.1
60th	20.7	20.7
80th	20.2	20.2
100th	20.5	20.5

* Beneficiary-level characteristics are weighted averages at the agency-level (weighted by the respective agency weights). For each characteristic, percentages are calculated for the baseline period (2015) from non-missing values (e.g., denominators differ across beneficiary characteristics). We used HHCAHPS beneficiary-level data to generate the statistics on beneficiary characteristics. These data did not contain information that was available for other data sources (e.g., rural status, chronic conditions) and as such, are not included above.

** Agency characteristic was used in the reweighting procedure to define the comparison group such that the exact balance between the intervention and comparison groups occurs by design.

6. Glossary

Term	Definition
PEP	HH episodes subject to a Partial Episode Payment (PEP) are identified by patient discharge status code in the HH claims RIFs (PTNT_DSCHRG_STUS_CD) equal to 06.
LUPA	HH episodes subject to a Low Utilization Payment Adjustment (LUPA) are identified by the LUPA indicator variable (CLM_HHA_LUPA_IND_CD) in the HH claims RIFs.
Outlier	Outlier payment adjustments are made for HH episodes representing a relatively high utilization of HH services. Episodes subject to an outlier payment are identified by the presence of a claim line value code (CLM_VAL_CD) equal to 17 in the HH claims RIFs.
Normal	An HH episode is considered normal if it did not receive a PEP, LUPA, or Outlier payment adjustment.
FFS	A beneficiary is considered full FFS for a given month if they are enrolled in Medicare Parts A and B and are not receiving HMO coverage, based on MBSF monthly enrollment indicators.
Predicted Probabilities	Episode-level values indicating the probability that the episode is included in the measure numerator, based on the measure-specific risk adjustment model.
Sequence of Episodes	Multiple HH episodes for the same beneficiary in which the subsequent episode starts within 60 days of the previous episode end date are considered to be part of the same episode sequence.
Sequence Start Date	Date on which the first episode in a sequence of HH episodes starts.
Sequence End Date	Date on which the last episode in a sequence of HH episodes ends.
Episode Start Date	Equivalent to HH claim "from" date.
Episode End Date	Equivalent to HH claim "through" date, except in the case when claim "through" date occurs after the claim "from" date on an ensuing claim for the same beneficiary. In this case, HH episode end date will be equal to the day prior to the next episode start date.
Transfer HHAs within 60 Days	If a beneficiary has multiple HH episodes for different HHAs in which one episode starts within 60 days of the previous episode, the first episode is flagged to indicate a transfer of HHAs within 60 days. These flagged episodes are excluded from the denominator of several measures in this report.