Medicare Incentives, Payment Reform, and Quality in the Home Health Care Sector

by

Jun Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Service Organization and Policy)
in the University of Michigan
2020

Doctoral Committee:

Professor Edward C. Norton, Chair
Professor Charles C. Brown
Associate Professor Jeffrey S. McCullough
Professor Andrew M. Ryan

Jun Li

lijununi@umich.edu

ORCID iD: 0000-0002-7829-0752

DEDICATION

To my four-legged friends and the people who took the time to be kind.

ACKNOWLEDGEMENTS

My gratitude goes toward the many people who have inspired, mentored, and supported me. First, I thank Edward Norton, my advisor, for always encouraging my research and professional development, through thoughtful feedback, advocating for me, and generosity with his time. I am grateful for having met him and fortunate to have him be a part of my life.

My dissertation has also been made better through the mentorship from every member on my committee. Each professor challenged me to improve in a different way and their insights have been crucial to the development of my dissertation.

I also thank Rachel Werner. At the ideation stage of my dissertation, she encouraged me and gave me access to her data. Her kindness is an inspiration and made this dissertation possible.

I am grateful for my family. In particular, I thank my life partner, Bedrich Aquino, for his unwavering support for my professional endeavors; my parents for their unconditional love; and my grandmothers for inspiring me to value education.

Finally, I am thankful for the generous funding from the Rackham Pre-Doctoral Fellowship from the University of Michigan and the Agency for Healthcare Research and Quality (1R36HS026836).

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES x

LIST OF APPENDICES xii

ABSTRACT xiii

CHAPTER 1 Value-Based Payments in Health Care: Evidence from a Nationwide Randomized Experiment in the Home Health Sector 1

1.1 Introduction 1
1.2 Background 4
1.3 Measures and data 12
1.4 Empirical Strategy 23
1.5 Results 27
1.6 Robustness checks and sensitivity analyses 33
1.7 Conclusion 40

CHAPTER 2 Medicare's Home Health Star Ratings Program Has Not Shifted Patients into Higher-Rated Agencies 44

2.1 Introduction 44
2.2 Methods 45
2.3 Results 52
2.4 Discussion 59

CHAPTER 3 Choosing Home Health Care on the Basis of Medicare Star Ratings: Is There Value in Picking the Best? 63

3.1 Introduction 63
3.2 Methods 64
3.3 Results 72
3.4 Discussion 83

APPENDICES 85

Appendix A 86
Appendix B 115
Appendix C 128

BIBLIOGRAPHY 134

LIST OF TABLES

TABLES

1.1 Quality performance outcomes targeted by year 1 (2016) of the Home Health Value-Based Purchasing Program. 8

1.2 Comparison of 2015 characteristics between treatment and control agencies under the Home Health Value-Based Purchasing program. 16

2.1 Characteristics of sample by star ratings and becoming the highest-ranked star option within across ZIP codes. 53

2.2 Regression discontinuity estimates of the effects of gaining a half star on patient admissions. 55

2.3 Balance tests across the star rating threshold. 57

2.4 Effects of becoming the highest-ranked option within a ZIP code on number of new patients. 59

3.1 Characteristics of patients using home health care in study cohort. 74

3.2 Differences in outcomes after treatment by a top agency versus lower-rated agency. 76

3.3 Differences in outcomes after treatment by a top agency versus lower-rated agency across subgroups. 80

3.4 Characteristics of complier population compared with patients in the full study sample. 82

A.1 Marginal incentive to improve on agency-reported measures targeted by year 1 of the Home Health Value-Based Purchasing program. 87

A.2 Expected marginal gains from improvement for agency-reported measures targeted by year 1 of the Home Health Value-Based Purchasing program. 88

A.3 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on agency-reported measures using a difference-in-differences design. 89

A.4 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on non-agency-reported measures using a difference-in-differences design. 90

A.5 Effect of marginal incentive size on agency-reported measures in year 1 of the Home Health Value-Based Purchasing Program using a difference-in-difference-in-differences design. 91

A.6 Effect of marginal incentive size on non-agency-reported measures in year 1 of the Home Health Value-Based Purchasing Program using a difference-in-difference-in-differences design. 92

A.7 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on coding manipulation measures using a difference-in-differences design. 93

A.8 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on health care utilization quality measures not a part of the program using a difference-in-differences design. 94

A.9 Sensitivity analyses on the aggregate treatment effects of the Home Health Value-Based Purchasing Program in year 1 using a difference-in-differences design. 95

A.10 Sensitivity analyses of effects of marginal incentive size on performance in the Home Health Value-Based Purchasing Program in year 1 using a difference-in-difference-in-differences design. 96

A.11 Sensitivity analysis of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on coding manipulation using a difference-in-differences design. 97

A.12 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on percentage of discharges to inpatient institutions (non-targeted measure) using a difference-in-differences design, stratified by patient type. 98

A.13 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on percentage of episodes followed by additional home health care (non-targeted measure) using a difference-in-differences design, stratified by patient type. 99

A.14 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on agency-reported outcome and agency-reported process measures using a cross-sectional design across total sample. 100

A.15 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on patient-survey and administrative-claims measures using a cross-sectional design across total sample. 101

A.16 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on agency-reported outcome and process measures using a cross-sectional design across propensity-score matched samples. 102

A.17 Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on patient-survey and administrative-claims measures using a cross-sectional design across propensity-score matched samples. 103

A.18 Effect of marginal incentive size on agency-reported outcome measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples. 104

A.19 Effect of marginal incentive size on agency-reported process measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples. 105

A.20 Effect of marginal incentive size on patient-survey measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples. 106

A.21 Effect of marginal incentive size on administrative-claims measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples. 107

A.22 Effect of marginal incentive size on agency-reported outcome measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample. 108

A.23 Effect of marginal incentive size on agency-reported process measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample. 109

A.24 Effect of marginal incentive size on patient-survey measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample. 110

A.25 Effect of marginal incentive size on administrative-claims measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample. 111

B.1 Star rating release dates and days included in analysis. 118

B.2 Measures included in the first six quarters of the home health star ratings program. 118

B.3 Functional form assessment of regression discontinuity design for all thresholds, 1 vs. 1.5 stars, and 1.5 vs. 2 stars. 119

B.4 Functional form assessment of regression discontinuity design 2 vs. 2.5 stars, 2.5 vs. 3 stars, and 3 vs. 3.5 stars. 120

B.5 Functional form assessment of regression discontinuity design 3.5 vs. 4 stars, 4 vs. 4.5 stars, and 4.5 vs. 5 stars. 121

B.6 Effects of removing outermost observations on regression discontinuity estimates. 122

B.7 Regression discontinuity estimates from a negative binomial 2 model. 123

B.8 Assessment of evidence of violation of parallel pre-trends. 124

B.9 Effects of becoming highest-ranked option within a ZIP code using a negative binomial-2 model. 124

B.10 Association of star ratings and number of patients using naïve linear regression. 125

C.1 Characteristics of ZIP codes by instrumental variable values. 129

C.2 Characteristics of home health agencies serving ZIP codes in the sample. 130

C.3 First-stage results using dichotomized instrumental variable. 131

C.4 Population characteristics across ZIP codes in the sample. 132

C.5 Alternative specifications of the first-stage regression using linear and linear spline differential distance as instruments. 133

LIST OF FIGURES

FIGURES

1.1 Treatment and control states under the Home Health Value-Based Purchasing program. 6

1.2 Overview of performance scoring rules in the Home Health Value-Based Purchasing program. 9

1.3 Home health agency eligibility and randomization in the Home Health Value-Based Purchasing program. 14

1.4 Binned scatter plot of marginal incentives to improve on four example measures. 20

1.5 Effect of the Home Health Value-Based Purchasing program, with units in standard deviations of the control group, on targeted quality measures. 28

1.6 Effects of the Home Health Value-Based Purchasing program by incentive size, with units in standard deviations of the control group, on targeted quality measures. 29

1.7 Effect of the Home Health Value-Based Purchasing program on coding manipulation measures. 31

1.8 Effect of the Home Health Value-Based Purchasing program on non-targeted quality measures. 32

1.9 Event study of the effect of the Home Health Value-Based Purchasing program on agency-reported targeted quality measures. 34

1.10 Event study of the effect of the Home Health Value-Based Purchasing program on non-agency-reported targeted quality measures. 35

1.11 Event study of the effect of the Home Health Value-Based Purchasing program on coding manipulation. 35

1.12 Event study of the effect of the Home Health Value-Based Purchasing program on non-targeted quality measures. 36

2.1 Relationship between home health star ratings and number of new patients per quarter pooling all thresholds. 49

A.1 Event studies of the effect of the Home Health Value-Based Purchasing model on the average hierarchical condition category risk score. 112

A.2 Event studies of the functional deficit level at start of care across agencies using a weighted regression. 113

A.3 Event studies of the change in functional deficit level between initial discharge and start of care at readmission across agencies using a weighted regression. 114

B.1 Relationship between unrounded star ratings and star ratings in the first six quarters of the Home Health Star Ratings program. 126

B.2 Distribution of unrounded star ratings in the first six quarters of the Home Health Star Ratings program. 127

LIST OF APPENDICES

APPENDIX

A. Appendix to Value-Based Payments in Health Care: Evidence from a Nationwide Randomized Experiment in the Home Health Sector 86

B. Appendix to Medicare's Home Health Star Ratings Program Has Not Shifted Patients into Higher-Rated Agencies 115

C. Appendix to Choosing Home Health Care on the Basis of Medicare Star Ratings: Is There Value in Picking the Best? 128

ABSTRACT

Federal policymakers have implemented two types of health care reforms in recent years to address poor quality of US health care. The first type uses payment incentives targeting health care providers and the second focuses on information transparency targeting consumers. Reforming home health care is particularly important. The quality of care delivered to the 3.4 million Medicare home health patients each year is highly variable. These patients have complex health needs, are aged, and have severely limited independence. Moreover, home health is the fastest growing health care sector in the US. Therefore, understanding the efficacy of mainstay reforms is paramount to guide policymakers toward the optimal policy for this population.

In Chapter one, I examine the effects of the Home Health Value-Based Purchasing program on home health quality. In this provider-directed incentive program, Medicare randomly selected a nationally representative group of agencies to compete on quality for financial rewards. The goal of the program was to achieve better home health care quality. Home health agencies in treatment states were rewarded or penalized based on their performance on agency-reported and non-agency-reported quality measures. The program improved agency-reported measures by approximately one percentage point, and performance gains follow a dose-response relationship with respect to incentive size. However, the performance gains in agency-reported measures did not reflect true quality improvement. I find that agencies manipulated their coding of patients and inflated their performance. Coding manipulation explained the entirety of the program's impact on agency-reported measures.

In Chapter two, I study the Home Health Star Ratings program, which provides public information to facilitate consumer decision-making. Like other quality disclosure programs, this program was designed to rectify insufficient competition on quality in the market. The assumption is that when patients lack information on quality when choosing providers, providers face little incentive to improve and compete on quality. Thus, the first-order goal of the program was to shift patients toward higher-rated home health agencies. I use a regression discontinuity design to determine whether agencies that received one more half-star obtained more new patients in the quarter following each star rating release. To further determine whether the star ratings had an effect on patients' choice of agencies, I used a difference-in-differences design to determine whether becoming one of the highest star-rated options in a ZIP code led to more patient volume. I find no evidence that receiving a higher star rating under the program meaningfully increased patient demand for higher-rated agencies.

In the final chapter, I evaluate whether the Home Health Star Ratings program contain meaningful information on quality for patients. I test whether receiving treatment from an agency with the highest number of stars in a patient's residential area result in better health outcomes. I use an instrumental variable approach to address the endogeneity of agency choice. Specifically, I exploit variation over time in the differential distance between the patient's residential ZIP code and the nearest top-rated agency and the closest lower-rated agency. I measure health as the number of days a patient spends alive and without use of additional health care in the 180 days following the end of home health care. I find that overall, patients treated by the top agency in their market spent 2.6 percent more days alive without use of health care, suggesting that the home health star ratings contain valid and meaningful information.

CHAPTER 1

Value-Based Payments in Health Care: Evidence from a Nationwide Randomized Experiment in the Home Health Sector

1.1 Introduction

Policymakers' use of value-based payments, also known as pay-for-performance, has increased rapidly to address both the high spending growth and low quality of health care. Value-based payment programs provide financial rewards to incentivize improvements in the quality and/or cost of health care provided to patients. Despite becoming a mainstay policy approach, studies examining these programs have reported minimal effects on quality (Rosenthal and Frank 2006; Ryan et al. 2016; Scott, Liu, and Yong 2018; Doran, Maurer, and Ryan 2017). These disappointing findings have led some to argue that these types of designs are too complex or that incentives are too small to elicit responses from providers (Figueroa et al. 2016; Ryan et al. 2017). However, value-based payment programs are typically implemented nationally at one time, leaving few options for comparable control groups. As a result, studies either do not include control groups or use control groups that differ in important ways from the treated groups (Christianson, Leatherman, and Sutherland 2008). Thus, it remains unclear whether the limited impact previously observed was due to programs' complexity, small incentives, or methodological limitations.

The Home Health Value-Based Purchasing (HHVBP) program is the first value-based payment program designed as a nationally representative experiment, presenting a unique opportunity to assess whether tying payment to performance is a viable policy intervention to improve health care quality. It focuses on home health care, which provides services to 4.5 million patients within their homes each year to help patients recover from illness and injury and to assist patients in their abilities to remain in the community (Ellenbecker et al. 2008; National Center for Health Statistics 2019). The Centers for Medicare & Medicaid Services (CMS) randomly selected nine states and assigned all eligible home health agencies in those states into the HHVBP program. For the first year, an agency under the program could have its 2018 Medicare reimbursement adjusted by a maximum of 3% upwards or downwards based on its performance in 2016. Performance was determined by both agency-reported and non-agency-reported quality measures. Agency-reported measures included outcome measures (whether patients improved their functional abilities between start of care and end of care) and process measures (whether agencies adhered to clinical guidelines); non-agency-reported measures included patient-reported experience and administrative-claims health care utilization measures. The explicit goal of the program is to achieve better quality, by rewarding better outcomes rather than solely volume of services (US Department of Health and Human Services, n.d.).

The only study that has examined the HHVBP program was an evaluation sponsored by CMS (Arbor Research Collaborative for Health and L&M Policy Research 2018). The evaluators found that agencies participating in the program had improved performance on several of the agency-reported measures, but not on the other measures targeted by the program. While the CMS evaluation estimated the overall effect of the program on targeted measures, they did not examine the relationship between incentive size and performance. This aspect is important as prior literature has found tremendous variation in incentive size under the precursor program, the Hospital Value-Based Purchasing program, that served as the blueprint for the HHVBP. In the Hospital Value-Based Purchasing program, the reward structure resulted in approximately one third of all hospitals facing no incentives to improve and incentives were heterogeneous (Norton et al. 2018). Thus, to ascertain that financial incentives drive improvement under the HHVBP, it is imperative to understand the dose-response relationship between incentives and performance.

Perhaps the most important gap left unexplored by the CMS evaluation is whether the improvements in agency-reported quality measures were in fact true improvements. Because the performance gains were limited to agency-reported measures, inaccurate documentation is a plausible alternative mechanism for agencies to improve their performance under the program without changing care delivery. In particular, the HHVBP's agency-reported outcome measures capture the proportion of patients with improved functional abilities between the start of care and end of care. Therefore, agencies could inflate their performance by exaggerating, or *up-coding*, patients' functional deficits at the start of care. The manipulation of coding rather than true quality improvement runs counter to CMS' goal to reward better quality of care. Further, coding manipulation misrepresents providers' performance and opportunities for coding manipulation erode agencies' incentives to invest in underlying quality (Fuloria and Zenios 2001). If CMS rewards agencies that manipulate their coding, then agencies that invest in true quality improvement may be unfairly penalized.

I conduct the first independent analysis of the program. First, I evaluate the HHVBP's effects on targeted outcomes. Second, to determine whether there is evidence that larger incentives lead to larger responses, I estimate the incentives faced by agencies under the program and examine whether the size of incentives affected agencies' performance. Third, I assess the extent to which the program affected underlying quality using a dual approach: I test whether agencies inflated their agency-reported measure performance through inaccurate reporting and whether the HHVBP reduced patients' use of subsequent health care services.

To test how incentive size affects quality performance, I use a non-parametric approach to estimate the incentives faced by each agency for each quality measure targeted by the program, or the *marginal incentives to improve*, following Norton, Li, Das, and Chen (2018). I find that the incentives were non-linear and heterogeneous across measures and agencies. The standard difference-in-differences model that accounts for changes over time and differences between treatment and control agencies is extended to include differences in the marginal incentives to improve—making the analysis a triple differences design.

I find evidence that the HHVBP program led to better performance on agency-reported measures, and that this effect was driven by agencies that faced larger marginal incentives to improve. Agencies under the program improved their performance across agency-reported measures by approximately one percentage point from an average baseline performance score of 68 out of 100 percent. For the average home health agency, a one percentage-point increase in agency-reported outcomes translates to four more patients out of approximately 400 patients per agency with better agency-reported quality.

To determine whether the program affected underlying quality, I examine for evidence that agencies engaged in coding manipulation to inflate their performance on the agency-reported measures. I assess the evidence of coding manipulation in two ways. In the first, I examine the coded functional deficit levels of patients at the start of home health care. In the second, I examine within-patient changes in coded function, focusing on patients who were discharged from home health care and then readmitted for a subsequent episode of home health care within one day. Because the transition occurs within one day, it is unlikely that the true functional abilities of the patients would change. There is also little reason for patients treated by HHVBP agencies to have more deficits following this transition than non-HHVBP agencies, absent of coding practice changes.

Agencies under the HHVBP program increased the coded functional deficit levels of their patients. There was a differential increase of 0.02 functional deficit levels among patients starting care in HHVBP agencies than those starting care in non-HHVBP agencies. The effects of the program on coded functional deficit levels persisted even after controlling for patient functional ability. Patients readmitted to an HHVBP agency had an increase of 0.07 deficit levels within one day of being discharged from a prior home health care episode, compared to patients not readmitted to an HHVBP agency.

To further consider whether the program led to better underlying quality, I assess whether the program improved patient outcomes across two quality measures that are plausibly related to improved quality targeted but are not rewarded under the program (Tao and Ellenbecker 2013; C.-L. Li et al. 2011; Luppa et al. 2010). I test whether patients used less subsequent home health care and whether they used less inpatient institutional care (i.e., inpatient rehabilitation facilities, skilled nursing facilities, hospitals, and hospice) after discharge from home health. Consistent with the notion that there was no improvement in true quality, I find no statistically significant effects of the HHVBP program on either measure.

A back of the envelope calculation based on my estimates of the impact of the HHVBP on targeted measures and coding practice change suggest that up-coding of functional deficit levels explains the entirety of the performance gains in agency-reported outcome measures. Together, this study is the first to provide evidence that agencies used coding manipulation rather than improving true quality to respond to the incentives under the HHVBP program. My results examining year one of the HHVBP program do not indicate that it is the right policy solution to improve patient welfare.

1.2 Background

1.2.1 Home Health Value-Based Purchasing program

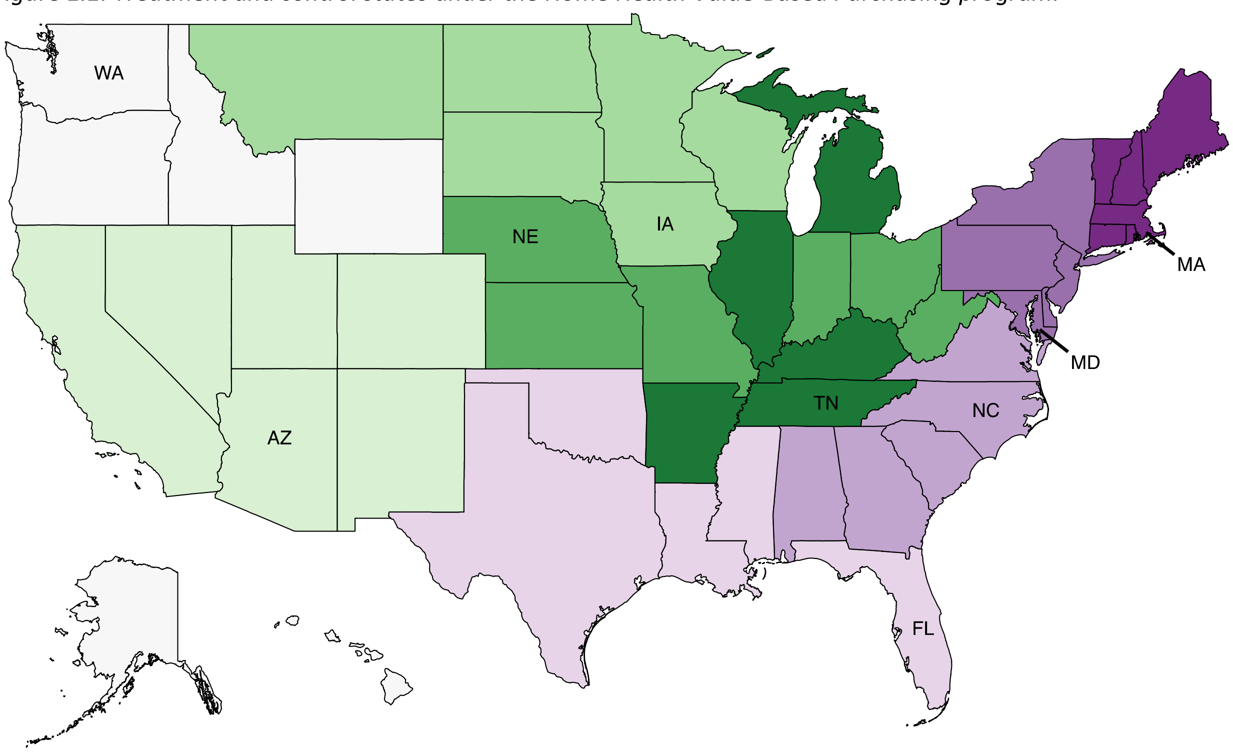
Value-based payments in home health care is based on the economic concept that linking payments to performance can align the incentives of the care provider, home health agencies, with the objectives of the payer, Medicare. Medicare, as the primary payer of home health care services and as social planner, wishes to maximize quality for a given level of spending, while home health agencies may wish to maximize profit. However, because quality is difficult to observe by patients, market forces do not work to improve quality, resulting in sub-optimal quality (J. Li and Norton 2019; Weisbrod 1991; Golden and Sloan 2008). As an intervention, the HHVBP uses financial rewards to incentivize agencies to achieve better quality in home health care.

Finding viable policy solutions to improve the quality of care in the home health sector is important for several reasons. There is substantial variation in home health quality and a growing concern that the existing volume-based reimbursement system does not provide sufficient incentives for home health agencies to provide high quality care (Medicare Payment Advisory Commission 2018; US Department of Health and Human Services, n.d.). Patients using home health care also tend to be disproportionately vulnerable, both medically and socially (Avalere 2015; Wolff et al. 2008). Thus, addressing quality deficits in this sector may alleviate health care disparities. Finally, home health care is the fastest growing health care sector in the US, making it even more urgent to ensure that the rising patient population is provided high quality care (Cuckler et al. 2018).

However, the evidence on whether value-based payments is effective in improving quality is mixed (Rosenthal and Frank 2006; Ryan et al. 2016; Scott, Liu, and Yong 2018; Doran, Maurer, and Ryan 2017; Institute of Medicine 2007). Unfortunately, value-based payment programs have generally been plagued by designs that have made rigorous evaluation difficult, low monetary rewards for providers, or relatively few providers exposed to the incentives (Rosenthal and Frank 2006; Richardson 2012; Christianson, Leatherman, and Sutherland 2008; Werner, Konetzka, and Polsky 2013). In contrast, Medicare's HHVBP uses randomization that facilitates evaluation, provides relatively large monetary rewards, and employs a reward structure that is designed to incentivize a broad swath of the provider population.

In November 2015, CMS selected nine states to participate in the HHVBP with the intent of creating a nationally representative treatment sample (Centers for Medicare & Medicaid Services, 2015). First, CMS divided states into nine regions based on a variety of characteristics, including geographic proximity, home health utilization rates, ownership model of agencies, proportion of home health users that are Medicare-Medicaid dually eligible beneficiaries, and average episodes of care per agency. Next, CMS used a random number generator to select one state within each of the nine regions to in the program. Within each region, each state had an equal probability of being selected. The program started on January 2016 and will last through December 2020 for Massachusetts, Maryland, North Carolina, Florida, Washington, Arizona, Iowa, Nebraska, and Tennessee (Figure 1.1). This study examines performance in the first year of the program.

Figure 1.1: Treatment and control states under the Home Health Value-Based Purchasing program.



Notes: Treatment states include Massachusetts, Maryland, North Carolina, Florida, Washington, Arizona, Iowa, Nebraska, and Tennessee. The treatment states were randomly selected with equal probability from each of the nine regions. These regions are not census regions and were identified by CMS based on a number of characteristics. Massachusetts (MA) was selected from a region also consisting of Vermont, Maine, Connecticut, Rhode Island, and New Hampshire; Maryland (MD) was selected from a region also consisting of Delaware, New Jersey, Pennsylvania, and New York; North Carolina (NC) was selected from a region also consisting of Alabama, Georgia, South Carolina, and Virginia; Florida (FL) was selected from a region also consisting of Texas, Oklahoma, Louisiana, Mississippi; Tennessee (TN) was selected from a region also consisting of Illinois, Kentucky, Arkansas, and Michigan; Iowa (IA) was selected from a region also consisting of North Dakota, South Dakota, Montana, Wisconsin, and Minnesota; Nebraska (NE) was selected from a region also consisting of Ohio, West Virginia, Indiana, Missouri, and Kansas; Arizona (AZ) was selected from a region also consisting of New Mexico, California, Nevada, Utah, and Colorado; Washing (WA) was selected from a region also consisting of Oregon, Alaska, Hawaii, Wyoming, Idaho

Agencies that met the program's inclusion criteria within the nine treatment states were mandatory participants. There were two inclusion criteria. First, only agencies that were Medicare certified in the baseline year (2015) and in the performance year (2016) were eligible for the program's first performance year. Second, agencies were only included if they had at least five measures that met minimum case requirements (e.g., 20 episodes for quality measures) in both the baseline and performance periods. CMS imposed a minimum number of cases to improve the reliability of the measures. Smaller and newer agencies that did not meet these inclusion criteria did not compete for rewards.

Among the measures targeted by the HHVBP program in its first performance year, 17 measures were assessed for achievement and improvement and three measures were assessed for submitting information to CMS (Table 1.1); in subsequent years of the program, these numbers changed. The measures assessed for achievement and improvement included a combination of agency-reported and non-agency-reported quality measures. To determine rewards, performance on each targeted measure was assigned a score (Figure 1.2) and then summed to obtain a Total Performance Score. Agencies were compared within each state to arrive at the score. A higher Total Performance Score translated linearly to a higher financial reward in terms of Medicare reimbursement adjustment. For performance in 2016, an agency in 2018 could have its total reimbursement adjusted by a maximum of 3% upwards or downwards. Thus, the range of reimbursement adjustment between the highest and lowest performing agencies was 6%. To put the payment incentives into perspective, 3% of Medicare reimbursement for the average home health agency in the US in 2016 was approximately $55,041 per year or $91 per patient (Centers for Medicare & Medicaid Services 2018a). Because Medicare pays for the largest share of home health care, a 3% adjustment is a substantial amount for most home health agencies (MedPac 2018). Moreover, by the last year of the five-year program, the adjustment will increase to 8% of Medicare reimbursement, with a range of 16% between the highest and lowest performing agencies.

Table 0.1: Quality performance outcomes targeted by year 1 (2016) of the Home Health Value-Based Purchasing Program.

Targeted quality outcomes examined in analysis
Agency-reported outcome measures (reported by agencies via OASIS)
Improvement in ambulation-locomotion
Improvement in bed transferring
Improvement in bathing
Improvement in dyspnea
Improvement in pain interfering with activity
Improvement in management of oral medications
Agency-reported process measures (reported by agencies via OASIS)
Pneumococcal polysaccharide vaccine ever received
Influenza immunization received for current flu season
Drug education on all medications provided to patient/caregiver
Patient-survey measures (calculated by Medicare-approved survey vendors)
How often the home health team gave care in a professional way
How well did the home health team communicate with patients
Did the home health team discuss medicines, pain, and home safety with patients
How do patients rate the overall care from the home health agency
Would patients recommend the home health agency to friends and family
Administrative-claims measures (calculated by CMS)
No unplanned hospitalization during the first 60 days of home health
No emergency department use during the first 60 days of home health

Notes: OASIS = Outcome and Assessment Information Set; CMS = Centers for Medicare & Medicaid Services. Performance-based measures are weighted equally

Figure 1.2: Overview of performance scoring rules in the Home Health Value-Based Purchasing program.

Measure Score Formula	Measure Score
Performance in 2016 ≤ Baseline in 2015 **AND** Performance in 2016 < Median of all agencies' performance in 2015	0 points
HIGHER OF **Improvement Points**: $10\times\left(\frac{Performance\ in\ 2016\ -\ Baseline\ in\ 2015}{mean\ of\ top\ decile\ of\ all\ HHAs'\ performance - Baseline\ in\ 2015}\right) - 0.5$ **OR** ***Achievement Points***: $10\times\left(\frac{Performance\ in\ 2016\ -\ Median\ of\ all\ agencies'\ performance\ in\ 2015}{mean\ of\ top\ decile\ of\ all\ agencies'\ performance - Median\ of\ all\ agencies'\ performance\ in\ 2015}\right) - 0.5$	1–9 points
Performance in 2016 ≥ Mean of the top decile of all agencies' performance in 2015	10 points

Notes: Each targeted measure receives a measure score, based on improvement points and achievement points. The measure scores are then summed to determine the Total Performance Score, which is then translated to a reward amount (bonus or penalty) using a linear conversion factor.

The program's financial reward structure also means that the program's incentives among agencies are heterogeneous. Financial rewards are determined by how much an agency has improved against its own baseline for each measure, in addition to how an agency compared to its peers for each measure (within-state competitors) in the performance year. In other words, for each measure, an agency's reward for a given unit of improvement varies by how well it performs in the baseline year (e.g., 2015) and how well its competitors perform in the subsequent year (e.g., 2016).

The purpose of the combined improvement and achievement scoring is to provide incentives to a larger proportion of the home health agencies compared to a system where only improvement or achievement is rewarded. The added benefit of this combination scoring system is that it deters strategic underperforming in one year to inflate an agency's improvement in another year. In other words, an agency could increase its improvement points by performing particularly badly in one year, but this behavior would be penalized through lower achievement points.

The theory behind value-based purchasing programs predicts that agencies under the program will have better performance than non-HHVBP agencies. Moreover, the heterogeneous incentives embedded in the program suggest that agencies facing larger incentives at the margin ought to perform better than agencies facing smaller incentives. These priors motivate examining the aggregate effects of the program as well as the relationship between incentive size and performance.

1.2.2 Perverse incentives in Medicare's existing reimbursement system

In addition to the incentives offered by the HHVBP, whether an agency has an incentive to improve on a particular measure also depends on the other incentives that it faces, including the existing incentives in Medicare reimbursement. However, not all of the other incentives are aligned with those in the HHVBP to improve care quality in the ways intended by the program. In particular, these interactions are likely to affect agencies' performance on the agency-reported and non-agency-reported measures (i.e., hospitalizations and emergency department (ED) visits during the home health episode) targeted by the program.

First, one must first consider how Medicare reimburses home health care to understand how perverse incentives in Medicare's reimbursement system can undermine the quality improvement incentives under the HHVBP. Under the Home Health Prospective Payment System, home health agencies are paid a fixed amount for each 60-day episode of care, conditional on a given patient's expected costs (Centers for Medicare & Medicaid Services 2015c). Information to calculate the expected costs of each patient is collected by home health agencies at the start of care using a standardized assessment known as the Home Health Outcome and Assessment Information Set (OASIS) (O'Connor and Davitt 2012; Centers for Medicare & Medicaid Services 2015b). Due to anticipated higher costs, agencies serving patients with more agency-reported functional deficits in their abilities at start of care to ambulate, bathe, or transfer from bed, are reimbursed at higher levels (Centers for Medicare & Medicaid Services 2015c). Therefore, existing incentives motivate agencies to exaggerate their patients' functional deficit levels as reported in the OASIS at the start of each episode of care.

Under performance year one of the HHVBP, three of the agency-reported measures examined the proportion of each agency's patients that improved their abilities to ambulate, bathe, and to transfer from bed between the start of care and end of care. As such, the HHVBP further reinforced an agency's incentive to up-code their patients' functional deficit levels at start of care; an agency could simultaneously raise their Medicare reimbursement rates and rewards under the HHVBP by coding their patients as initially having more functional deficits (Medicare Payment Advisory Commission 2019a). Thus, all agencies have an incentive to up-code the functional deficits of patients at start of care, with the HHVBP compounding this perverse incentive.

While the existence of coding manipulation has yet to be documented in the HHVBP prior to this study, coding manipulation has been found in other value-based payment programs. Bastani, Goh, and Bayati (2019) investigated hospitals' response to penalties tied to hospital-acquired infections in Medicare's Hospital-Acquired Condition Reduction Program. Like the measures in the HHVBP, the hospital-acquired infections were self-reported by hospitals. The authors found that hospitals operating in states with more stringent reporting regulations were associated with lower instances of coding manipulation. In another study, investigators found that hospitals' coded patient severity of illness for conditions targeted by the Hospital Readmissions Reduction program increased more relative to non-targeted conditions, indicating that hospitals strategically used coding to boost their performance (Sukul et al. 2019). Although examining different care sectors, these studies indicate that coding manipulation is one plausible undesirable response to value-based payment incentives.

The second set of perverse incentives pertains to the effect of the HHVBP program on agencies' performance on the non-agency-reported, administrative-claims quality measures. Year one of the HHVBP provided incentives for agencies to decrease hospitalizations and ED visits during the home health episode. However, under the standard payment system, hospitalizations and ED used while under the home health agency's care are reimbursed to the agency by Medicare separately. Furthermore, services provided by the hospital may even substitute for services that the agency otherwise would have to provide (David, Rawley, and Polsky 2013). Thus, the benefit to agencies of improving on these measures may be limited, even with possibility of earning rewards under the HHVBP. Compounding this disincentive is that an intervening hospitalization or ED visit could provide the necessary justification for the agency to request physician certification for additional home health care, which is revenue increasing for agencies (Centers for Medicare & Medicaid Services 2019a; Konetzka, Stuart, and Werner 2018). Therefore, while the HHVBP generated incentives for home health agencies to reduce patients' hospitalizations and ED visits, the program was likely insufficient to overcome the existing incentives in reimbursement to their overuse. Thus, one would expect that the agencies under the HHVBP program respond differently on the non-agency-reported administrative-claims measures than the agency-reported measures.

In summary, while the theory behind value-based payment predicts that agencies under the program would have better performance than non-HHVBP agencies, and that the effects are likely to be greater among agencies facing larger marginal incentives, the existing incentives in Medicare reimbursement affect how agencies respond to the various measures targeted by the HHVBP program. Specifically, one would expect that HHVBP agencies are more likely to exaggerate a patient's functional deficit levels through up-coding than non-HHVBP agencies. Further, the HHVBP program likely impart a different effect on the non-agency-reported measures of hospitalization and ED use than the agency-reported measures, due to countervailing incentives in existing Medicare reimbursement policies.

1.3 Measures and data

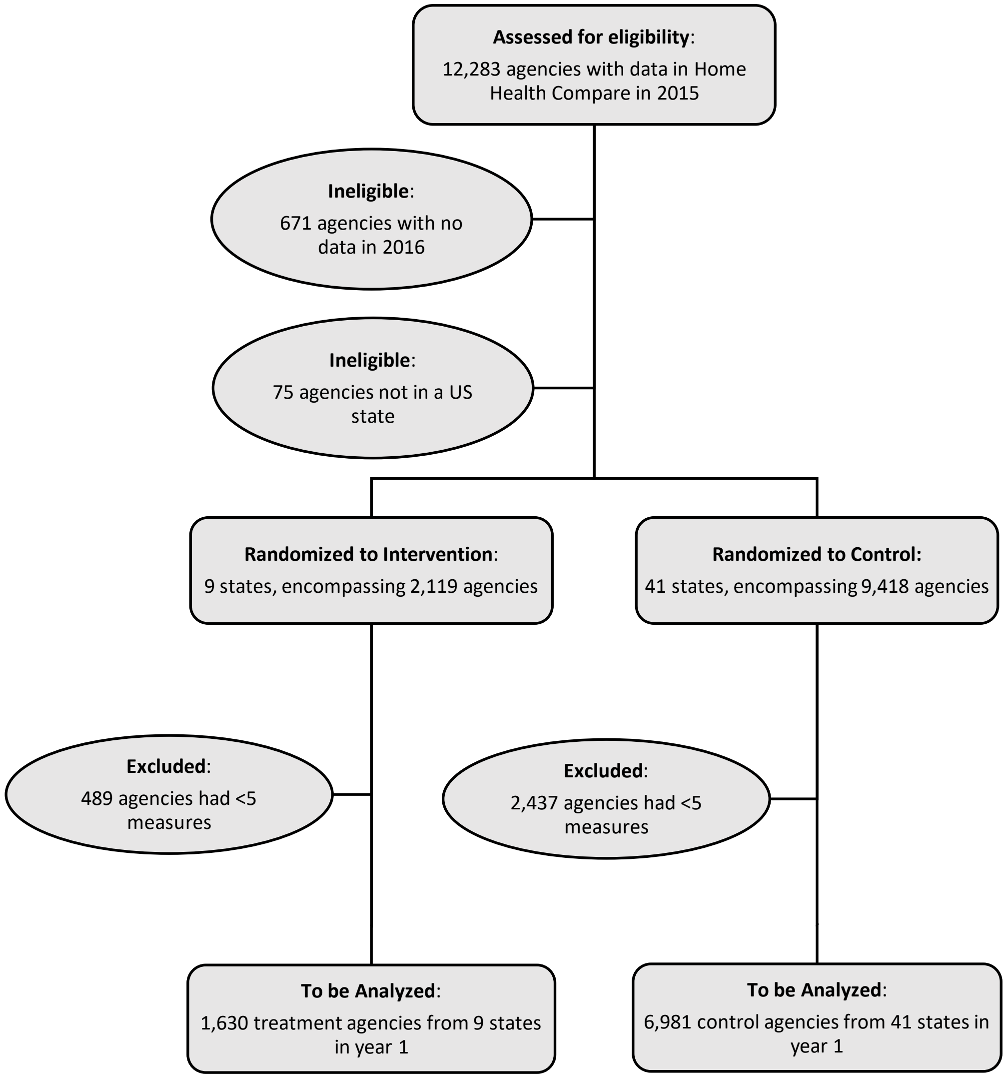
This study examines whether the HHVBP program led to better quality in the first performance year in 2016 by comparing HHVBP agencies to non-HHVBP agencies, between 2015 and 2016. Several hypotheses are tested using a variety of measures and data. To assess the HHVBP's aggregate effects on targeted performance outcomes, I use agency-level quality data on each of the targeted measures of the program. Agency-level, measure-specific quality performance data also provide the necessary information to calculate marginal incentives for improvement for each agency and measure, which enables me to examine the relationship between the size of incentives and agencies' quality performance for each targeted measure. Finally, to assess the extent that the program affected true quality, I use patient-level data to examine whether agencies up-coded patients' functional deficit levels on average and whether agencies improved on non-targeted measures of agency-level quality.

1.3.1 Study sample

The main sample in this study consist of the agencies eligible for the program in 2016. Agencies operating within each of the nine randomly selected treatment states were eligible to be in the treatment group. I replicate CMS' HHVBP eligibility criteria using 2015 and 2016 data from Home Health Compare. Home Health Compare provides quality performance information for the universe of Medicare-certified home health agencies in the US.

There were 12,283 agencies with data in Home Health Compare for 2015, of which 671 no longer had data in Home Health Compare for 2016 (Figure 1.3). Using the same inclusion criteria for both participating and non-participating agencies, I exclude an additional 75 agencies from the analyses because they had no chance of being assigned to treatment (i.e., Washington, D.C., Puerto Rico, Guam, Northern Mariana Islands, US Virgin Islands). Of the 11,537 agencies remaining, 8,611 agencies had at least five measures with that met CMS' minimum case requirements for eligibility in both 2015 and 2016 (e.g., 20 cases for OASIS/claims quality measures and 40 patient experience surveys). This yields 1,630 agencies in the treatment states and 6,981 agencies in the control states.

Figure 1.3: Home health agency eligibility and randomization in the Home Health Value-Based Purchasing program.



While randomization guarantees independence between treatment and covariates in expectation, randomization involving a modest sample of 50 states, may not resolve selection bias (Deaton & Cartwright, 2018). To assess the comparability of treatment and control agencies after randomization, I compare conditional means in 2015 baseline quality performance and agency characteristics that were likely linked to the quality performance of agencies in 2016 (Bruhn & McKenzie, 2009). Patient characteristics of agencies examined include patient demographics, and the proportion of patients typically associated with lower profit margins. Patients associated with lower profit margins for home health agencies include patients residing in the bottom quartile of the average ZIP code income per state, patients discharged from an inpatient facility within 14 days before admission to home health, patients with poor control of clinical conditions, patients with overall high health risk, patients receiving intravenous therapy or parenteral nutrition, and patients with traumatic wounds or ulcers (Centers for Medicare & Medicaid Services, n.d.). I use 100% home health OASIS data from 2015 to obtain these patient characteristics.

I also check for balance in a number of organizational characteristics of the home health agencies, including the number of patients admitted to the agency, ownership model, freestanding status, and rural mix of each agency's catchment areas. I use the 2015 Provider of Service file to obtain ownership and freestanding status. I follow a two-step process to determine the rural mix of each agency's catchment areas. I first use ZIP codes of patients served by each agency to obtain counties served using the US Department of Housing and Urban Development's algorithm; I obtain agencies' rural status based on whether the counties served by an agency were rural, using the US Department of Agriculture Urban Influence Codes (Probst et al. 2014).

While the HHVBP agencies were similar to the non-HHVBP agencies in many regards, there were also some important differences (Table 1.2). For example, the treatment group had more agencies that served urban counties only; more patients; more patients who were white and of Hispanic ethnicity; and more patients with traumatic wounds or ulcers and significant bathing needs (Panel A, Table 1.2). Treatment agencies also tended to perform better than control agencies across the agency-reported outcome measures (Panel B, Table 1.2). These descriptive statistics suggest a need to account for differences between agencies to achieve unbiased estimates. Throughout the analysis, I use a difference-in-differences design to account for these potential sources of confounding in all specifications.

Table 0.2: Comparison of 2015 characteristics between treatment and control agencies under the Home Health Value-Based Purchasing program.

	Control (1)	Treatment (2)	
Panel A. Sample characteristics			
Number of states	41	9	
Number of agencies	6,981	1,630	
Number of agencies per state (Median)	87	125	
Panel B. Organizational characteristics			
Ownership			
For-profit, percent	77.5	74.5	
Not-for-profit, percent	18.1	18.6	
Government, percent	4.4	6.9	
Setting			
Freestanding, percent	89.5	89.6	
Rural catchment areas			
Only serve metropolitan counties, percent	38.7	45.9	
Panel C. Patient Characteristics			
Distinct patients admitted (No.)	638.7	872.9	***
	(1642.4)	(1440.5)	
Admissions by payer source			
Medicare FFS, percent	78.7	77.6	
	(22.1)	(24.8)	
Medicaid, percent	9.4	9.1	
	(15.6)	(18.4)	
Medicare Advantage, percent	15.7	16.0	
	(17.8)	(20.9)	
Private, percent	2.0	2.0	
	(5.5)	(6.4)	
Medicare admissions			
Average Age, mean	75.3	77.1	***
	(4.4)	(4.2)	
Female, percent	61.7	61.8	
	(7.8)	(8.1)	
White, percent	67.4	74.8	***
	(30.8)	(29.0)	
Hispanic, percent	10.5	14.7	
	(19.5)	(28.0)	
Reside in low-income ZIP codes, percent	32.1	25.5	***
	(25.8)	(22.9)	
Lower-profit margin Medicare admissions			
Discharged from acute care, percent	52.9	55.6	
	(24.0)	(23.4)	
Poor control of clinical conditions, percent	6.7	4.5	***
	(6.2)	(3.9)	
Overall high risk, percent	32.3	32.7	
	(21.8)	(21.4)	
IV therapy or parenteral nutrition, percent	2.6	2.5	
	(5.4)	(4.6)	
Traumatic wounds or ulcers, percent	9.8	11.0	***
	(7.8)	(7.9)	
Significant bathing needs, percent	17.6	19.6	
	(12.3)	(12.5)	

Notes: *** P < 0.05; IV = Intravenous; Standard deviations in parentheses
Comparisons are adjusted for region and state-level clustered standard errors.

	Control (1)	Treatment (2)	
Panel B. Baseline quality performance.			
Agency-reported outcome measures			
Ambulation-Locomotion	63.2	65.6	
	(13.0)	(11.4)	
Bed Transferring	58.3	61.4	
	(14.7)	(12.0)	
Bathing	67.0	69.8	
	(14.5)	(12.4)	
Dyspnea	63.5	67.4	***
	(18.5)	(15.6)	
Pain Interfering with Activity	66.7	70.9	
	(17.5)	(13.2)	
Management of Oral Medications	52.3	52.7	
	(14.5)	(14.7)	
Agency-reported process measures			
Pneumococcal Polysaccharide Vaccine	68.7	67.8	
	(22.8)	(23.0)	
Influenza Immunization	65.5	65.5	
	(19.2)	(19.9)	
Drug Education	94.9	94.4	
	(8.0)	(8.6)	
Patient-survey measures			
Professional care	88.6	88.7	
	(4.1)	(3.7)	
Communication with patients	85.9	85.8	
	(4.8)	(4.3)	
Specific care issues	83.5	82.8	
	(5.7)	(5.6)	
Overall care rating	84.5	84.4	
	(6.4)	(6.0)	
Would recommend agency	79.6	79.8	
	(8.2)	(7.2)	
Administrative-claims measures			
No unplanned Hospitalization	84.3	84.1	
	(4.1)	(3.7)	
No Emergency Department Use	87.5	88.0	
	(4.2)	(3.8)	

Notes: *** $P < 0.05$; IV = Intravenous; standard deviations in parentheses Comparisons are adjusted for sampling region and state-level clustered standard errors.

1.3.2 Heterogeneous incentives under the HHVBP

In addition to examining the aggregate effects of the HHVBP program on targeted quality performance, I also assess the relationship between the size of marginal incentives to improve to agencies' quality performance for each targeted measure. To do so, I exploit both the random treatment group selection and a calculated measure of marginal incentives for each agency and targeted quality measure.

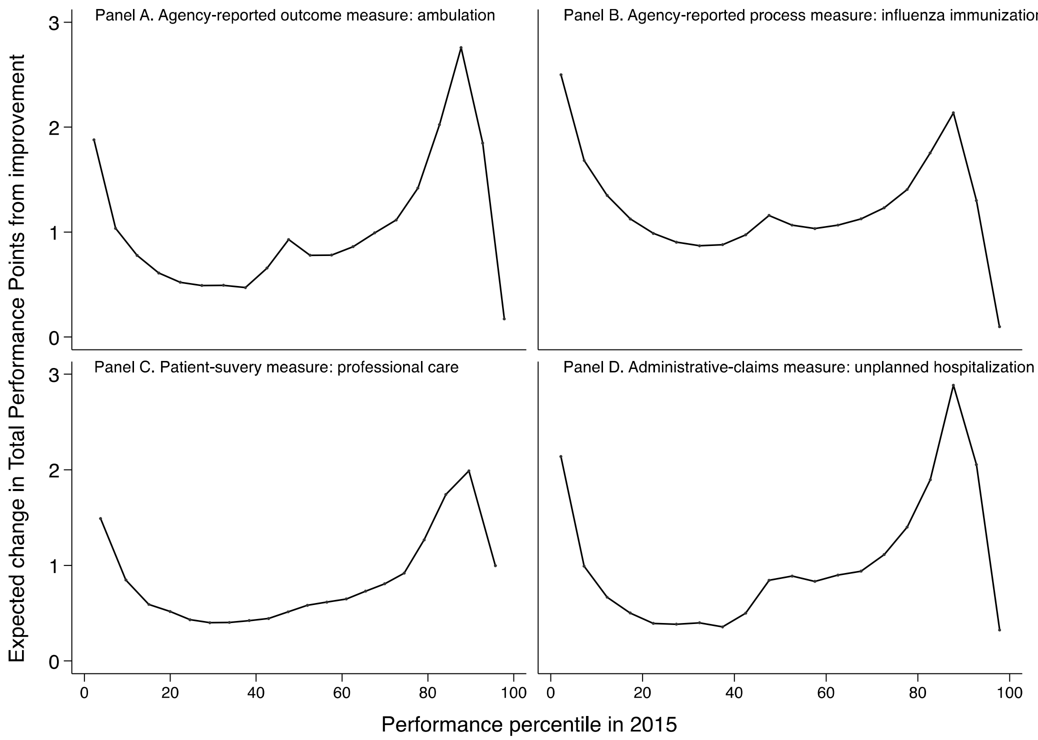
I use a non-parametric approach to estimate the measure-specific incentive size for each agency, following the general method outlined by Norton, Li, Das, and Chen (2018). In the HHVBP program, how much an agency improved on a measure was determined from comparing an agency's performance in 2016 against the agency's 2015 baseline performance. To evaluate how well an agency performed relative to its peers, the program compared each agency's performance rate in 2016 against the 2016 performance of other agencies within the same state. Thus, prior to the start of the program, each agency in 2015 could have estimated its expected rewards from improvement, or marginal incentive for improvement, for each measure targeted by the program by considering how much its Total Performance Score would change if it improved the measure by a modest increment.

I construct the measure-specific marginal incentive for improvement for each agency following a number of steps. First, I compute an initial Total Performance Score for each agency, for both control and treatment states, using their 2015 performance. Second, I compute an improved Total Performance Score for each agency across both treatment and control groups in my sample as if the agency improved its performance rank by a decile for each measure. I do this by first ranking each agency within each state on its 2015 performance for each measure. I then assign each agency a new hypothetical performance rate that is taken from the performance rate of an agency that was one-decile higher in rank. For example, for a state with 100 agencies, the worst (100th ranked) agency receives a hypothetical decile-improved performance rate of the 90th ranked agency. Agencies in the top 10th percentile (1st–10th) are assigned the performance of the best performing agency in the state. I then replicate CMS' procedure to calculate an improved Total Performance Score. Finally, for each agency and measure, I calculate the difference between the improved Total Performance Score and the initial Total Performance Score. This expected increase in Total Performance Score is my measure of marginal incentive for improvement for each targeted measure under the program.

Calculating the marginal incentive for improvement as the expected increase in Total Performance Score from one decile in improvement has several benefits. A one-decile in improvement is modest in terms of performance and is standardized across measures. The average marginal incentive for improvement from one decile in improvement was approximately one more Total Performance Score point for each measure. The mean baseline Total Performance Score was 23 (SD=16) (Appendix Tables A.1–A.2). It is also consistent with the scoring structure of the HHVBP, where agencies compete with one another for points based on a ranking system. Thus, a one-decile improvement in a measure is large enough to be meaningful while also allowing me to simulate the incentives faced by agencies under the program. Finally, while a decile in improvement represents incentives for agencies to gain rewards, it is also generally consistent with the incentives for agencies to avoid penalties. Using a similar approach, I simulate each agency's expected decrease in Total Performance Score from a decile decline in performance. I find that agencies' incentives to avoid penalties are highly correlated with their incentives to improve for those in the upper half of the performance distribution. For agencies in the bottom half of the performance distribution, they face no incentives to avoid penalties because agencies are given 0 achievement scores when they perform below the median. Combined, it suggests that where the incentive to avoid penalties are not highly correlated with the incentive to improve, the incentives to improve dwarfs the incentives to avoid penalties.

Approximately 94% of agencies had positive incentives (or 6% with zero marginal incentives for improvement), but the incentives were heterogeneous (Figure 1.4). The marginal incentives to improve were larger for those in the bottom two percentiles of performance and in the 80_{th} to 90_{th} percentiles. For agencies at the top of the distribution, there was little room for improvement and their marginal incentives for improvement were zero. The distribution of competing agencies was another reason for some agencies to face no incentives. In these cases, improving rank might not increase an agency's score. These general patterns in incentives across the performance distribution were similar to patterns found in the precursor Medicare Value-Based Purchasing program for hospitals, although there were substantially fewer agencies with zero incentives to improve in this program (Norton et al. 2018).

Figure 1.4: Binned scatter plot of marginal incentives to improve on four example measures.



Notes: Scatter plots are presented in bins of 20.

1.3.3 HHVBP targeted quality outcomes

The primary outcomes in this study are the 16 measures of quality targeted by the HHVBP (Table 1.1). I gather these measures from Medicare's Home Health Compare website, which provides agency-level performance information corresponding to the HHVBP program's baseline year 2015 and performance year 2016 (Centers for Medicare & Medicaid Services, 2019). Nine measures were self-reported by the home health agencies to CMS via OASIS; of which six measures captured the percent of each agency's patients with improved ability to function in activities of daily living between start of care and end of care; and three measures focused on how well the agencies adhered to clinically recommended processes of care. Among the non-agency-reported measures, another five examined patient satisfaction collected from patient surveys, which were administered by independent and Medicare-approved survey vendors. The last two measures focused on decreased hospitalization and ED use during the home health episode. These two non-agency-reported measures were calculated by CMS using Medicare administrative-claims data (Health Care Financing Administration 2000; Centers for

Medicare & Medicaid Services 2015c). Finally, there were four additional measures included in the HHVBP but not currently available to the public and therefore excluded from my analysis.

1.3.4 HHVBP's effects on true underlying quality: coded functional deficit levels

To discern whether the effects of the HHVBP on targeted quality were driven by coding manipulation by agencies, I examine how patients' functional status was coded in the OASIS. Because HHVBP performance on six of the agency-reported measures was determined by whether patients had improved functioning between the start of care and end of each home health care episode, a straightforward method for agencies to improve their measured performance was to report more functional deficits at baseline and/or fewer functional deficits at discharge. As mentioned previously, coding patients as having more functional deficits at the start of care would also improve an agency's Medicare reimbursement – more deficits translate to greater resource use and thus higher reimbursement by Medicare (Medicare Payment Advisory Commission 2019a). Therefore, agencies faced incentives to up-code from both the HHVBP program and existing Medicare reimbursement policy. To check for these undesirable up-coding practices, I examine two outcomes. I compare HHVBP agencies to non-HHVBP agencies, before and after the HHVBP program's implementation on 1) the agency-reported functional deficits of patients at the start of care, and 2) the difference in agency-reported functional deficits between discharge and readmission to home health care within one day.

To calculate the first measure, I use the functional deficits coded in OASIS assessments at the start of care for Medicare fee-for-service patients, and I only include episodes that were also included in the HHVBP measure specifications for the agency-reported outcome measures (Centers for Medicare & Medicaid Services 2016). Higher functional deficit levels represent greater disability. The functional abilities include ambulation (levels 1–6), bed transfer (1–5), bathing ability (1–6), dyspnea (1–4), pain interference (1–4), and medication management ability (1–3). Second, to measure the change in functional deficit levels between discharge and readmission, I focus on the difference in functional deficit levels of patients who were discharged and then admitted for another episode of care within one day of discharge. Only patients with a discharge and readmission within one day are included in this measure and readmissions are not restricted to the same agency.

These two measures offer complementary insights. The first measure examines differences in average functional deficits of patients at the start of care by exposure to the HHVBP program; if there are differential increases in average functional deficits at the start of care, this would suggest coding manipulation but does not eliminate the possibility that patients treated by the HHVBP program had coincidental worsened functional abilities. However, the second measure, by exploiting home health readmissions for the same patient, holds constant patients' true functional abilities, which are unlikely to change within the short duration of one day (between discharge from the initial episode and readmission for the subsequent episode). It also provides more concrete information about unsavory coding practice changes because if HHVBP agencies were coding more accurately (which could be suggested by the findings involving the first measure), they should not have a larger change—more specifically a greater increase in patients' functional deficits between start of care at the readmission and the initial episode discharge—when compared to agencies not in the HHVBP. Moreover, agencies in the HHVBP faced an incentive to rate patients as having more functional deficits at entry into home health care and an incentive to rate patients as having fewer functional deficits at exit. Therefore, if the difference in patients' agency-rated functional health between discharge and readmission is large and indicates greater relative deficits at the start of readmission compared to the end of the initial discharge, this provides evidence consistent with the notion that agencies were strategically up-coding. I again examine deficit levels for ambulation, bed transfer, bathing ability, dyspnea, pain interference, and medication management ability using a difference-in-differences approach.

For both functional deficit coding measures, there may be rating variation introduced by each agency's staff. It is well known that determining patient function is subjective and varies by the rater (O'Connor and Davitt 2012). Thus, to minimize the undue influence of interrater differences among agencies with few patients, I restrict both measures to agencies with at least 10 Medicare patients per year. The mean functional deficit levels did not vary by the number of patients per year, only the variance. Therefore, with a larger denominator per agency, the noise from interrater variability is reduced and leads to more reliable measures of how each agency rates its patients.

1.3.5 HHVBP's effects on true underlying quality: non-targeted quality measures

To discern whether HHVBP led to true improvements in care delivery, I also examine quality measures that were not targeted by the program but are plausibly related to improved quality.

The rationale behind examining non-targeted measures is that, if the performance improvements among HHVBP agencies were true improvements that benefited patients' functional independence, one might expect fewer patients to use additional health care services (Tao and Ellenbecker 2013; C.-L. Li et al. 2011; Luppa et al. 2010). I test whether the program had an effect on 1) the rate at which patients were discharged to an inpatient institutional facility at the end of their home health care episode and 2) the rate at which patients used additional home health care.

First, I determine the rate at which patients were discharged to an inpatient institution by examining all discharge assessments that flagged patients with an admission to inpatient facilities (i.e., hospital, rehabilitation facility, nursing home, hospice) at the home health discharge. Second, I examine the rate at which patients used additional home health care by calculating the proportion of each agency's Medicare fee-for-service patients who were either certified for additional home health care before the end of the initial 60-days episode of care or started a new home health episode within one day of discharge. I only include Medicare fee-for-service beneficiaries in order to track patients across agencies.

1.4 Empirical Strategy

This study examines the effect of the HHVBP program on the quality of care provided by home health care agencies by comparing treatment to control agencies before and after the start of the program. The analysis begins with testing whether the program's introduction had an aggregate effect on targeted measures of quality. The study next tests whether there was a direct relationship between each agency's marginal incentives to improve and performance on the targeted measures. Finally, to discern whether the program's effects on performance reflected improvements in underlying quality, I test for the program's effects on functional deficit coding and the program's effects quality measures not targeted by the program. I describe the empirical specifications used to examine these tests below.

Prior to looking at the data on the outcomes for the treatment group, the majority of the analysis presented in this paper has been pre-specified and registered in the AEA RCT Registry as AEARCTR-0004159 (J. Li 2019). Pre-registering this study minimizes issues of *p*-hacking and provides transparent documentation of the development of the study.

1.4.1 Aggregate effects of HHVBP

Equation (1) examines the aggregate effect of the program by assessing the changes in outcomes among agencies in the HHVBP states relative to those in non-HHVBP states before and after the implementation of the HHVBP. I use a difference-in-differences design to assess the relative change in outcomes while controlling for fixed differences across agencies and secular time trends (Greene, 2018).

$$Y_{hst}^{m} = \beta_0^m + \beta_1^m HHVBP_s + \beta_2^m Post_t + \beta_3^m HHVBP_s \times Post_t^m + \epsilon_{hst}^m \quad (1)$$

The data are constructed at the agency h by year t level. In this equation, $HHVBP_s = 1$ if agency h operated within a state s participating in HHVBP and 0 otherwise; $Post_t = 1$ if period t occurred in performance year 2016 and 0 if in baseline year 2015.

The dependent measure Y_{hst}^{m} denotes the outcomes examined in this study, indexed by m. The first set of outcomes are the 16 measures of quality targeted by the HHVBP program. The second set of outcomes include measures to discern whether the program had an effect on the coding practices among agencies. These measures include six measures that capture the average functional deficits per patient per agency at the start of care; and six measures that reflect average change in functional deficits between the start of care at a home health readmission and discharge from an initial home health episode. The last set of measures focus on two measures of quality not targeted by the HHVBP program: percent of agencies' discharges to inpatient institutions and percent of agencies' patients receiving additional home health episodes. I run Equation (1) separately for each of the 30 outcomes.

Each estimate of β_3^m is of primary interest because it provides the change in measure m in HHVBP agencies relative to non-HHVBP agencies comparing 2016 versus 2015. The null hypothesis is $\beta_3^m = 0$. If the HHVBP program increased the outcome, then $\beta_3^m > 0$, and if the policy decreased the outcome, then $\beta_3^m < 0$. The standard errors are clustered at the state level since treatment was assigned at the state level (Abadie, Athey, Imbens, & Wooldridge, 2017).

1.4.2 Relationship between incentive size and performance under the HHVBP

To test whether agencies with larger marginal incentives to improve respond differently to the program than agencies with smaller marginal incentives to improve, I use a difference-in-difference-in-differences approach in Equation (2).

$$\begin{aligned} Y_{hst}^{m} = {} & \theta_7^m(HHVBP_s \times Post_t \times Incentive_h^m) + \theta_6^m(HHVBP_h \times Post_t) \\ & + \theta_5^m(HHVBP_s \times Incentive_h^m) + \theta_4^m(Post_t \times Incentive_h^m) \\ & + \theta_3^m Incentive_h^m + \theta_2^m Post_t + \theta_1^m HHVBP_s + \theta_0 + v_{hst}^m \end{aligned} \quad (2)$$

Equation (2) is similar to Equation (1) but it includes a continuous variable $Incentive_h^m$. The variable $Incentive_h^m$ reflects the marginal incentive to improve, measured as the expected increase in Total Performance Score points if agency h improved its 2015 performance on dependent measure m by one decile within its state s; $Incentive_h^m$ for control agencies is the expected marginal gains in Total Performance Score Points had they been participants in HHVBP. I run 16 regressions using Equation (2), one for each of the 16 targeted measures under the HHVBP.

The estimand of interest is θ_7^m, which estimates the marginal effect of the HHVBP program for agencies with positive $Incentive_h^m$ on each measure m. If HHVBP agencies with larger incentives perform better than HHVBP with smaller incentives, then $\theta_7^m > 0$. If HHVBP agencies with no incentives to improve do not perform better than non-HHVBP agencies, then $\theta_6^m = 0$. Standard errors are again clustered at the state level to reflect the level at which treatment was assigned.

1.4.3 Handling multiple outcomes: Standardized treatment effect indices

This paper reports the effects of the HHVBP in terms of summary indices for a majority of the outcomes of interest. Summary indices are aggregations of the treatment effects across related outcomes. Aggregating to indices removes the issue of false positives arising from conducting a large number of statistical tests and also have the added benefit of increased power. Furthermore, drawing conclusions from aggregated effects across related constructs has more policy salience since it is more important to know whether the program led to improved quality of a particular type, e.g., functional abilities, than if the program led to improvements on a specific measure, such as bathing ability. Therefore, I use the indices where applicable to draw conclusions about the effects of the program. The individual, unaggregated effects with *p*-values adjusted for multiple hypothesis testing are also reported in the appendix (Westfall and Young 1993).

To produce the summary indices, I perform omnibus tests on whether there was any overall effect of the HHVBP program on a particular domain, or set, of related outcomes. Outcomes are considered related if they measure similar constructs. They also correlate within each domain. The 16 targeted measures of the program are grouped into four mutually exclusive domains: 1) six agency-reported functional ability outcomes, 2) three agency-reported clinical process measures, 3) five patient experience survey measures, and 4) two health care utilization measures derived from Medicare administrative claims. The summary indices for the effects of the program on functional deficit coding are also reported. For each of the two measures (functional deficits of patients at start of care and the difference in functional deficits between discharge and start of care at readmission), the effects from six functional abilities including ambulation, bed transfer, bathing ability, dyspnea, pain interference, and medication management ability are aggregated.

I followed the steps outlined in Kling, Liebman, and Katz (2007) to obtain the standardized treatment effects, which are summary indices of the average treatment effects across the outcomes of each domain, standardized by the standard deviations of the respective measures in the control group. I estimate the treatment effects for each outcome using the specifications (Equations 1 and 2) described in sections 1.4.1 and 1.4.2, standardize them, and average them, while accounting for the covariance of the treatment effect estimates. The mean effect size for a given domain for a set of M outcomes is denoted by Equation (3).

$$Index_M = \sum_{m \in M} \frac{1}{M} \frac{\xi_m}{\sigma_m} \tag{3}$$

In this equation, σ_m is the standard deviation of outcome m in the control group and ξ_m is the treatment effect estimate (β_3^m from Equation (1), or θ_7^mor θ_6^m from Equation (2) for measure m). The term $\frac{\xi_m}{\sigma_m}$ puts the treatment effects in terms of the standard deviations of the control group, and $\frac{1}{M}$ averages over the standardized treatment effects for a set of related outcomes. To account for the covariance of the treatment effect estimates, I estimate seemingly unrelated regressions for all m within a domain M and clustered errors at the state level. Each $Index_M$ is interpreted as the average treatment effect for domain M in standard deviations of the control group.

The standardized treatment effect of HHVBP across the three agency-reported clinical process measures, for example, is estimated by taking the average of $\beta_3^{Pneumococcal\ vaccine}$ divided by the standard deviation of the control group's performance on pneumococcal vaccine, $\beta_3^{influenza\ vaccine}$ divided by the standard deviation of the control group's performance on influenza vaccine, and $\beta_3^{drug\ education}$ divided by the standard deviation of the control group's performance on drug education from Equation (1).

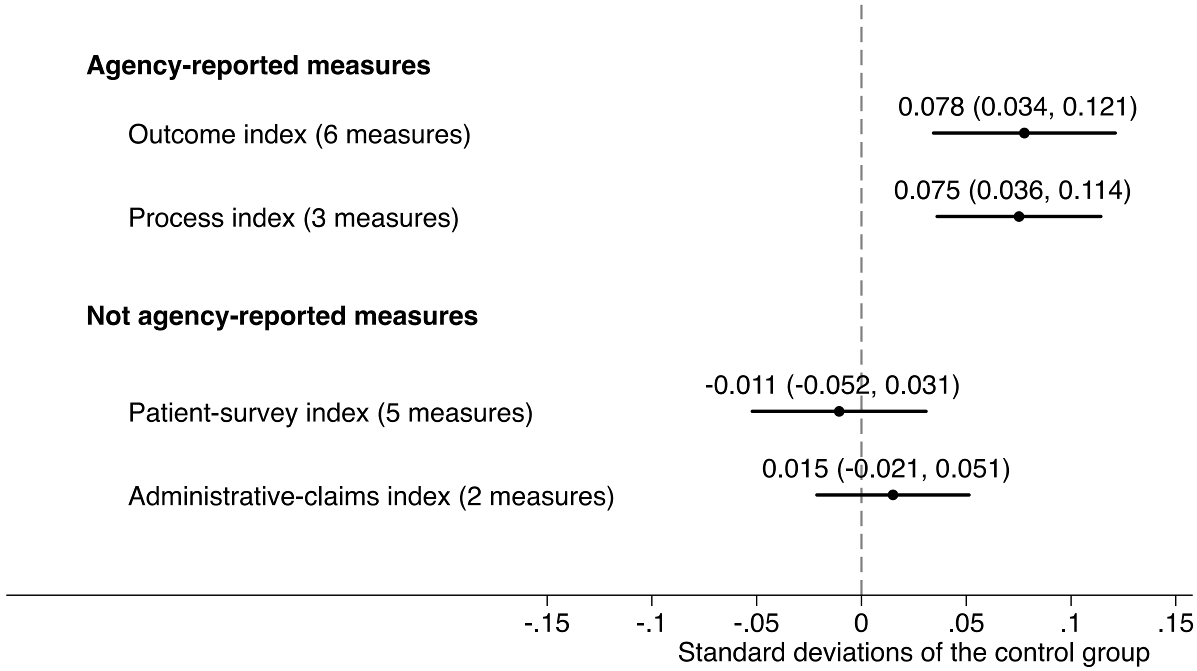
I consider the two quality outcomes not targeted by the HHVBP (percent of discharges to inpatient institutions and percent of patients receiving additional home health episodes) to capture distinct types of additional health care use and therefore do not aggregate them into a single summary index. In total, I use ten aggregate tests and two individual tests to draw conclusions about the effects of the HHVBP program.

1.5 Results

1.5.1 Aggregate effects of HHVBP program on targeted quality performance

The results indicate that in aggregate, the HHVBP program led to better performance on agency-reported measures (outcomes and process measures), but not on the non-agency-reported measures from patient surveys and administrative claims (Figure 1.5). The positive effects of the program on agency-reported outcomes and process measures were both statistically significant and similar in magnitude. The effects of HHVBP on standardized treatment effects for agency-reported outcomes were 0.078 (SE=0.022) of a standard deviation increase in performance and 0.075 (SE=0.020) for process measures. This effect is approximately a one percentage-point increase in performance, estimated by taking the average of the point estimates of the respective domains (Appendix Tables A.3). To put these estimates into more context, these effects indicate that for the average agency, approximately four out of 400 more patients per agency had improved quality as measured by the HHVBP program.

Figure 1.5: Effect of the Home Health Value-Based Purchasing program, with units in standard deviations of the control group, on targeted quality measures.



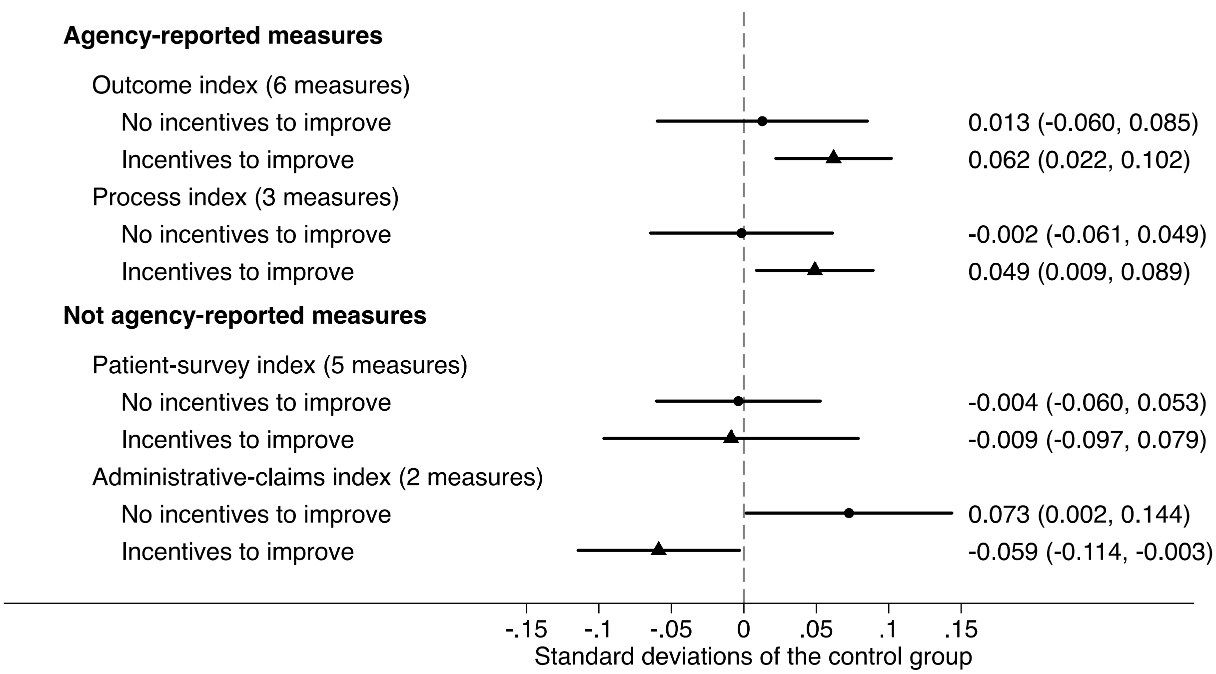
Notes: The effects, with 95 percent confidence intervals, are standardized treatment effects estimated from coefficient estimates from a difference-in-differences model (Equation 1) for each measure domain.

The HHVBP program had no discernable effect on the patient-survey and administrative-claims measures (Appendix Table A.4). The estimates indicate a non-significant decrease in patient-survey measures, with an effect of −0.011 (SE=0.021) of a standard deviation. This estimate corresponds to approximately −0.05 percentage points. Given the perverse incentives in Medicare's existing reimbursement system to overuse hospitalizations and ED, it was also unsurprising to find that the program did not lead to reductions in hospitalizations or ED use. For administrative-claims measures, the effect was 0.015 (SE=0.019), which corresponds to a statistically non-significant increase of approximately 0.24 percentage points for unplanned hospitalizations and decrease of 0.12 percentage points for ED visits.

1.5.2 Effects of HHVBP on targeted quality performance by incentive size

The results in Figure 1.6 demonstrate that the incentive size matters for quality performance. When HHVBP agencies had no marginal incentives to improve, their performance on the agency-reported measures was not statistically significant (Figure 1.6). In contrast, HHVBP agencies that faced marginal incentives to improve performed better on the agency-reported measures than those with no incentives to improve (Appendix Table A.5). The standardized treatment effect on agency-reported outcome measures was 0.062 (SE=0.020) of a standard deviation increase in performance for each unit increase in expected Total Performance Score, which is approximately a one percentage-point increase in performance. The effect of having marginal incentives to improve on agency-reported process measures was similar: 0.049 (SE=0.021) of a standard deviation increase in performance or approximately a one percentage-point increase in performance. Together, these findings show that the aggregate effects of the HHVBP program on agency-reported quality measures were driven by agencies that faced marginal incentives to improve.

Figure 1.6: Effects of the Home Health Value-Based Purchasing program by incentive size, with units in standard deviations of the control group, on targeted quality measures.



Notes: The effects, with 95 percent confidence intervals, come from the difference-in-difference-in-differences model (Equation 2) comparing treatment agencies with marginal incentives to improve to agencies with no marginal incentives to improve under the first year of the HHVBP program.

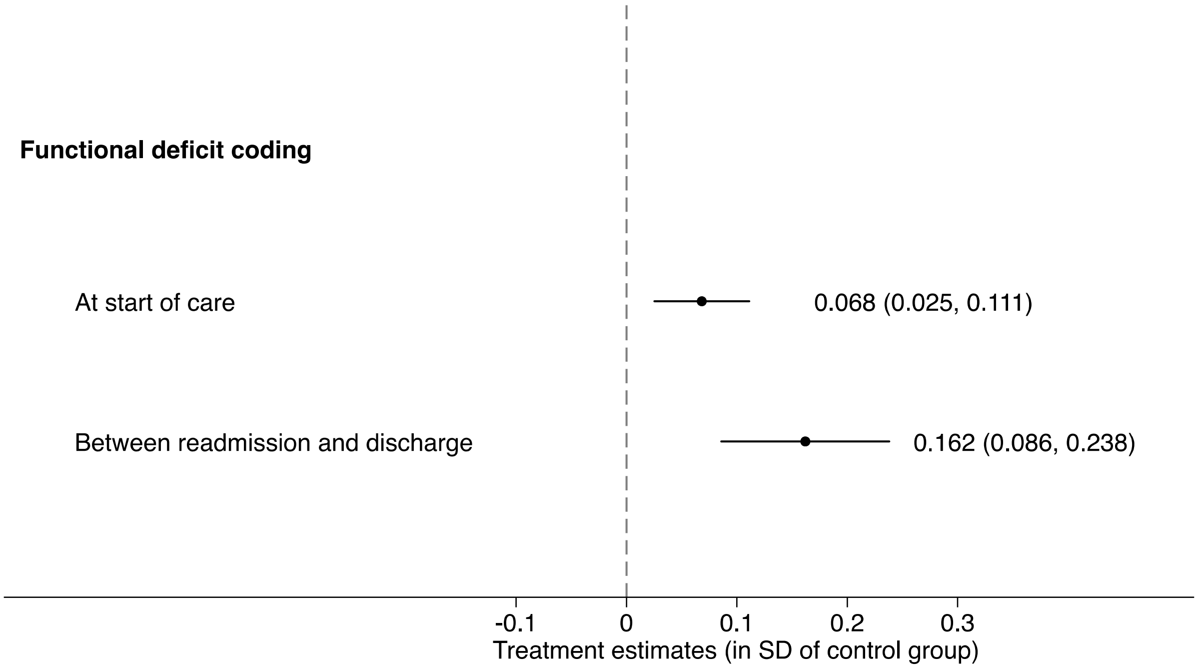
The relationship between the incentive size and performance showed a different pattern for the non-agency reported quality measures targeted by the program (Appendix Table A.6). No statistically significant effects of incentive size were observed for patient-survey measures. Furthermore, the effects of incentive size on performance on administrative-claims measures (unplanned hospitalizations and ED use) were of the opposite direction. For the two administrative-claims measures, performance increased on administrative-claims measures among HHVBP agencies with no expected marginal incentives to improve (0.073 (SE=0.036) of a standard deviation), and agencies facing larger marginal incentives to improve performed worse, -0.059 (SE=0.028) of a standard deviation.

1.5.3 Improvements in underlying quality versus coding manipulation

One possible reason that the HHVBP program led to improved agency-reported quality performance was because agencies made true improvements in underlying quality. An alternative explanation may be that the improvements were the results of coding manipulation by home health agencies. To disentangle these two mechanisms, I examine the effects of the HHVBP program on measures related to functional deficit coding using Equation (1) and two measures of quality not targeted by the HHVBP program. I then conduct a back of the envelope calculation to determine how much of the performance gains in agency-reported outcomes were due to coding practice changes.

I find consistent evidence that patients treated by HHVBP agencies had more coded functional deficiencies at the start of care than agencies not in the program (Figure 1.7); the differential effect of the HHVBP program on the average level of functional deficits at the start of care was 0.068 (SE=0.022) of a standard deviation or approximately an increase of 0.02 of a functional deficit level (Appendix Table A.7). Within-patient changes in coded functional deficits between discharge and readmission to home health care within one day also indicate an increase in coded deficits; even among patients who are unlikely to have had an actual change in deficits during the one-day transition from the initial episode to the readmission episode, patients under the HHVBP had a differential increase of 0.162 (SE=0.039) of a standard deviation in coded deficits, which is approximately an increase of 0.07 of a functional deficit level.

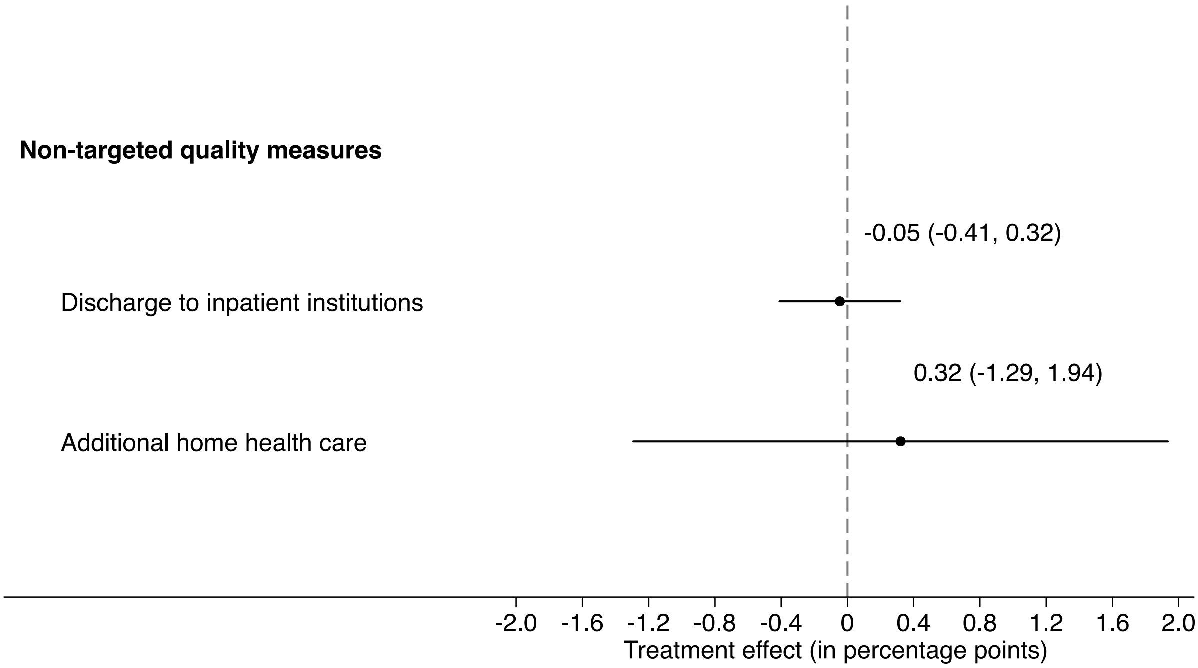
Figure 1.7: Effect of the Home Health Value-Based Purchasing program on coding manipulation measures.



Notes: The effects are estimates with 95 percent confidence intervals from the difference-in-differences model (Equation 1). The estimates for functional deficits are standardized treatment effect indices across functional abilities for ambulation, bed transfer, bathing ability, dyspnea, pain interference, and medication management ability.

I also find that the HHVBP did not lead to better quality of care when assessed using two non-HHVBP measures (Figure 1.8). In fact, the HHVBP program had no effect on patients' discharge rate to inpatient institutionalizations (-0.05 (SE=0.18) percentage points) nor on their use of additional home health care (0.32 (SE=0.80) percentage points) (Appendix Table A.8).

Figure 1.8: Effect of the Home Health Value-Based Purchasing program on non-targeted quality measures.



Notes: The effects are estimates with 95 percent confidence intervals from the difference-in-differences model (Equation 1).

Given evidence of coding manipulation and that HHVBP did not appear to improve true quality of care, the natural follow-up question is to what extent this coding manipulation affected the apparent improvements under HHVBP. Thus, I estimate how much of the performance effect could be explained by coding manipulation using a back of the envelope calculation. The basic goal of this exercise is to determine whether up-coding functional deficits by a modest amount, one or two deficit levels, match the estimated performance gains in HHVBP agency-reported measures.

To begin, recall that the agency-reported outcome measures examine the percentage of patients with improved functional abilities between start and end of care. Therefore, a one percentage-point increase in the percentage of patients with improved agency-reported outcomes means that there were approximately four more patients per agency with improved agency-reported outcomes due to the HHVBP. This figure comes from an estimate of 395 patients per HHVBP agency in 2015 across the six agency-reported measures.

Next, using the estimated effect of the HHVBP on functional deficit levels measured at the start of care from Equation 1, an average increase of 0.02 functional deficit levels means that total functional deficit levels increased by approximately 7.9 per agency (0.02×395). If each agency up-coded by one deficit level per patient, this translates to approximately eight patients ($8 \div 1$) or a two percentage-point increase ($8 \div 395$) in the percent of patients with improved functional abilities under the HHVBP. Up-coding by two levels per patient would mean that four patients or one percentage point more patients would have been considered to have improved by the HHVBP. In other words, if agencies up-coded by a modest amount of one or two functional deficit levels, this up-coding explains the entirety of the performance improvement on agency-reported outcomes measures. While this calculation is estimated with some error, the performance gains estimated from a 0.02 functional deficit coding increase exceed or match the one percentage point performance effects of the HHVBP. Further, 90 percent of functional deficits were coded at or below the middle severity level in 2015, indicating that there was room to up-code patients. In total, these results indicate that the HHVBP did not lead to true improvements in quality and that the apparent gains in quality performance was driven by coding manipulation.

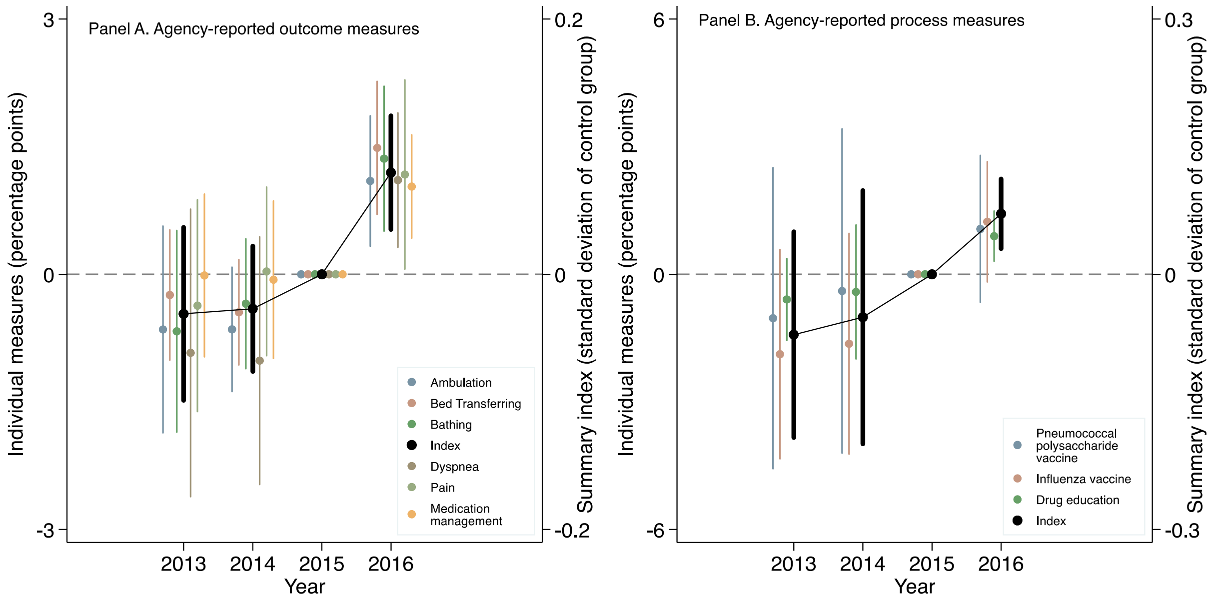
1.6 Robustness checks and sensitivity analyses

1.6.1 Parallel trends

To assess the validity of my empirical approach, I conduct several robustness checks. I first check for evidence of violation of parallel trends in the pre-intervention period for each of the outcomes. The estimates from a difference-in-difference model is an unbiased estimate of the effect of the HHVBP program if, absent the program, the average change in the outcomes between treatment and control agencies would have been the same for treatment and controls (Angrist and Pischke 2009).

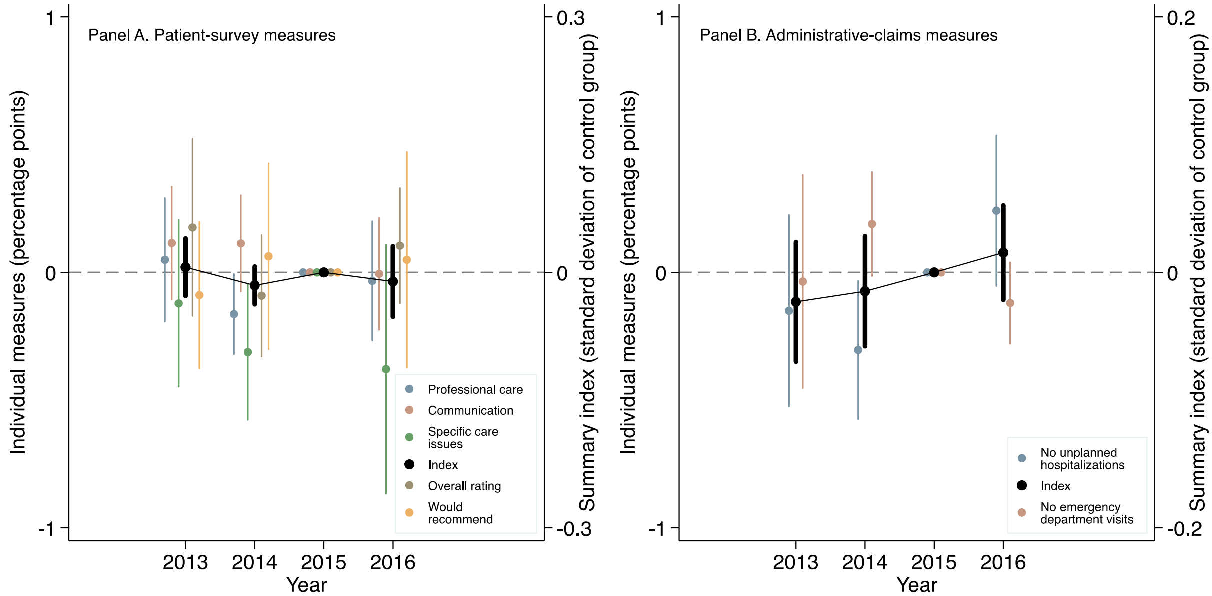
To test for parallel trends, I construct the data in the same fashion as my main analysis, but I add additional years of pre-intervention data to conduct event studies. The presence of a treatment effect before 2016 would suggest that there may be diverging pre-trends. For each of the 16 targeted quality measures, I examine performance from 2013 through 2016 (Figures 1.9–1.10). I use a similar regression model as equation (1), but I include period dummies for 2013 and 2014 in addition to 2016. For the functional deficit coding and non-HHVBP quality measures, I add in 2014 data and a period dummy for 2014 in addition to 2016 (Figures 1.11–1.12). I find no evidence of a treatment effect before 2016.

Figure 1.9: Event study of the effect of the Home Health Value-Based Purchasing program on agency-reported targeted quality measures.



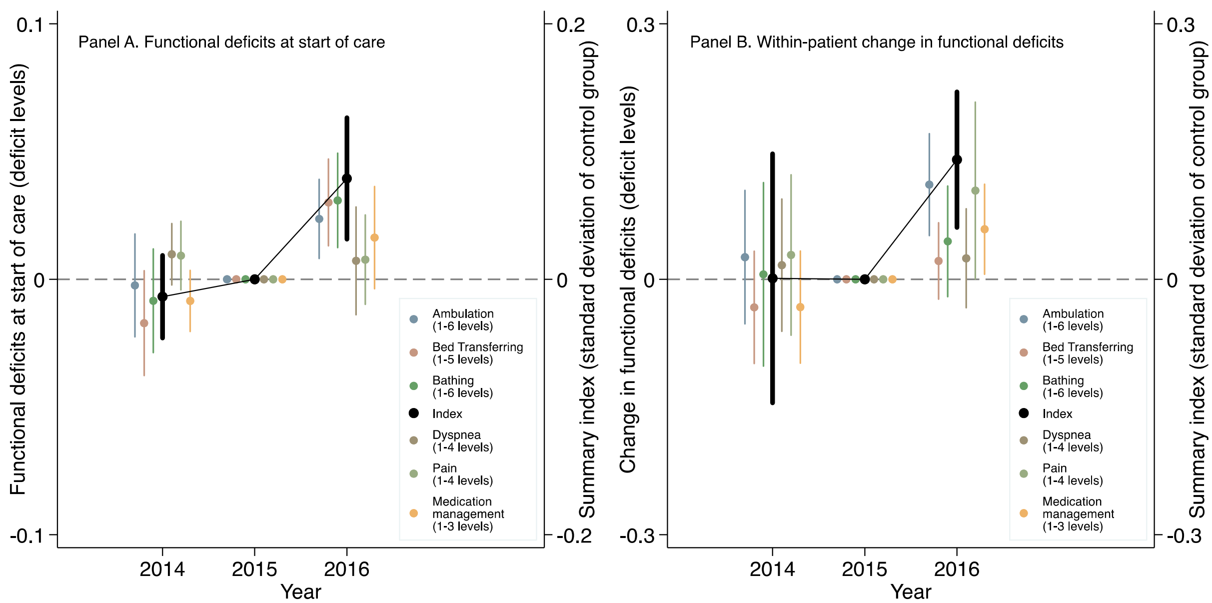
Notes: The effects are treatment estimates with 95 percent confidence intervals from the difference-in-differences model (Equation 1) plus dummy variables for 2013, 2014, and 2016. The left y-axis corresponds to the individual measures on the percent of patients with improvements in ambulation, bed transferring, bathing, dyspnea, pain, and medication management in panel A and percent of patients with pneumococcal polysaccharide vaccine, influenza vaccine, and drug education in panel B. The right y-axis corresponds to the index for each domain as standardized treatment effects in units of the standard deviation of the control group.

Figure 1.10: Event study of the effect of the Home Health Value-Based Purchasing program on non-agency-reported targeted quality measures.



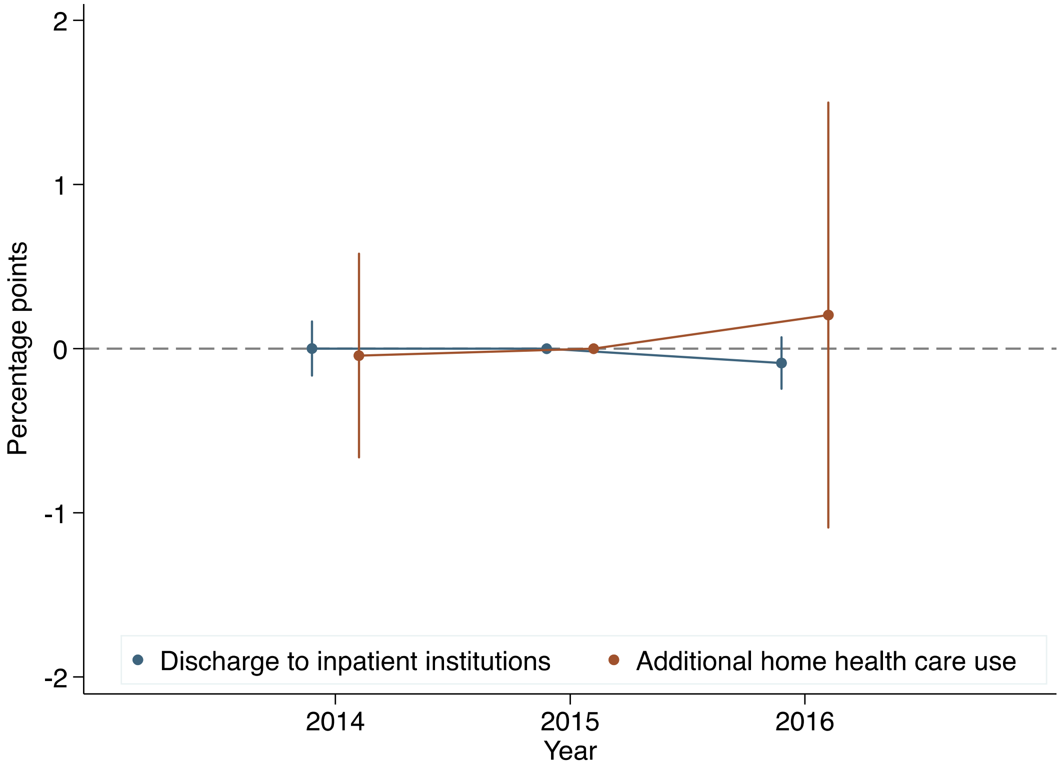
Notes: The effects are treatment estimates with 95 percent confidence intervals from the difference-in-differences model (Equation 1) plus dummy variables for 2013, 2014, and 2016. The left y-axis corresponds to the individual measures on the percent of patients rating agencies highly for professional care, communication, specific care issues, overall rating, and would recommend in panel A and percent of patients without unplanned hospitalizations and no emergency department visits in panel B. The right y-axis corresponds to the index for each domain as standardized treatment effects in units of the standard deviation of the control group.

Figure 1.11: Event study of the effect of the Home Health Value-Based Purchasing program on coding manipulation.



Notes: The effects are treatment estimates with 95 percent confidence intervals from the difference-in-differences model (Equation 1) plus dummy variables for 2014 and 2016. The left y-axis corresponds to the individual agency-reported functional deficits among patients for ambulation, bed transferring, bathing, dyspnea, pain, and medication management. The right y-axis corresponds to the index for the six functional abilities as standardized treatment effects in units of the standard deviation of the control group. Panel A displays the differential effects of the program on the functional deficit levels of patients at the start of care. Panel B displays the differential effects of the program on the within-patient change in functional deficit levels between readmission for the subsequent episode and the discharge from the initial episode.

Figure 1.12: Event study of the effect of the Home Health Value-Based Purchasing program on non-targeted quality measures.



Notes: The effects are treatment estimates with 95 percent confidence intervals for each outcome from the difference-in-differences model (Equation 1) plus dummy variables for 2014 and 2016.

Another way that the parallel trends assumption may be violated is if the composition between the treatment and control agencies differed over time. This could occur if the patient population of agencies in the HHVBP program changed, such as if HHVBP agencies avoided sicker patients. If the patient population under the HHVBP were healthier, patients may appear to have more health gains (Murtaugh et al. 2008). I use agency-level average hierarchical condition category (HCCs) scores to measure patient health. HCCs are a commonly used risk score in the health services literature derived from Medicare fee-for-service claims data. Each HCC score is based on the beneficiary's age, sex, Medicaid eligibility status, original reason for qualifying for Medicare, institutional residential status, and the beneficiary's diagnoses from the prior year (Pope et al. 2011). Because the HCCs are derived from each patient's historical claims data encompassing a variety of health care providers, this measure is unlikely to be manipulated by home health agencies. I use the Home Health Public Use Files for years 2014 to 2016 to obtain agency-level average HCC scores (Appendix Figure A.1). Again, I find no evidence that patient composition changed over time.

1.6.2 Spillovers within chain organizations and controlling for covariates

One may also be concerned that spillover effects among agencies within one parent company, i.e., chain organization, could lead to contamination between treatment and controls, which would bias the treatment effects toward the null. For instance, a parent company with one agency in HHVBP and one agency not in HHVBP may choose to change its operations for both companies in response to the HHVBP program. The importance of spillover effects depends on how agencies have modified their behaviors under the HHVBP. If a chain organization developed a set of coding guidelines in response to the HHVBP, for instance, they could disseminate the guidelines to agencies in non-HHVBP states at minimal cost. However, spillover effects across agencies may be more limited, if the chain organization hired additional staff to provide more one-on-one patient care.

To determine whether spillovers within chain organizations may have biased the estimates toward the null, I exclude agencies in chains that spanned HHVBP and non-HHVBP states and I re-estimate Equations (1) and (2). Approximately 27% of the home health agencies in this study were part of a chain organization. Among these 2,208 agencies, 59% belonged to an organization that serves both HHVBP and non-HHVBP states. I use the Healthcare Cost Report Information System data from fiscal year 2015 to identify home health agencies that shared chain identifier numbers.

The treatment estimates are robust to excluding chain organizations that operated both HHVBP and non-HHVBP agencies. Similar to my main results, I find that home health agencies in HHVBP states improved on agency-reported outcome (standardized treatment effect=0.067, SE=0.023) and process measures (standardized treatment effect=0.070, SE=0.034) relative to home health agencies in non-HHVBP states, before and after the implementation of the program (Appendix Table A.9). I again find that home health agencies with marginal incentives to improve performed better on agency-reported outcome measures (standardized treatment effect=0.085, SE=0.024) and agency-reported process measures (standardized treatment effect=0.037, SE=0.021) (Appendix Table A.10).

The main estimates are also robust to controlling for time varying control variables and controlling for the marginal incentives to improve on all other measures in the program (Appendix Table A.9–A.10). The set of time-varying covariates include each agency's annual total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that were associated with 10 to 20% lower profit margins for home health agencies (Centers for Medicare & Medicaid Services). The measures of marginal incentives to improve are not time varying and are constructed from 2015 data; thus, it is unsurprising that the estimates are not affected.

1.6.3 Robustness of coding manipulation and non-HHVBP quality measures

In examining for evidence of coding practice change in my main analysis, I focus on agencies with at least 10 patient episodes in a given year. For functional deficit levels at the start of care, this restriction resulted in 8,241 out of 8,576 (96%) agencies. For the within-patient change in functional deficit levels for patients who were readmitted, this restriction resulted in 636 out of 5,750 (11%) agencies. This means that the external validity of the finding is potentially limited if the smaller sample is not representative of the broader population. Reassuringly, I do not find that the mean functional deficit levels vary by the number of patients per year, only the variance, suggesting that the smaller sample was representative of the larger sample and that agencies with larger denominators just had more reliable information. Nonetheless, I also estimate the effects of the program using a weighted regression, weighting by the number of patients included in each measure denominator for each agency. This approach allows me to include the entire sample while giving more weight to agencies with more reliable information.

I find consistent effects between the weighted regression and the restricted denominator approach used in the main analysis (Appendix Tables A.11). Functional deficits at start of care increased by 0.093 (SE=0.026) of a standard deviation of the control group and the change in functional deficits between discharge and readmission increased by 0.06 (SE=0.027) of a standard deviation. While these results are consistent between the two approaches, the weighted regressions showed evidence of violation of parallel trends (Appendix Figures A.2–A.3), making it the less desirable approach.

To discern whether the program had an effect on underlying quality, I also examine two non-HHVBP targeted measures: the effects of the HHVBP program on the rates that Medicare patients were discharged to inpatient institutions and the rates that they used additional episodes of home health care. However, home health agencies serve post-acute care patients who are recovering after an inpatient stay in addition to community dwelling, non-post-acute care patients. Unlike community dwelling patients, post-acute care patients are likely to recover and regain independence faster (Murtaugh et al. 2008). Therefore, measurable effects may be concentrated among post-acute care patients. As a sensitivity check, I examine these two non-HHVBP measures stratified by the type of patients. For both groups, I am unable to reject the null hypothesis that the HHVBP program had an effect (Appendix Tables A.12–A.13).

1.6.4 Alternative models to examine the effect of HHVBP on targeted quality measures

In this study's main approach, I estimate the effects of the program using both a double-differences and a triple-differences design. However, I also conduct a cross-sectional comparison of 2016 performance between HHVBP and non-HHVBP agencies, controlling for performance from 2015 and random assignment regional strata (see pre-analysis plan for more details). This alternative design does not rely on the assumption of parallel trends but requires that treatment and control agencies are comparable except for treatment assignment. Balance checks indicated that there were potentially important differences between the treatment and control group (Table 2). For instance, HHVBP agencies tended to perform better than non-HHVBP agencies on the agency-reported outcome measures. Furthermore, adjusting for covariates yielded smaller and sometimes directionally inconsistent effects, suggesting that differences between agencies may be confounding the estimates.

Because there appears to be evidence of imbalance in covariates between treatment and control groups, I also estimate the effects of the program on targeted quality using measure-specific propensity-score-matched samples. This approach assumes that matching removes unobserved differences between the treatment and control groups, conditional on the distribution of observable characteristics (Stuart 2010). I use 1:1 propensity-score matching without replacement to obtain a subset of agencies that were similar in the distribution of covariates that predict treatment assignment (Leuven & Sianesi, 2018). The covariates I use to create the matched samples included 2013 and 2014 lagged performance outcomes; rural status; ownership status; freestanding status; patient demographics (percent white, percent Hispanic, average age); percent of admissions that were discharged from acute care; percent of admissions from the bottom income quartile of each state; and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies (Centers for Medicare & Medicaid Services). A comparison of balance shows that propensity-score matching result in more similar groups of treatment and control agencies on pre-treatment characteristics than the unrestricted sample (see pre-analysis plan).

Both sets of cross-sectional designs yield results similar to the difference-in-differences model: in the aggregate, the HHVBP program led to improved performance on the agency-reported outcome and process measures and no statistically significant effects on the other non-agency-reported measures (Appendix Tables A.14–A.17). When examining the effects of marginal incentive size on quality performance, analyses using propensity-score matched samples with common support were generally consistent with the difference-in-difference-in-differences model results, although some estimates are sensitive to covariate adjustment suggesting that there may be confounders due to differences across agencies (Appendix Tables A.18–A.21). I observe no statistically significant effects using the unrestricted total sample (not propensity-score matched) comparing 2016 performance between HHVBP and non-HHVBP agencies (Appendix Tables A.22–A.25).

1.7 Conclusion

In this study, I examine whether a randomized value-based purchasing program implemented in the home health care sector led to improved quality performance. This study provides evidence that home health agencies responded to the incentives in the program—by achieving higher performance on agency-reported outcome and process measures where they faced marginal incentives to improve. However, the quality improvement attributed to the HHVBP was driven by coding manipulation as opposed to true improvements in care delivery.

This study produces a number of policy-relevant findings. First, I demonstrate that home health agencies responded to the incentives in the value-based payment program. Specifically, because the study quantified agencies' marginal incentives to improve and related the incentive size to quality performance, it shows that value-based payment programs affect performance through the mechanism of financial incentives. Furthermore, unlike studies of precursor value-based payment programs, these findings are estimated from a nationally representative experiment, which avoids many of the methodological shortcomings common in the value-based payment literature. These findings should inform the ongoing debate on whether value-based payment elicits responses from health care providers.

Second, this study examines whether agencies achieved true quality improvement, by assessing for coding manipulation under the HHVBP. Using detailed patient assessment data, I find that the HHVBP resulted in up-coding of functional deficits by home health agencies. Up-coding served two purposes: increase an agency's Medicare payments through the existing reimbursement system and raise an agency's apparent performance on the HHVBP. Combined with the finding that patients treated by HHVBP agencies were no less likely to be discharged to inpatient facilities nor less likely to use additional home health care, these findings suggest that the program increased up-coding rather than underlying quality. Furthermore, a back of the envelope calculation shows that coding manipulation explains the entirety of the HHVBP's apparent performance gains.

This study is limited in five major areas. First, while the HHVBP program has released information on the eligibility, inclusion, and scoring rules of the HHVBP, the exact agencies eligible for the program and their scores received are not released publicly. Reassuringly, the total number of HHVBP agencies in this study (N=1,626) and the CMS sponsored evaluation conducted by Arbor Research Collaborative for Health (N=1,587) is similar (Arbor Research Collaborative for Health and L&M Policy Research 2018). Furthermore, the scoring calculations are conducted in a consistent manner across all agencies. Therefore, any differences are unlikely to result in biased estimates. Moreover, my findings on the aggregate effects of the program on targeted quality measures are consistent with those from CMS' evaluators. While they only examined whether the program as a whole had an effect on the targeted measures and did not consider the heterogeneous incentives faced by agencies nor changes in coding practices, they also found that performance improved on the agency-reported measures and not on the non-agency-reported measures (Arbor Research Collaborative for Health and L&M Policy Research 2018).

Second, while examining marginal incentives from improvement provides proxy information on the incentives in the program, a full examination requires information on marginal costs as well as how quality is produced. If home health agencies vary in marginal costs from improvement, their decisions to respond are more likely to be driven by marginal returns from improvement rather than just marginal benefits.

Third, care quality is a notoriously difficult construct to measure and there are many ways one could discern quality improvement in the home health context. While the non-HHVBP measures used in this study (additional health care utilization following the end of a home health episode) are one way to gauge quality, these measures may not be sensitive to all care delivery improvements. Previous studies suggest, however, that patients with better functional abilities are less likely to require additional health care services (Tao and Ellenbecker 2013; C.-L. Li et al. 2011; Luppa et al. 2010). Furthermore, these non-HHVBP measures come from OASIS data which are not always uniformly reliable (Wolff et al. 2008) as agencies have some discretion in which assessment to file and how to complete individual fields on the assessments. These shortcomings of the OASIS data, however, are unlikely to affect the estimates obtained from this study, as these measures are not targeted by the HHVBP and thus reliability is unlikely to differ systematically between the treatment and control group.

Finally, while this study has assessed for evidence of coding manipulation, it does not assess whether there exists a dose-response relationship between incentive size and extent of manipulation. The two measures of up-coding used in this study together provide an understanding of the aggregate effects of the program on coding manipulation, but individually suffer from various disadvantages. While the first measure can be assessed across all home health agencies, it does not account for patient factors. This poses an issue because changes in the patient population among agencies within the treatment group cannot be reliably disentangled from changes in coding. While the second measures accounts for patient factors, it cannot be assessed across all home health agencies due to the relative infrequent occurrence of patients who end and then start a new episode of care. Thus, a full examination of a dose-response relationship between incentive size and up-coding would require using an alternative set of measures that is also sufficiently powered to address these relationships. Future research should focus on developing more robust measures of coding behavior to provide a better understanding of the relationship between incentive size and coding manipulation.

To improve the program, policymakers can consider several approaches to address coding manipulation. One policy solution is to remove rewards for agency-reported measures of quality and increase targeting of measures less prone to gaming like administrative-claims or patient-reported outcomes measures. Another is to down weight the agency-reported measures relative to non-agency-reported measures in the overall rewards calculations. A third is for CMS to consider increasing oversight of documentation practices. Finally, although these results suggest that the first year of the program did not achieve better underlying quality, future years of the program may yield more meaningful effects as the rewards more than double in size. With larger rewards, the marginal benefits from intervening on the underlying quality may make investing in true quality more worthwhile for agencies.

CHAPTER 2

Medicare's Home Health Star Ratings Program Has Not Shifted Patients into Higher-Rated Agencies

2.1 Introduction

The Home Health Star Ratings program is part of a prominent national movement to provide objective information to help consumers compare provider quality. In addition to the Home Health Star Ratings program implemented on July 2015, the Centers for Medicare & Medicaid Services (CMS) has also implemented separate star rating program for nursing homes, hospitals, dialysis facilities, physicians, and Medicare Advantage plans. The mechanisms of these programs to achieve better quality are simple and are predicated on patients choosing agencies based on ratings. In one mechanism, if patients can identify and choose high-quality agencies, then quality improves through reallocation. In another mechanism, if patients choose higher-rated agencies and agencies believe that demand exists for quality, then agencies will be incentivized to compete on quality, leading to better quality overall. To facilitate patient choice, the Star Ratings program uses a composite 5-star format. The format is designed based on the notion that an easy-to-read presentation improves information transparency and makes it easier for patients to choose (Faber et al. 2009; Peters et al. 2007; Hibbard and Peters 2003). Each agency is rated 1 to 5 stars in half-star increments every quarter. The rating is a summary score of an agency's performance on nine quality measures relative to other agencies in the nation (Centers for Medicare & Medicaid Services 2015a).

Evidence on the effects of the Home Health Star Ratings program on patient choice remains sparse. A small and growing body of literature suggests that the 5-star format has had modest success at eliciting consumer demand for highly-rated nursing homes and Medicare Advantage plans (Werner, Konetzka, and Polsky 2016; Perraillon et al. 2017; Werner et al. 2012; Darden and McCarthy 2015). However, the effects of the star ratings in the home health setting may differ from these other health care sectors. For instance, unlike institutional health care settings, home health patients are likely less able to anchor their perceptions of quality based on physical presence of the providers, so star ratings information in home health could potentially have greater impact. However, care for home health is delivered to patients at their homes, so switching out of an unsatisfactory agency may be less costly for patients, and thus, the added value of star ratings information may be lower. Therefore, whether patients actually use the home health Star Ratings information to choose higher-rated agencies is unknown.

My objective is therefore to assess whether the Home Health Star Ratings program succeeded in guiding Medicare patients toward higher-rated agencies – the essential rate-limiting step for the program to achieve better quality in the market. Using national data with a regression discontinuity design, this study compares agencies that are virtually identical but are barely on the opposite sides of an arbitrary star threshold (i.e., rounding of a continuous, underlying score to nearest half star), to assess whether having one more half star leads to more patient admissions for an agency. Furthermore, because patients choose agencies from those serving their residential area, patients' response to a particular rating may depend on the distribution of competing agencies within a local market. Thus, I further examine whether becoming one of the highest star-rated agencies in a ZIP code leads to more patient volume. If the star ratings increased patient demand for higher-rated agencies, then gaining one more half star or becoming the best within a ZIP code should lead to more patient volume.

In additional to examining the overall effects, I also assess whether patient demand for higher-rated agencies may differ across the star rating distribution. For example, a patient may place more value on a change from 3 to 3.5 stars, from average to above average, than from 1 to 1.5 stars, worst to slightly less bad. Specific to whether becoming one of the highest star-rated agencies in a ZIP code leads to more patient volume, I also test whether the effects differ for agencies that became the best star-rated option in a ZIP because other agencies became worse relative to the national distribution, if the effects differ for agencies became best with little to no quality change, or whether the effects differ depending on the number of competing agencies in a ZIP code.

2.2 Methods

2.2.1 Data source and study population

The study examines the universe of fee-for-service Medicare beneficiaries served by home health agencies in one of the 50 US states and Washington, DC that received a star rating from July 2015–December 2016. Patient data come from the home health Outcome and Assessment Information Set (OASIS), which includes service use dates and residential ZIP code information collected for all adult patients admitted to home health agencies (Centers for Medicare & Medicaid Services 2015b). I linked the OASIS data to the Master Beneficiary Summary File to identify patients' Medicare enrollment status, Medicaid enrollment status, race and ethnicity, sex, age, and end stage renal disease status.

I obtained the six home health star ratings and rating release dates from Home Health Compare. These ratings were released beginning in July 2015 on a quarterly basis and were calculated based on care outcomes provided to patients between October 1, 2013 through December 31, 2015. Unrounded ratings underlying the publicly displayed ratings in Home Health Compare were obtained through a special Freedom of Information Act data request to CMS.

I gathered the state and territory, years in operation, and for-profit status of the included agencies from Home Health Compare. I used the fiscal year 2014 and 2015 Healthcare Cost Report Information System to identify agencies affiliated with a chain organization in the year before each star rating release.

2.2.2 Outcome variables: New Medicare fee-for-service patients

My primary objective is to determine whether the star ratings program resulted in more patients using highly rated agencies. If patients use and respond to star ratings, then agencies with more stars should obtain more new patients. Therefore, I look at the number of new Medicare fee-for-service patients in the quarter following each star rating release in a quarter. I examine new patients to each home health agency because existing patients may be responding to prior experience as opposed to star ratings information. I focus on Medicare fee-for-service patients since they are given the freedom to choose any Medicare-certified provider and their choice of agencies are not constrained by insurance networks, which allows for a more accurate measure of patient choice in response to the star ratings.

To identify new patients admitted to each home health agency in each quarter following a star rating release, I include patients without prior use of the same home health agency in the year before the start of home health care. For each quarterly release, I then calculated the agency-level total number of new patients, excluding the day of release and up to, but not including, the subsequent date of release (Appendix Table B.1).

2.2.3 Statistical analysis: one more half-star and highest-ranked in ZIP code

2.2.3.1 One more half-star

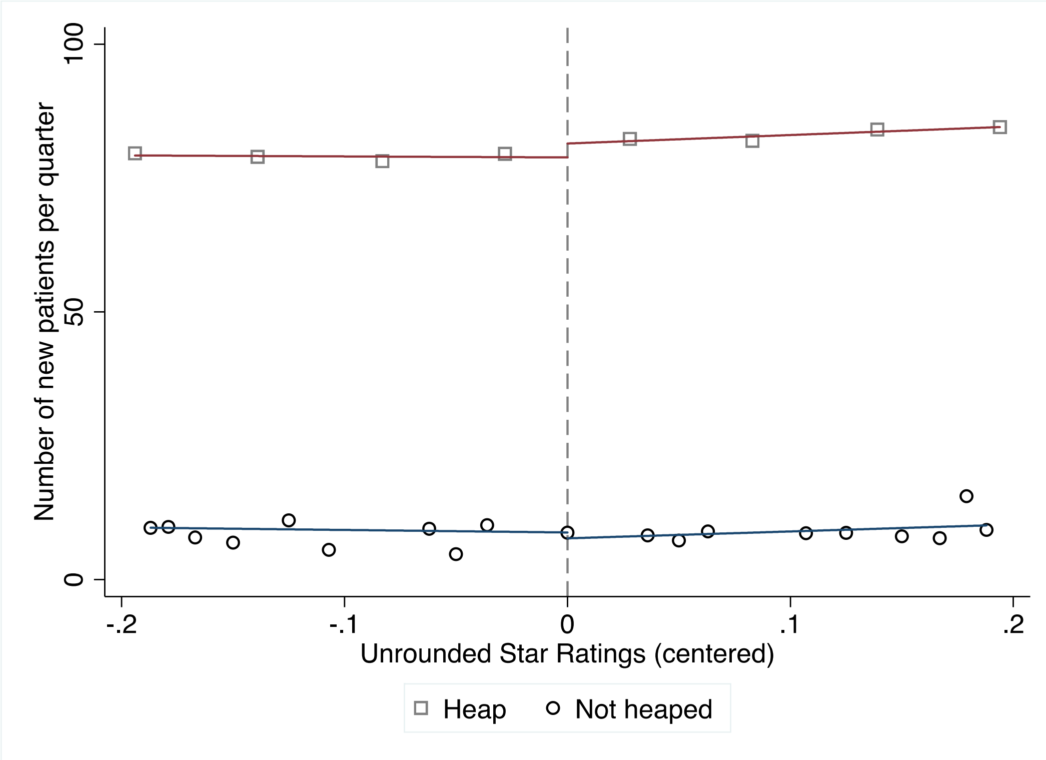
To isolate the effects of one more half-star on new patient admissions, this study uses a regression discontinuity design that leverages CMS's composite star ratings assignment rules. For each of the nine measures used to calculate the star ratings (Appendix Table B.2), CMS first gave each agency a measure-specific star rating, rounded to the nearest half decimal point. Next, CMS averaged the measure-specific stars to arrive at a composite, unrounded star rating. Unrounded star ratings were then rounded to the nearest half point to arrive at the final composite rating. For instance, an agency receives 2.5 stars if the agency's unrounded composite rating is 2.251 and receives 2 stars if its rating is 2.249. Therefore, as shown in Appendix Figure B.1, the discontinuity is the rounding threshold cutoff (gray dashed line), the running variable is the unrounded star ratings, and treatment is an additional half star.

I examine the effects of an additional half star on admissions in two ways. First, I combine the thresholds and six quarterly releases into one sample. Each unrounded star rating that is up to, but not including ±0.25 on either side of a threshold is included and centered at the rounding threshold. Combining the thresholds makes the assumption that the effects of one more half star are constant across the various thresholds (Cattaneo et al. 2016). Conceptually, the effects may be homogeneous since crossing the threshold always results in the same treatment (a half star more) and choosing an agency with a higher rating is presumably always better than an agency with a lower rating, all other factors equal.

However, I also test for presence of heterogeneous responses across thresholds. Thus, in the second way, I examine each threshold separately to determine whether having an additional half-star yield different effects across the distribution of the stars. Heterogeneity may be plausible if, for example, patients place more value on a change from 3.5 to 4 stars, from average to above average, than from 1 to 1.5 stars, worst to slightly less bad. For each threshold, agencies with unrounded star ratings of ±0.25 on either side of a threshold is included and centered at each rounding threshold.

By construction the unrounded star ratings will cluster around certain values because CMS takes the average of rounded numbers (Appendix Figure B.2). Furthermore, for each agency, only measures with sufficient patients in the measure denominators were used to calculate the overall ratings. Thus, the overall star ratings for smaller agencies tended to reflect fewer measures and this also leads to patterns in the running score. Thus, the unrounded star ratings displayed presence of heaping and the heaped unrounded star ratings tend to belong to agencies of larger size (Figure 2.1). Combining heaped and non-heaped points may induce bias when heaping is non-random (Barreca, Lindo, and Waddell 2016). This can be an issue because there may be a disproportionate share of heaped versus non-heaped observations across the treatment and control groups. As evidenced in Appendix Figure B.2, the non-heaped observations tend to skew toward lower values of the unrounded star ratings, suggesting that non-random heaping may induce bias if ignored. Thus, following the approach outlined by Barreca, Lindo, and Waddell (2016), I separately conduct the combined threshold analysis and threshold-specific analyses by heap and non-heap points, where the non-heap points are unrounded star ratings that occur less than 1 percent of the time.

Figure 2.1: Relationship between home health star ratings and number of new patients per quarter pooling all thresholds.



Notes: Pooled threshold sample includes agencies with unrounded star ratings that are centered at the rounding threshold and up to, but not including ±0.25 on either side.

To model the effects, I use a parametric approach because the running variable is discrete as opposed to continuous (Lee and Card 2008). In my preferred specification, Equation (1), I use ordinary least squares regression to estimate the level shifts in the cross-sectional relationship between in the number of new patients per agency per quarter and an additional half star in rating (see Appendix for various specifications used in this study). This specification performed the best among other alternatives in terms of Akaike information criterion, such as adjusting for quadratic functions of the unrounded star ratings and allowing for different slopes across the rounding threshold (Appendix Tables B.3–B.5). It is also preferred due to small sample concerns and because higher-order polynomial specifications are subject to overfitting problems.

$$Pats_{jq} = \beta_0 + \beta_1 Star_{jq} + \beta_2 running_{jq} + X\beta_{jq} + \epsilon_{jq} \quad (1)$$

In this specification, $Pats_{jq}$ is the number of new patients per agency j in quarter q; $Star_{jq}$ is equal to 1 if the observation received a higher star rating (right of the threshold) or 0 otherwise; and $running_{jq}$ is the unrounded star rating for agency j in quarter q, centered at the rounding threshold. β_1 provides the marginal effect of having one more half-star at the rounding threshold.

I control for a number of agency-level covariates, $X\beta_{jq}$. This vector includes the number of new patients in the prior quarter, number of Medicare patients in the prior quarter, the number of measures used to calculate the star ratings, the number of measures used to calculate the star ratings. These variables are important because they reduce variance greatly. For example, size is highly correlated from one quarter to the next. I also include agency organizational characteristics (agency age, chain affiliation, for-profit status) and prior quarter Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease). Finally, I also include star rating release fixed effects. Standard errors were clustered at the home health agency level.

In sensitivity analyses, I test the robustness of the estimates without adjustment for covariates and with the data points further away from the threshold removed. To check for evidence of manipulation at the threshold, I test for balance in agency-level characteristics. I also formally test for manipulation of the unrounded star ratings using the Frandsen manipulation test for discrete running variables (Frandsen 2017). Similar to McCrary's test (McCrary 2008), the test proposed by Frandsen is based on smooth approximations to the running variable's distribution around the threshold. Finally, I also examine the relationship between treatment and outcome using negative binomial-2 count model to reflect the "count" nature of the dependent variable.

2.2.3.2 Highest-ranked option in ZIP code

In addition to examining the effects of having one more half-star on an agency's new patient volume across the nation, I extend the analysis by also assessing whether local factors matter. Because patients select home health agencies from their choice set specific to their residential locality and that patients may pay more attention to agencies that are the highest-ranked option within their ZIP code, this study also assesses whether agencies that become the highest-ranked option, between the first and second release of star ratings (October 8, 2015 release versus July 16, 2015 release), receive more patient volume. For instance, in a ZIP code where the highest-rated option is 4 stars, does becoming a 4-star agency make a difference on the agency's patient volume?

Using a difference-in-differences design, I compare new patient admissions among agencies in ZIP codes where the agencies become the highest ranked star option to agencies that do not become the highest ranked star option, before and after second star rating release. I focus on the change between the first- and second-star ratings release because it informs whether the program was effective during its early implementation period. This results in a duration of 83 days before the second set of star ratings and 111 days after the second set of star ratings. I restrict the agency-ZIP code observations to agencies that had at least 2 distinct Medicare patients in the prior year in a given ZIP code to capture each agency's service area (Porell, Liu, and Brungo 2006), ZIP codes with at least 2 home health agencies, and agencies that had a star rating in both quarters so that it is a balanced panel. To put these restrictions in context, the average patient in a US ZIP code in the year before the second set of star ratings is served by 12.77 (SD=29.07) agencies without any sample exclusions and 17.29 (SD=22.24) agencies following all exclusions.

I use ordinary least squares to obtain estimates of the effect of becoming the highest ranked option on new patient volume (Equation 2). The linear specification allows me to remove potential sources of confounding, namely time-invariant differences between the treatment and control groups and common time shocks.

$$Pats_{jzq} = \alpha_0 + \alpha_1 Quarter2_q + \alpha_2 Highest_{jz} + \alpha_3 Quarter2_q \times Highest_{jz} + X\beta_{jzq} + \epsilon_{jq} \qquad (2)$$

In this specification, $Pats_{jzq}$ is the number of new patients per agency j in ZIP code z in quarter q; $Quarter2_q$ takes the value of 1 for quarter 2 and 0 for quarter 1; $Highest_{jz}$ takes the value of 1 if an agency became the highest-ranked option within the ZIP code and 0 otherwise. α_3 provides the marginal effect of becoming the highest-ranked option within a ZIP code.

I include, $X\beta_{jq}$, a vector of time varying covariates. This set includes an agency-level variable that captures each agency's unrounded star ratings. It includes a ZIP-code level variable that captures the number of home health competitors in each ZIP code. It also includes characteristics of patients treated within each ZIP-agency pair: the percent of full Medicare-Medicaid dually eligible patients, percent of partial Medicare-Medicaid dually eligible patients, percent of patients discharged from an inpatient source in the past 14 days, percent of patients with end stage renal disease, percent female, percent white, percent black, percent Hispanic, mean age of Medicare patients, and percent of Medicare fee-for-service patients. I cluster standard errors at the home health agency level.

To discern whether there may be heterogeneous effects from becoming the highest-rated option, I examine several subsamples. First, I again examine whether the effects of becoming the highest-rated option differ depending on the distribution of star ratings in the first release. Second, I focus on agencies that had no change in star ratings between the two releases, meaning that the agencies that became best in a ZIP code did so because other agencies became worse relative to the national distribution. Third, I examine a sample of agencies that had a small change in their unrounded star ratings (within range of 0.111 unrounded star ratings) and had a star rating change between the first and second release. For the agencies in this sample, their ratings within the national distribution changed little (suggestive of small to no quality change) while their observable rating changed. The effect of becoming the highest-ranked option in this sample then provides an estimate of the effects of the stars as opposed to changes in quality. Finally, I allow the effects of becoming the highest-ranked option to vary by market competitiveness, measured as the total number of agencies serving each ZIP code. I focus on the bottom quartile of number of agencies per ZIP (seven or fewer agencies) and the top quartile of number of agencies per ZIP (more than 28 agencies).

In robustness checks, I use an additional quarter of baseline data to check for evidence of violation of parallel trends. I also test the robustness of the estimates by using negative binomial-2 models as the outcome is a count of patients.

2.3 Results

2.3.1 One more half-star

There was a total of 8,998 agencies after pooling the thresholds across the heaped sample and 1,818 agencies across the non-heaped sample (Table 1, Panel A).

Table 2.1: Characteristics of sample by star ratings and becoming the highest-ranked star option within across ZIP codes.

	Panel A				Panel B	
	Heaped (8,998 agencies)		Not Heaped (1,818 agencies)		Highest-ranked option in ZIP code (8,610 agencies)	
	1/2-star more	1/2-star fewer	1/2-star more	1/2-star fewer	Became highest	Not highest
N						
Number of observations	21,527	21,209	2,392	1,865	5,993	183,510
Number of agencies	8,270	8,246	1,358	1,194	855	8,542
Agency characteristics						
Number of measures, mean (SD)	9 (0)	9 (0)	8 (1)	8 (1)	9 (0)	9 (0)
Years in operation, mean (SD)	18 (13)	18 (13)	13 (11)	12 (11)	22 (14)	19 (13)
For Profit (percentage)	76	76	84	84	71	77
In chain organization (percentage)	19	18	5	5	47	40
Medicare Patient characteristics						
New fee-for-service patients, mean (SD)	85 (173)	81 (140)	9 (14)	9 (16)	4 (7)	3 (6)
Medicare patients, mean (SD)	288 (560)	278 (456)	80 (161)	80 (162)	12 (22)	10 (18)
Percent fee-for-service, mean (SD)	81 (18)	81 (18)	75 (28)	75 (29)	78 (28)	80 (28)
Mean age, mean (SD)	76 (4)	76 (4)	72 (7)	72 (7)	75 (8)	75 (9)
Percent female, mean (SD)	62 (9)	62 (9)	62 (14)	62 (14)	60 (30)	61 (31)
Percent white, mean (SD)	67 (30)	67 (30)	47 (34)	48 (34)	79 (32)	71 (37)
Percent black, mean (SD)	17 (23)	17 (22)	32 (32)	31 (32)	11 (24)	15 (29)
Percent Hispanic, mean (SD)	12 (22)	12 (21)	16 (25)	15 (23)	7 (21)	10 (25)
Percent full dual enrolled, mean (SD)	32 (24)	32 (24)	46 (29)	45 (30)	23 (30)	27 (34)
Percent partial dual enrolled, mean (SD)	8 (8)	8 (8)	12 (14)	12 (14)	7 (17)	7 (17)
Percent inpatient admission, mean (SD)	39 (21)	39 (21)	25 (21)	25 (21)	51 (33)	47 (34)
Percent ESRD patients, mean (SD)	4 (4)	4 (4)	5 (8)	5 (6)	3 (11)	4 (13)
Number of agencies in ZIP, mean (SD)					14 (17)	23 (26)

Notes: Heap points are observations with unrounded star ratings that occur at least 1 percent of the time and not heaped are observations with unrounded star ratings that occur less than 1 percent of the time.

Agencies with one more half star were similar to agencies with one half star fewer with respect to organizational characteristics and patient characteristics, although agencies with one more half star had fewer patients on average than those with one half star fewer.

A comparison of the characteristics of the heaped to non-heaped samples suggest that there were systematic differences across these two types of agencies and there is a potential need to adjust for compositional differences at heap points in the regression discontinuity analysis. Heaped agencies had been in operation longer (an average of 18 years versus 13 years), had more measures (9 versus 8), smaller percentage of for-profit agencies (76 versus 84), and more agencies belonging to a chain organization (19 percent versus 5 percent). Agencies in the heaped sample tended to be larger than those in the non-heaped sample. At baseline, there were almost nine times more new patients treated by agencies in the heaped sample and almost four times more Medicare patients. Patients treated by agencies in the heaped sample also differed from those treated by agencies in the non-heaped sample in a number of ways. Patients treated by agencies in the heaped sample tended to be older (76 versus 72 years), more likely to be white (67 percent versus 48 percent), and more likely to have had a preceding inpatient discharge (39 percent versus 25 percent). There were fewer patients that were black (17 percent versus 32 percent), Hispanic (12 percent versus 16), full Medicare-Medicaid dual enrollees (32 percent versus 46 percent), and partial Medicare-Medicaid dual enrollees (8 percent versus 12 percent).

The regression discontinuity estimates from pooling all thresholds indicate that the effect of receiving one more half star on the number of new patients was small and non-significant for both heaped and non-heaped samples (Table 2.2). Gaining a half star was associated with a decrease of 0.57 patients (SE = 0.80), or less than 1 percent from a baseline of 83 patients, in the heaped sample and an increase of 0.34 patients (SE=0.66), less than 4 percent from a baseline of 9 patients, in the non-heaped sample. The results were similarly not statistically significant and small in magnitude when the sample was restricted to narrower windows around the cutoff.

Effects of having one more half star across each of the individual thresholds also suggest small and non-significant effects. For the heaped sample, the point estimates suggest that having one more half star is associated with a change of approximately 1 percent in new patients for most of the thresholds. The exceptions are the 1 versus 1.5 and 1.5 versus 2 stars, where the effects were larger (8 percent decrease and 3 percent increase), although imprecisely estimated. The threshold-specific point estimates in the non-heaped sample are larger in magnitude but noisy. Gaining a half-star from 3.5 to 4 stars, for instance, is associated with a 32 percent increase in new patient volume.

Table 2.2: Regression discontinuity estimates of the effects of gaining a half star on patient admissions.

		Heap			Not Heaped		
	Covariates	N agencies	Prior Quarter New Patients, mean (SD)	Treatment Estimate (SE)	N agencies	Prior Quarter New Patients, mean (SD)	Treatment Estimate (SE)
Pooled (-0.25, 0.25)							
		8,998	83 (157)		1,818	9 (15)	
	No			2.59 (3.18)			-1.16 (1.00)
	Yes			-0.57 (0.80)			0.34 (0.66)
By threshold [-0.25, 0.25]							
1 vs. 1.5		174	17 (18)		112	8 (10)	
	No			17.05 (11.26)			0.22 (2.53)
	Yes			-1.47 (5.59)			-0.13 (1.79)
1.5 vs. 2		1,086	27 (38)		413	7 (10)	
	No			3.09 (3.63)			-1.11 (1.20)
	Yes			0.81 (1.20)			-0.12 (0.83)
2 vs. 2.5		2,592	39 (61)		602	7 (10)	
	No			4.00 (2.67)			0.58 (1.56)
	Yes			-0.28 (0.93)			-0.64 (1.03)
2.5 vs. 3		4,053	69 (179)		575	9 (12)	
	No			8.75 (4.86)			-1.91 (1.78)
	Yes			-0.99 (1.20)			0.53 (1.04)
3 vs. 3.5		4,713	94 (165)		436	10 (14)	
	No			6.88 (6.32)			1.18 (2.48)
	Yes			-0.96 (1.45)			0.87 (1.93)
3.5 vs. 4		4,097	108 (182)		297	12 (18)	
	No			3.16 (7.67)			0.71 (3.93)
	Yes			0.96 (2.08)			3.84 (2.57)
4 vs. 4.5		2,582	99 (147)		172	14 (26)	
	No			-1.77 (8.19)			-13.86 (7.19)
	Yes			1.00 (1.97)			-8.89 (4.49)
4.5 vs. 5		1,183	71 (109)		76	16 (36)	
	No			-5.25 (7.72)			-7.35 (13.40)
	Yes			-0.93 (1.81)			23.45 (14.75)

Notes: Estimates obtained from ordinary least squares regression assessing level shifts in the cross-sectional relationship between in the number of new patients per agency per quarter and an additional half star in rating. Covariates include the number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Standard errors were clustered at the home health agency level.

The results remained not statistically significant and with small magnitudes when estimated without covariate adjustment and when the outermost data points were removed (Appendix Table B.6). I find no evidence of imbalance across the covariates for the heap sample and some evidence of violations in the non-heaped sample (the number of measures used to calculate the star ratings, beta=0.17 and SE=0.06, and prior quarter new patients, beta = –1.78 and SE=0.91) (Table 2.3). Furthermore, a formal test for evidence of manipulation of the unrounded star ratings when the ratings are discrete using the Stata command *rddistestk* with k=0 for heaped data and k<0.001 indicate no evidence of manipulation, with p-values exceeding 0.90 in both cases (Frandsen 2017). Finally, estimates were similar when using negative binomial-2 models as opposed to ordinary least squares (Appendix Table B.7).

Table 2.3: Balance tests across the star rating threshold.

Covariate	Heap: N agencies	Heap: Treatment Estimate (SE)	Not heaped: N agencies	Not heaped: Treatment Estimate (SE)
Agency characteristics				
Number of measures	8,998	-0.002 (0.003)	1,818	0.17 (0.06)
Years in operation	8,998	0.42 (0.26)	1,818	0.03 (0.69)
For Profit	8,998	-0.01 (0.01)	1,818	0.00 (0.02)
In chain organization	8,664	0.00 (0.01)	1,353	0.02 (0.02)
Medicare Patient characteristics				
New patients	8,998	3.24 (3.16)	1,818	-1.78 (0.91)
Medicare patients	8,998	14.09 (10.39)	1,818	0.78 (8.64)
Percent fee-for-service	8,987	-0.35 (0.36)	1,810	-1.75 (1.85)
Mean age	8,987	-0.04 (0.09)	1,810	-0.26 (0.43)
Percent female	8,987	-0.09 (0.17)	1,810	-0.65 (0.82)
Percent white	8,987	-0.19 (0.60)	1,810	0.76 (2.19)
Percent black	8,987	0.37 (0.45)	1,810	-2.45 (2.05)
Percent Hispanic	8,987	-0.29 (0.44)	1,810	1.71 (1.49)
Percent full dually enrolled	8,987	0.22 (0.48)	1,810	-1.57 (1.88)
Percent partial dually enrolled	8,987	-0.03 (0.17)	1,810	1.38 (0.89)
Percent inpatient admission	8,987	0.03 (0.43)	1,810	0.93 (1.34)
Percent ESRD patients	8,987	-0.01 (0.08)	1,810	0.54 (0.45)

Notes: Estimates obtained from ordinary least squares regression assessing level shifts in the cross-sectional relationship between individual covariates and an additional half star in rating.

2.3.2 Highest-ranked option in ZIP code

There was a total of 8,610 agencies in the analysis of whether becoming the highest-ranked option in a ZIP code between the first- and second-star rating releases led to more new fee-for-service patients (Table 1, Panel B). There were 855 agencies that became the highest-rated star option within at least one of their ZIP codes served and 8,542 agencies that were not the highest-rated star option within their ZIP codes served.

Agencies that became the highest-ranked option in a ZIP code were generally similar to agencies that did not, although there were some differences. Agencies that became the highest-ranked option in a ZIP code tended to have been in operation for longer (22 years versus 19), more likely to be in a chain organization (47 percent versus 40), and were larger, with more new fee-for-service patients (4 patients versus 3 patients) and more Medicare patients (12 patients versus 10 patients) in the quarter before becoming the highest-ranked option. They also faced fewer competing agencies (14 versus 23) in the ZIP codes that they served.

The types of patients served by agencies that became the highest-ranked option were again similar to patients served by agencies that did not become highest-ranked. However, patients in agencies that became the highest-ranked option tended to be white (79 percent versus 71 percent), with a higher percentage of patients with a preceding inpatient stay (51 percent verses 47 percent), and a lower percentage of full Medicare-Medicaid dual enrollees (23 percent versus 27).

The difference-in-differences estimates suggest that becoming the highest-ranked star option in a ZIP code is not associated with more new patients (Table 2.4). Overall, agencies gain 0.092 (SE=0.088) new patients, or approximately 3 percent from a baseline of approximately 3.24 patients, after becoming the highest-ranked star option; however, the effect is not statistically significant.

The effects are more variable across various baseline star ratings and generally not statistically significant. For instance, becoming the highest-ranked star option in a ZIP code starting from 3 stars in the first quarter is associated with a gain of 0.03 (SE=0.23) new patients, or approximately 1 percent, but the effect is not statistically significant. For agencies that started from 4.5 stars in the first quarter, becoming the highest-ranked star option in a ZIP code is associated with a decrease of 0.57 (SE=0.20) or 17 percent in new patients.

Among observations with no star rating change between the first and second quarter, becoming the highest-ranked agency is also associated with a positive but statistically non-significant effect of 0.27 (SE=0.16) additional new patients. For observations with a star rating change but small change in unrounded star ratings, the effect is a statistically significant decrease of 0.43 (SE=0.20) new patients.

Finally, the effects remain statistically non-significant and small across ZIP codes with seven or fewer agencies (0.04 new patients, SE=0.13) and more than 28 agencies (–0.21 new patients, SE = 0.13).

Table 2.4: Effects of becoming the highest-ranked option within a ZIP code on number of new patients.

	N, Agency-ZIP pairs (agencies)	Prior Quarter New Patients, mean (SD)	Highest Rated, estimate (SE)	Quarter 2, estimate (SE)	Highest Rated x Quarter 2, estimate (SE)
Overall	189,503 (8,610)	3.24 (5.82)	0.25 (0.20)	0.946 (0.022)	0.092 (0.088)
Quarter 1 rating					
2	10,948 (730)	1.92 (3.18)	-0.38 (0.24)	0.311 (0.039)	-0.17 (0.68)
2.5	26,413 (1,454)	2.57 (4.73)	-0.21 (0.32)	0.628 (0.037)	-0.07 (0.22)
3	45,880 (1,975)	3.32 (5.93)	0.44 (0.39)	0.972 (0.058)	0.03 (0.23)
3.5	50,317 (1,997)	3.70 (6.43)	0.33 (0.40)	1.148 (0.047)	0.03 (0.18)
4	33,943 (1,345)	3.56 (6.23)	0.59 (0.35)	1.042 (0.047)	0.05 (0.18)
4.5	16,491 (719)	3.26 (5.76)	-0.69 (0.21)	1.073 (0.081)	-0.57 (0.20)
No rating change	133,897 (5,764)	3.41 (6.08)	0.47 (0.32)	1.023 (0.028)	0.27 (0.16)
Rating change	5,784 (224)	3.30 (5.53)	-0.92 (0.42)	0.97 (0.12)	-0.43 (0.20)
7 or fewer agencies	54,513 (4,827)	3.95 (7.09)	0.34 (0.31)	1.18 (0.03)	0.04 (0.13)
More than 28 agencies	46,746 (4,078)	2.06 (3.55)	-0.27 (0.17)	0.50 (0.03)	-0.21 (0.13)

Notes: All models control for linear terms for unrounded star ratings, number of home health competitors in each ZIP code, percent of full Medicare-Medicaid dually eligible patients in each ZIP-agency pair, percent of partial Medicare-Medicaid dually eligible patients in each ZIP-agency pair, percent of patients discharged from an inpatient source in the past 14 days in each ZIP-agency pair, percent of patients with end stage renal disease in each ZIP-agency pair, percent female in each ZIP-agency pair, percent white in each ZIP-agency pair, percent black in each ZIP-agency pair, percent Hispanic in each ZIP-agency pair, mean age of Medicare patients in each ZIP-agency pair, and percent of Medicare fee-for-service patients in each ZIP-agency pair. Standard errors are clustered at the home health agency level.

In robustness checks, I find no evidence of violation of parallel pre-trends (Appendix Table B.8) and results were generally similar when estimated with using negative binomial-2 models (Appendix Table B.9). For example, the marginal effect of becoming the highest-ranked option is 0.101 (SE=0.055) new patients. An exception is the effects when examining ZIP codes with seven or fewer agencies; in this instance, becoming the highest-ranked option is associated with a statistically significant increase of 0.13 new patients (SE=0.08).

2.4 Discussion

I find no evidence that having a higher star rating under the Home Health Star Ratings program meaningfully increased the number of new patients treated by home health agencies with higher star ratings in the next quarter. Estimates suggest that having one more half-star is associated with non-statistically significant decrease of 0.57 new patients per agency, from a baseline of 83 new patients per agency per quarter among larger agencies and a non-statistically significant increase of 0.34 new patients per agency among smaller agencies, from a baseline of 9 new patients per agency per quarter. Even for agencies that became the highest-rated option within ZIP codes served, the effects are similarly small in magnitude and again not statistically significant, at 0.092 more new patients per agency from a baseline of 3.24 patients per quarter. Results were similar across baseline star ratings, across different types of agencies, and across varying levels of market competitiveness. Together, these findings suggest that Medicare's star ratings program is limited in its effectiveness in guiding patients toward higher-rated agencies.

This study provides the first estimates of the effects of the star ratings in the home health context in the program's and finds non-statistically significant effects that are small in magnitude, around 1 to 4 percent, in the program's first year and a half. Previous investigators have found that providing publicly available quality report cards in the form of summary star ratings have led to increased patient demand in the nursing home sector and for Medicare Advantage plans. In nursing homes, two papers have found that higher-rated facilities increased patient admissions in the range of 2 to 6 percent (Werner, Konetzka, and Polsky 2016; Perraillon et al. 2017). The program for Medicare Advantage health plans was associated with similarly positive and significant effects on plan enrollment (Darden and McCarthy 2015; McCarthy and Darden 2017; Reid et al. 2013). For both nursing homes and Medicare Advantage, having higher ratings appeared to have the biggest effect for those with above average star ratings. In contrast, gaining a higher rating in the home health sector did not appear to have clear relationships across varying star thresholds. One reason for these differences between the home health and nursing home sectors may be because patients choosing home health selects from those that serve their residential area as opposed to a nursing home which attracts patients to its location. In other words, a higher rating does not necessarily translate to a higher rank within the patient's actual choice set. However, my results examining ZIP-code level rankings finds similarly lackluster responses.

There are several potential explanations for the lack of an observable effect of the program on patient demand. The low responsiveness may be because of preconceived notions of a low variance in quality of home health care. Conversations with several hospital discharge planners from various hospitals suggest that home health agencies are commonly perceived as similar in quality, unlike the case for nursing homes. Thus, consumers and/or their surrogates may believe that there is little to gain from choosing a higher-rated agency.

Additionally, the star ratings information may not be meaningful and therefore is distrusted by patients and/or their surrogates. While the measures chosen for the star ratings reflect the health outcome goals for home health care and are applicable to a substantial portion of home health patients and agencies (Appendix Table B.2), they may suffer from other disadvantages. From July 2015 through December 2016, eight out of the nine quality measures used to determine the star ratings were reported by home health agencies. Therefore, the star ratings may be vulnerable to gaming, and patients and their surrogates may not find the information useful in gauging true quality. Thus, the utility of the star ratings information may be low.

Furthermore, it is possible that patients do not respond to a half star difference, because, for example, they assume some inherent measurement error. Instead, they may only respond to larger differences. I find, however, that a simple linear regression examining the number of patients to star ratings is only associated with an increase of 0.91additional patients for each additional star. Additionally, there does not appear to be a monotonic relationship between star ratings and patient volume – patient volume is positively correlated with number of stars for up to 4 stars, but then is negatively correlated for highly rated agencies (4.5 to 5 stars) (Appendix Table B.10).

Finally, studies suggest that consumer awareness of publicly available reports on home health quality is low (Baier et al. 2015), which would hinder the program's effectiveness if a general lack of awareness persisted into the program's early years.

There are several limitations in this paper worthy of mention. First, this paper focuses on fee-for-service Medicare patients new to a given agency from July 2015 through December 2016. While focusing on fee-for-service Medicare patients has the benefit of eliminating constraints on the patients' choice set imposed by insurance network rules, it limits the generalizability of the findings to other populations. It is possible that other patients, such as younger patients or future generations of Medicare patients, may have different responses to the star ratings information. Furthermore, this study does not explore the reason behind lack of observable response and instead focuses on the first-order policy effects of the program: whether it shifted patients toward higher-rated agencies or not. It remains important to determine whether there is a perceived lack of value to the information, a lack of awareness of the existence of data, or general mistrust of the information.

These results are important for several reasons. Updating the information on a quarterly basis entails significant resources from the federal government. Thus, discerning whether it is achieving greater competition on quality in the market is essential for understanding the program's effectiveness. As home health care use increases with payment reform and an aging demographic that increasingly desires to age in place (Keohane et al. 2018; Medicare Payment Advisory Commission 2019b; Cuckler et al. 2018), finding ways to ensure patients are informed when choosing home health and that the market is competitive on quality will be crucial to achieving better quality care.

CHAPTER 3

Choosing Home Health Care on the Basis of Medicare Star Ratings: Is There Value in Picking the Best?

3.1 Introduction

Home health is one of the nation's fastest growing health care sectors, providing services to more than 3.4 million Medicare patients each year (Keohane et al. 2018; Cuckler et al. 2018; Medicare Payment Advisory Commission 2019b). Since July 2015, the Centers for Medicare & Medicaid Services (CMS) has released quarterly star ratings summarizing home health agencies' quality of care (Centers for Medicare & Medicaid Services 2018b). CMS designed the ratings to help patients differentiate agencies in terms of quality. To further help patients choose high quality care, CMS mandated in November 2019 that inpatient providers such as hospitals provide quality information to patients choosing post-discharge care (Centers for Medicare and Medicaid Services (CMS) 2019). Anecdotal evidence suggest that many hospitals rely on the star ratings information to fulfill this requirement. With increased reliance on the star ratings program to differentiate home health quality and a rapid proliferation of home health care use during the past decade, it has become critical that the star ratings provide meaningful information for consumers.

It is unclear, however, whether the star ratings contain meaningful information. One concern is that a majority of the measures used to construct the ratings are based on data collected and self-reported by the home health agencies to CMS. Thus, the star ratings may be constructed from data that can be manipulated by the home health agencies. Another concern is that these ratings are based on lagged quality, and outcomes of patients treated by an agency more than one year ago may be irrelevant one year later. A third concern is that the ratings may be inaccurate because they do not fully account for differences in patient health status across agencies. While the quality measures used to create the star ratings are adjusted for baseline observable differences between patients, there may be important unobservable factors omitted from the star ratings calculations. For example, if patients with low ability to manage their medical conditions disproportionately use higher quality agencies, then this sorting could lead to lower ratings for the agency. On the other hand, an agency could selectively admit more patients likely to yield better measures and artificially inflate its star ratings.

Given that the star ratings information is intended to guide patient choice, it is imperative that patient outcomes are better if they choose a highly rated home health agency. Therefore, I investigate whether being treated by a home health agency with the highest number of stars—a *top agency*—in a given patient's residential ZIP code matters for health outcomes. I examine a national sample of Medicare fee-for-service beneficiaries who received home health care from July 2015 through April 2016. I use a quasi-experimental instrumental variable method with variation in patients' proximity to the nearest top agency over four quarterly updates of star ratings. I assess quality using an objective measure of patient health: days patients spend alive without health care use in the 180 days after discharge from home health.

To further understand the effects of being treated by top home health agencies, I examine several subgroups of the patient population. First, I assess whether receiving home health care from a top agency improves patient health across two subgroups. One subgroup consists of ZIP codes where the top agencies have a maximum of 3.5 stars (national average or below) and a second consists of top agencies with 4 stars or more (above national average). These subgroups provide more nuanced information about how outcomes for patients in low- and high-quality markets vary. Further, I assess how patient outcomes are affected when treated by top agencies in ZIP codes where all other agencies are on average 1 star lower, and where all other agencies are on average lower by more than 1 star. In other words, does receiving treatment from agencies with more stars matter? Finally, I examine whether the effects of being treated by top agencies are experienced by both patients referred from an inpatient setting (i.e., post-acute), such as individuals discharged from a hospital, and those who came from the community. Community-entry patients may have a different trajectory of health than post-acute home health patients. Thus, it is important to understand whether the treatment effect differ for different types of patients.

3.2 Methods

3.2.1 Study population and data

This study used data from 100 percent Outcome and Assessment Information Set (OASIS) to identify patients who started and finished home health care from a star-rated home health agency between July 17, 2015 and July 4, 2016. Approximately 80 percent of home health agencies receive a rating each quarter (Centers for Medicare & Medicaid Services 2015a). The OASIS data set contains home health assessments, dates of service, and patients' residential ZIP codes. Patients were included if they were continuously enrolled in fee-for-service Medicare in the 180 days before the start of care and 180 days after the end of care. Patients were also included if they died after starting home health care. Furthermore, patients must have had a residential ZIP code within the 50 states and Washington DC.

The star rating of each home health agency was determined based on the rating assigned by CMS to each agency at the time of each patient's start of care. These quarterly star ratings were obtained from Home Health Compare. To avoid ambiguity on the timing of the rating updates, data from the exact day that star ratings were released were excluded.

To understand health care use outside of the initial home health episode, this study also examined several other data sources. I used the 100 percent MedPAR claims data to identify each patient's hospitalization, inpatient rehabilitation facility, and long-term care hospital use. I used the Minimum Data Set (MDS), which contains nursing home assessment information, to identify days each patient spent in a nursing home. Finally, I used the Medicare Beneficiary Summary File, which identifies each patient's date of death, enrollment status in Medicare and Medicaid, supplemental drug coverage, and demographics information, such as date of birth, sex, and race.

I supplemented these data with the 2010 US Census and the American Community Survey 2012-2016 file to understand variation across ZIP codes in terms of characteristics including population density, rurality, education, labor force participation and unemployment, household income, race, and languages spoken at home.

3.2.2 Outcome variables

The primary objective of this study is to determine whether patients treated by home health agencies that have the highest star ratings in their ZIP code have better outcomes than if they were treated by lower-rated agencies. The main way that I examine patient outcomes is by measuring the number of days a patient spends alive and without use of institutional or home health care in the 180 days following the end of home health care. I assess this outcome using instrumental variables and controlling for observable patient characteristics (more details below).

This outcome measure is similar to other quality measures used in recent literature that focus on the number of days patients spend in good health overall, as opposed to outcomes during specific acute episodes or events (Groff, Colla, and Lee 2016; Bell et al. 2019; Burke et al. 2019; Lynn and Brock 2019; Cornell et al. 2019). This measure of days spent alive and free from health care use is unambiguous in terms of welfare implications for patients and for society. For patients, it represents independence and being alive; for society, health care use reflects resource use. This measure also aligns with the explicit objectives of home health care, which is to maintain patients' independence, avoid premature institutionalization, and increase patients' ability to remain in the community (Ellenbecker et al. 2008; Luppa et al. 2010).

In addition to examining the overall days patients spend alive and without health care use, I break down the number of days spent by each of the components of the overall measure. I calculate days in institutions (hospitalizations, inpatient rehabilitation facilities, long-term care hospitals, and nursing homes), days in home health, and days in death. This allows me to determine which patient states are driving the results.

Finally, I also examine the number of days in the initial home health care episode to understand how top agencies differ from other agencies. I measure this initial investment by home health agencies because these treatment decisions could enhance or offset differences in resource use after the end of care.

3.2.3 Instrumental variable

The goal of this study is to assess whether patients treated by a top home health agency within their residential area get better outcomes than if they were treated by a lower-rated home health agency. The ZIP code is the smallest unit of residential area for which patients can compare agencies on star ratings on Home Health Compare. The average ZIP code has 7 agencies serving it (i.e., there are at least seven different agencies treating at least one patient residing in that ZIP code). Hence, this study focuses on the options available to patients at the ZIP-code level.

One method to analyze whether outcomes vary by the type of agency is by comparing patients treated by top agencies to those treated by agencies with lower ratings, controlling for observable characteristics of patients. The matching of patients to home health agencies, however, is not random, and important unobservable characteristics of patients may confound effect estimates from a naïve comparison of conditional outcomes. Therefore, this study uses an instrumental variable approach, where the instrument approximates random assignment of patients to treatment (care from top home health agencies vs. all other home health agencies), conditional on covariates.

The instrument in this study is differential distance, which has been commonly used in the health services literature to assess both acute and post-acute care outcomes (Werner et al. 2019; Cornell et al. 2019). Distance was measured using the geodetic distance in miles between the centroid of two ZIP codes: patient's residential ZIP code and home health agency's administrative address. Differential distance was defined as the difference between the distance from a patient's ZIP code to the nearest top home health agency serving the patient's ZIP code and the distance from a patient's ZIP code to all other lower-rated home health agencies serving the patient's ZIP code. For example, in a ZIP code where patients are treated by a 1-star agency and two 3-star agencies, differential distance would be calculated as the distance from the patient's ZIP code to the nearest 3-star agency minus the distance from the patient to the 1-star agency. The differential distance is then dichotomized, such that patients who are closer to a top agency are given a value of 1 and patients who are closer to or equidistant to a lower-rated agency are given 0. For the average patient, the mean distance is 14.6 miles to the closest top agency and 6.9 miles to the closest non-top agency (Appendix Table C.1). Note that the star ratings are updated each quarter, and therefore, patients' differential distance may change quarterly as well.

There are two assumptions that must be met for the instrumental variables approach to be valid. First, the instrument (distance) must be correlated with the treatment of interest (care from top agency). Qualitative evidence suggests that distance is correlated with patient's choice of home health care (Baier et al. 2015). When faced with a dearth of information, patients sometimes choose agencies that are closest to them, even though home health services are delivered to patients at their homes at no additional cost (e.g., travel or out-of-pocket) and not provided at the agency's administrative address. This correlation is also observed in data. Patients were disproportionately treated by agencies located in their residential ZIP code. For example, 10.7 percent of patients in the sample were treated by agencies located in the patient's ZIP code, while only 4.2 percent of agencies serving the patient's ZIP code has an administrative address in the same ZIP code (Appendix Table C.2). Finally, first-stage results also suggest that proximity was strongly associated with treatment by a top home health agency (F=1942), and exceeds the standard benchmark of an F-statistic of 10 (Staiger and Stock 1997). The likelihood of receiving care from a top home health agency was 23.72 percentage points higher for patients living closer to a highest-rated agency than not (Appendix Table C.3).

Second, the instrument must not be associated with the outcome except through treatment assignment. First, I argue that on institutional grounds, distance to an agency should not affect health outcomes directly as care is delivered to patients at their homes. However, there may still be a correlation between distance and patient outcomes if, for example, high-income patients choose to locate themselves in the neighborhood of good home health care or if high quality home health agencies choose to locate near high-income patients (Y. Wang et al. 2017). In one way to assess whether this assumption is likely violated, I compare the association between the instrument and observable characteristics across patients, conditional on residential ZIP codes. The instrumental variable approach yields more balanced characteristics, particularly after conditioning on ZIP codes, indicating the need to control for ZIP code fixed effects. In another way, I conduct a falsification test to assess whether the instrument is correlated with drug coverage that could plausibly influence patients' health outcomes. Evidence of strong correlations between the instrument and other potential mechanisms (e.g., presence and quality of drug coverage) that influence health outcomes would suggest biased estimates. Reassuringly, I find a first-stage F-statistic of 0.27, providing suggestive evidence that the second criterium is met.

3.2.4 Statistical analysis

I first tested for differences in outcomes between treatment by top agencies and lower-rated agencies using multivariable ordinary least-squares regression.

$$Outcomes_{itz} = \delta_0 + \delta_1 D_{itz} + X_i\gamma + \theta_z + \epsilon_{itz} \quad (1)$$

In equation 1, $D_{itz} \in \{0,1\}$ equals 1 if patient i was treated by a top agency in quarter t and 0 if the patient was treated by a non-top agency in the ZIP code z in quarter t; $Z_{itz} \in \{0,1\}$ equals 1 when the patient in quarter t lives closer to a top agency in ZIP code z and 0 for living equidistant or closer to any other agency in the ZIP code. $X_i\gamma$ is a vector of patient-level characteristics, which include age, sex, race/ethnicity, full and partial Medicare and Medicaid eligibility status, end-stage renal disease status, whether their home health care was preceded by an inpatient stay, functional impairment at the start of home health care, and number of days spent in acute care, long-term care hospitals, inpatient rehabilitation, nursing homes, and home health in the 180 days before the start of home health. This regression also includes θ_z, a vector of ZIP code fixed effects. Adjusting for ZIP code fixed effects controls for potential differences between patients who live closer to higher-rated agencies than those who live farther for reasons that may be correlated with health outcomes. Errors are clustered at the ZIP code level.

In the main approach, using 2-stage least-squares regression, I first predict the type of agency the patient in ZIP code z uses in quarter $t \in \{1,4\}$:

$$D_{itz} = \alpha_0 + \alpha_1 Z_{itz} + X_i\gamma + \theta_z + \epsilon_{itz} \quad (2)$$

$D_{itz} \in \{0,1\}$ equals 1 if patient was treated by a top agency in quarter t and 0 if the patient was treated by a non-top agency in the ZIP code z in quarter t; $Z_{itz} \in \{0,1\}$ equals 1 when the patient in quarter t lives closer to a top agency in ZIP code z and 0 when equidistant or closer to a non-top agency. $X_i\gamma$ is the vector of patient characteristics described above and θ_z is a vector of ZIP code fixed effects. Errors are clustered at the ZIP code level.

In the second stage, I estimate the following to predict patient outcomes (e.g., the number of days a patient spends alive without health care in the 180 days after discharge from the initial home health episode):

$$Outcomes_{itz} = \beta_0 + \beta_1 \widehat{D}_{itz} + X_i\gamma + \theta_z + \epsilon_{itz} \quad (3)$$

3.2.5 Additional analyses

To provide a fuller understanding of the effects, I re-estimated the 2-stage least squares regression across various subpopulations: 1) I examined the effects stratified by the star value of the top home health agencies in a ZIP code, separately focusing on ZIP codes with top ratings at or below the average across the nation (1 to 3.5 stars) and ZIP codes with top ratings above the average across the nation (4 to 5 stars). This sub analysis sheds light on whether effects differ by the overall quality of agencies in a market; 2) I assessed the effects stratified by ZIP codes with a large difference in ratings between the top agencies and non-top agencies, and by ZIP codes with a small difference in ratings between the top agencies and non-top agencies. This stratification provides insight into whether health outcomes differ when patients are treated by agencies with more stars. The median differential in ratings is 1 star. Subsequently, I divided the sample into ZIP codes with rating differentials less than 1 star and ZIP codes with differentials of at least 1 star. I only examine ZIP codes where the top agencies have ratings between 3.5 and 4.5 stars to ensure common support across the rating differential; 3) I assessed whether the effects differed by home health patients who were community-entry versus post-acute (i.e., referred to home health from a hospital, skilled nursing facility, inpatient rehabilitation facility, or long-term care hospital). Previous research has found that community-entry patients differ from post-acute care patients (Murtaugh et al. 2008; WWAMI Rural Health Research Center 2018; Wysocki and Cheh 2019). Community-entry patients tend to have lower functional abilities, higher cognitive impairments, and are more likely to be dually eligible for Medicare and Medicaid (WWAMI Rural Health Research Center 2018; Wysocki and Cheh 2019). Thus, these two types of patients may have different outcomes.

I also tested the robustness of the results using alternative specifications of the instrument. In the first-stage equation, rather than using a dichotomized instrument, I use a linear measure of differential distance as well as a linear spline specification of the differential distance. Additionally, I examined whether using a linear model in the first-stage regression using the main dichotomized instrument to predict treatment probability was appropriate. I find that predicted probabilities were within bounds and I find high correlations ($\rho = 0.99$) between the predictions from the linear and nonlinear model.

Finally, because the estimates from the instrumental variable approach applies only to patients for whom treatment at a top agency was chosen solely because they lived closer to a top agency (known as compliers), and not for those who would always be treated by a top agency regardless of proximity (known as always takers), nor those who would never choose a top rated agency regardless of proximity (known as never takers), I profile the complier population using the methods described by Baiocchi, Cheng, and Small (2019). Their approach uses the variation from the first-stage regression across binary covariate groups to estimate the prevalence of each covariate for the complier population.

3.2.6 Limitations

There are several limitations in this study. While I assess whether treatment by the top agency affects patient outcomes, I do not quantify how much it matters for each unit increase in star ratings. However, for the average patient, making a decision about which home health care agency to select is constrained by the options available to them. Thus, focusing on the benefits of receiving treatment by a top agency relative to other options in each patient's choice set reflect the realistic situations in which patients make their choice.

Relatedly, the estimates that I obtain from this study are average effects among compliers—patients who are borderline indifferent to which agencies they are treated by. This raises two potential issues. One is of averages. Average effects may not be true for all patients. To address this limitation, this study explores the possibility of heterogeneous effects across different contexts for compliers through a variety of subgroup analyses.

The second issue is one of selection (Kowalski 2018, 2016). Selection is also sometimes framed as an external validity problem because of potential heterogeneous selection into treatment based on unobserved factors, such as if patients select into treatment based on whether they expect to gain from receiving treatment. I test for evidence of selection and quantify the directionality and magnitude, using the approach by Black, Joo, LaLonde, Smith, and Taylor (Black et al. 2017). This approach tests the null hypothesis that the difference in untreated outcomes between compliers and never takers of treatment is equal to zero. Because these two populations do not receive treatment, a difference in their outcomes reflect a difference in selection (not of treatment effects). I find that never users spend 2.07 more days alive and without care than untreated compliers (p=0.011), indicating that compliers are adversely selected. In other words, my estimates may be overstating the true effects of being treated by top-rated agencies by approximately 2.07 days; however, this selection bias is 55 percent of the estimated effect found in the study, indicating that some but not all of the effect could be due to selection. Similarly, to examine evidence of selection between always takers and compliers who receive treatment, I compare the difference in treated outcomes between always takers and compliers. I find no differences between the always takers and compliers that are statistically or substantively meaningful.

Another limitation of this study is that it focuses on the first year of the program and that the results may not generalize to future years of the program. This is particularly relevant if the star ratings lead to lower variance in actual quality among agencies, not just the CMS imposed distribution of quality ratings. If the variance of actual quality decreases, then it would be reasonable to expect that the differential gains from receiving treatment by top agencies also decrease over time.

3.3 Results

There were 1,870,017 Medicare home health patients residing in 22,332 ZIP codes from July 17, 2015 and July 4, 2016 in my cohort. Compared with patients treated by a non-top agency, those treated by a top agency were older (0.081 [SE=0.034] years), less likely to be black (–0.14760 [SE=0.064] percentage points), and more likely to be non-white, non-black, and non-Hispanic (0.234 [SE=0.042] percentage points) (Table 3.1).

Patients treated by top agencies were also less likely to be fully dual eligible for Medicare and Medicaid (–0.79 [SE=0.10] percentage points), less likely to have end-stage-renal disease (–0.118 [SE=0.035] percentage points), more likely to have been admitted from an inpatient setting

(2.32 [SE=0.17] percentage points), and had spent fewer days in institutional care (–0.321 [SE=0.071] days) and home health (–3.12 [SE=0.13] days) in the 180 days before the start of home health. Finally, except for a lower likelihood of having a pressure ulcer (–0.144 [SE=0.038] percentage points), patients treated by top agencies tended to have more health and functional limitations as documented by the home health agency at the start of care compared to patients treated by a lower rated agency.

Patient characteristics were more similar and with smaller magnitude of differences when compared across the instrument, suggestive that the dichotomized differential distance resulted in a more balanced sample between the treatment and comparison groups. Nonetheless, some differences remain. For instance, patients living closer to a top agency were still less likely to be fully dual eligible for Medicare and Medicaid (–0.42 [SE=0.14] percentage points), more likely to have been admitted from an inpatient setting (1.67 [SE=0.28] percentage points), and had spent fewer days in nursing homes (–0.262 [SE=0.082] days) and home health care (–3.00 [SE=0.44] days) in the prior 180 days. At the start of home health, patients closer to top agencies had fewer issues with urinary and bowel incontinence than patients living equidistant to farther from a top agency.

Table 3.1: Characteristics of patients using home health care in study cohort.

	Actual treating agency (1,870,017 patients 22,332 ZIP codes)				**Instrumental Variable** (1,870,017 patients 22,332 ZIP codes)			
	Lower-rated agency	**Top vs. Lower-rated agency**			**Far/equidistant to top agency**	**Close vs. far/equidistant to top agency**		
	Mean	Difference		*p*-value	Mean	Difference		*p*-value
Demographics								
Age	77.0	0.081	(0.034)	0.016	76.9	-0.028	(0.041)	0.488
Female (%)	62.0	-0.09	(0.10)	0.387	62.0	-0.04	(0.17)	0.809
White (%)	79.0	-0.023	(0.082)	0.778	79.2	0.09	(0.12)	0.438
Black (%)	11.4	-0.147	(0.064)	0.021	11.3	-0.028	(0.092)	0.759
Hispanic (%)	6.3	-0.064	(0.042)	0.126	6.1	-0.015	(0.065)	0.821
Other race/ethnicity (%)	3.4	0.234	(0.042)	0.000	3.4	-0.047	(0.059)	0.428
Medicaid coverage								
Full dual-eligible (%)	18.6	-0.79	(0.10)	0.000	18.4	-0.42	(0.14)	0.003
Partial dual-eligible (%)	5.7	0.029	(0.048)	0.541	5.8	-0.160	(0.079)	0.044
End-stage renal disease (%)	3.1	-0.118	(0.035)	0.001	3.1	0.001	(0.059)	0.985
Referred from inpatient institution (%)	63.0	2.32	(0.17)	0.000	63.3	1.67	(0.28)	0.000
Health care use before start of care (days)								
Any institutional care	15.5	-0.321	(0.071)	0.000	15.4	-0.29	(0.10)	0.003
Acute care hospital	5.4	-0.050	(0.019)	0.008	5.3	-0.038	(0.030)	0.200
Long-term care hospital	0.3	-0.0262	(0.0062)	0.000	0.3	-0.001	(0.010)	0.890
Inpatient rehabilitation	0.4	0.0358	(0.0079)	0.000	0.5	0.012	(0.010)	0.252
Nursing home	9.5	-0.281	(0.064)	0.000	9.4	-0.262	(0.082)	0.001
Home health	16.8	-3.12	(0.13)	0.000	16.3	-3.00	(0.44)	0.000
Health status at start of care								
Has pressure ulcer (%)	3.5	-0.148	(0.038)	0.000	3.6	-0.126	(0.065)	0.054
Shortness of breath (0 to 4)	1.3	0.1388	(0.0048)	0.000	1.4	-0.0035	(0.0040)	0.384
Urinary incontinence or catheter (0 to 2)	0.6	0.0212	(0.0018)	0.000	0.6	-0.0058	(0.0021)	0.006
Bowel Incontinence (0 to 5)	0.4	0.0108	(0.0034)	0.010	0.4	-0.0115	(0.0042)	0.006
Ability to dress upper body (0 to 3)	1.5	0.0737	(0.0029)	0.000	1.5	0.0047	(0.0029)	0.101
Ability to dress lower body (0 to 3)	1.8	0.0864	(0.0029)	0.000	1.9	0.0101	(0.0029)	0.000
Ability to bathe (0 to 6)	3.2	0.1513	(0.0050)	0.000	3.2	0.0216	(0.0051)	0.000
Toilet transferring ability (0 to 4)	1.2	0.1006	(0.0039)	0.000	1.2	0.0064	(0.0037)	0.081
Bed transfer ability (0 to 4)	1.4	0.0856	(0.0035)	0.000	1.4	0.0091	(0.0030)	0.003
Ambulation ability (0 to 6)	2.6	0.1164	(0.0036)	0.000	2.7	0.0117	(0.0037)	0.002

Notes: Each row represents a separate regression that is adjusted for ZIP code fixed effects. Standard errors are clustered at ZIP code level. Measures of health status at start of care comes from information reported by home health agencies where larger values represent worse health.

Although the analyses control for ZIP code fixed effects, I still examine the characteristics of across ZIP codes to understand the context facing patient decisions. The average ZIP code in the sample was served by an average of 7.21 (SE=0.03) home health agencies (Appendix Table C.2). Of these, 1.489 (SE=0.004) agencies were in the top category with an average of 4.15 (SE=0.002) stars, approximately 1.035 (SE=0.002) stars more than the average star rating of the non-top agencies in each ZIP code.

ZIP codes with higher star ratings among their top agencies differed from ZIP codes with lower ratings among their top agencies. ZIP codes with higher star ratings among their top agencies were served by more agencies and had greater star differentials between the top and non-top agencies. For example, in ZIP codes where the top agency was rated 5 stars, the rating differential between the top and mean non-top agencies was an average of 1.537 (SE=0.003) stars. In comparison, the differential is 0.707 (SE=0.002) stars for ZIP codes where top agencies had 3.5 stars or fewer.

ZIP codes with higher star ratings among the top agencies also had different population characteristics than ZIP codes with lower ratings among top agencies (Appendix Table C.4). For instance, compared to ZIP codes with lower ratings among their top agencies, ZIP codes with top agencies ranked 5 stars had greater population density, more households with urban residence, a lower percentage of households that only spoke English, and greater income inequality. Together the differences at the patient- and ZIP code levels between those treated by top versus all other agencies confirm a need for adjustment in regression analyses.

Shifting to results, multivariable regression (Equation 1) suggests that patients spent significantly more days alive and without any care when treated by top agencies (Table 3.2). Being treated by a top agency resulted in 2.23 (95% CI, 2.01 to 2.44; $P < 0.0001$) more days alive and without any health care from a baseline of 144.8 days. The majority of the 2.23-day gain was driven by less time being deceased (–1.252 days; 95% CI, –1.40 to –1.10; $P < 0.0001$), using home health (–0.621 days; 95% CI, –0.76 to –0.49; $P < 0.0001$), and using a nursing home (–0.250 days; 95% CI, –0.32 to –0.18; $P < 0.0001$). Patients treated by a top agency also spent –2.69 days fewer in the initial home health episode (95% CI, –2.92 to –2.46; $P < 0.0001$), compared to a baseline of 50.7 days.

Table 3.2: Differences in outcomes after treatment by a top agency versus lower-rated agency.

	Multivariable regression				Instrumental Variable Regression			
Outcome Variable	Mean	Coefficient (SE)	*p*-Value	95% CI	Mean	Coefficient (SE)	*p*-Value	95% CI
Days in initial home health episode	50.7				50.4			
No covariates		-4.34 (0.22)	0.000	(-4.78, -3.91)		-15.16 (1.23)	0.000	(-17.57, -12.74)
Covariates, no FE		-2.08 (0.14)	0.000	(-2.36, -1.79)		-6.07 (0.78)	0.000	(-7.60, -4.55)
Covariates and FE		-2.69 (0.12)	0.000	(-2.92, -2.46)		-10.50 (0.94)	0.000	(-12.35, -8.65)
Days alive without health care	144.8				145.2			
No covariates		3.65 (0.15)	0.000	(3.36, 3.94)		9.99 (0.72)	0.000	(8.56, 11.41)
Covariates, no FE		2.53 (0.11)	0.000	(2.33, 2.74)		4.09 (0.50)	0.000	(3.11, 5.07)
Covariates and FE		2.23 (0.11)	0.000	(2.01, 2.44)		3.75 (0.79)	0.000	(2.20, 5.29)
Any days in an institutional setting	7.4				7.3			
No covariates		-0.373 (0.042)	0.000	(-0.46, -0.29)		-0.09 (0.19)	0.630	(-0.48, 0.29)
Covariates, no FE		-0.268 (0.039)	0.000	(-0.35, -0.19)		0.21 (0.18)	0.257	(-0.15, 0.56)
Covariates and FE		-0.354 (0.041)	0.000	(-0.43, -0.27)		-0.73 (0.30)	0.015	(-1.32, -0.14)
Acute care hospital	2.5				2.5			
No covariates		-0.128 (0.014)	0.000	(-0.15, -0.10)		-0.073 (0.067)	0.282	(-0.20, 0.06)
Covariates, no FE		-0.112 (0.012)	0.000	(-0.14, -0.09)		-0.013 (0.060)	0.824	(-0.13, 0.10)
Covariates and FE		-0.075 (0.013)	0.000	(-0.10, -0.05)		-0.101 (0.094)	0.279	(-0.28, 0.08)
Long-term care hospital	0.2				0.2			
No covariates		-0.0461 (0.0043)	0.000	(-0.05, -0.04)		-0.142 (0.019)	0.000	(-0.18, -0.10)
Covariates, no FE		-0.0392 (0.0041)	0.000	(-0.05, -0.03)		-0.104 (0.018)	0.000	(-0.14, -0.07)
Covariates and FE		-0.0208 (0.0044)	0.000	(-0.03, -0.01)		0.030 (0.032)	0.352	(-0.03, 0.09)
Inpatient rehabilitation facility	0.1				0.1			
No covariates		0.0010 (0.0045)	0.827	(-0.01, 0.01)		-0.025 (0.024)	0.306	(-0.07, 0.02)
Covariates, no FE		-0.0024 (0.0039)	0.535	(-0.01, 0.01)		-0.017 (0.021)	0.413	(-0.06, 0.02)
Covariates and FE		-0.0082 (0.0031)	0.007	(-0.01, -0.002)		-0.002 (0.024)	0.936	(-0.05, 0.04)
Nursing home	4.6				4.5			
No covariates		-0.200 (0.035)	0.000	(-0.27, -0.13)		0.15 (0.16)	0.357	(-0.16, 0.46)
Covariates, no FE		-0.115 (0.033)	0.000	(-0.18, -0.05)		0.34 (0.15)	0.020	(0.05, 0.63)
Covariates and FE		-0.250 (0.035)	0.000	(-0.32, -0.18)		-0.66 (0.25)	0.009	(-1.15, -0.17)
Days in home health	16.4				16.1			
No covariates		-3.22 (0.12)	0.000	(-3.45, -2.99)		-10.49 (0.62)	0.000	(-11.71, -9.28)
Covariates, no FE		-1.690 (0.072)	0.000	(-1.83, -1.55)		-5.06 (0.37)	0.000	(-5.79, -4.32)
Covariates and FE		-0.621 (0.069)	0.000	(-0.76, -0.49)		-2.91 (0.49)	0.000	(-3.87, -1.94)
Days deceased	11.4				11.4			
No covariates		-0.058 (0.074)	0.436	(-0.20, 0.09)		0.60 (0.33)	0.068	(-0.04, 1.25)
Covariates, no FE		-0.577 (0.074)	0.000	(-0.72, -0.43)		0.76 (0.34)	0.024	(0.10, 1.42)
Covariates and FE		-1.252 (0.078)	0.000	(-1.40, -1.10)		-0.11 (0.56)	0.846	(-1.20, 0.98)

Notes: In each model, the sample include 1,870,080 patients and 22,333 ZIP codes. Means for multivariate regression and instrumental variable regression are the unadjusted mean for each outcome for patients treated by non-top agencies and for patients located far from or equidistant to a top agency, respectively. Covariates include patient age, sex, race/ethnicity, full and partial Medicare and Medicaid eligibility status, end-stage renal disease status, whether their home health care was preceded by an inpatient stay, functional impairment at the start of home health care, and number of days spent in acute care, long-term care hospitals, inpatient rehabilitation, nursing homes, and home health in the 180 days before the start of home health. FE represents ZIP code fixed effects.

In instrumental variables regression, the point estimates suggest effects that are 44 percent larger than that observed in the multivariable regression. Patients treated by top agencies spent 3.75 days alive and without any health care (95% CI, 2.20 to 5.29; *P* <0.0001) from a baseline of 145.2 days. This was driven by 2.91 fewer days in home health care (95% CI, –3.87 to –1.94; *P* <0.0001) and 0.66 fewer days in nursing homes (95% CI, –1.15 to –0.17; *P*=0.009). Unlike the multivariable regression, the instrumental variable regression did not find evidence of an effect on the number of days deceased. Finally, patients treated by a top agency spent 10.5 days fewer in the initial home health episode (95% CI, –12.35 to –8.65; *P* <0.0001), approximately a 20 percent decrease from the baseline. The estimates were similar to sensitivity analyses using alternative specifications: with an instrument of linear differential distance and a linear spline specification of the differential distance (Appendix Table C.5).

The association between days spent alive without care and treatment by a top home health agency was further investigated in several additional subgroup analyses (Table 3.3). Across ZIP codes with at most 3.5 stars among the top agencies, the effect was non-significant (–0.75 days; 95% CI, –5.35 to 3.85; *p*=0.749). The effect among ZIP codes with top agencies with above average star ratings (4 to 5 stars) yielded a statistically significant effect of 6.51 more days that patients spent alive and without care (95% CI, 4.15 to 8.87; *P* <0.0001), from a baseline of 145.23 days.

In results that examines whether a larger star differential between the top and non-top agencies affect outcomes, I find that patients spent 7.80 more days alive and without health care use (95% CI, 4.13 to 11.47; *P* <0.0001) when treated by top agencies in ZIP codes with a larger star differential (Table 3.3). In contrast, the effects were smaller and non-significant for patients in ZIP codes with differentials less than 1 star (0.50 days; 95% CI, –1.82 to 2.83; *p*=0.671). This provides evidence that being treated by agencies with more stars matter for patient outcomes.

In the last subgroup analysis, I find that both post-acute and community-entry home health patients experienced a beneficial effect from being treated by a top agency, contrary to concerns that these patients may fare differently (Table 3.3). Patients that were admitted to home health care following an inpatient stay spent 3.40 more days alive and without health care use (95% CI, 1.80 to 5.00; *P* <0.0001) when treated by top agencies. The effects were primarily driven by decreased use of home health (–2.23 days; 95% CI, –3.11 to –1.36; *P* <0.0001) and nursing homes (–0.69 days; 95% CI, –1.22 to –0.17; *p*=0.010). Patients admitted from the community spent 5.60 days more days alive and without health care use in the 180 days following the end of the initial home health episode (95% CI, 2.30 to 8.89; *p*=0.001). This was driven by fewer days in home health (–4.94 days; 95% CI, –7.20 to –2.67; *P* <0.0001).

Table 3.3: Differences in outcomes after treatment by a top agency versus lower-rated agency across subgroups.

	Top star rating in ZIP code		Difference in ratings (top vs. others)		Home health patient type	
	1 to 3.5 stars (244,515 patients 8,775 ZIP codes)	4 to 5 stars (1,625,565 patients 18,215 ZIP codes)	<1 star (581,378 patients 13,223 ZIP codes)	1+ star (794,257 patients 14,517 ZIP codes)	Post-acute (1,199,955 patients 21,249 ZIP codes)	Community-entry (657,378 patients 18,890 ZIP codes)
Days in initial home health episode						
Unadjusted Mean (SE)	53.96 (0.51)	50.02 (0.20)	46.57 (0.25)	52.51 (0.28)	41.67 (0.11)	65.50 (0.32)
Coefficient (SE)	-2.90 (2.51)	-16.27 (1.47)	-3.39 (1.34)	-22.20 (2.45)	-8.51 (0.87)	-14.37 (1.74)
95% CI	(-7.81, 2.01)	(-19.15, -13.38)	(-6.02, -0.76)	(-27.00, -17.40)	(-10.20, -6.81)	(-17.79, -10.96)
Days alive without health care						
Unadjusted Mean (SE)	144.73 (0.26)	145.23 (0.12)	148.24 (0.14)	144.63 (0.15)	151.036 (0.079)	135.09 (0.18)
Coefficient (SE)	-0.75 (2.35)	6.51 (1.20)	0.50 (1.19)	7.80 (1.87)	3.40 (0.82)	5.60 (1.68)
95% CI	(-5.35, 3.85)	(4.15, 8.87)	(-1.82, 2.83)	(4.13, 11.47)	(1.80, 5.00)	(2.30, 8.89)
Days in an institutional setting						
Unadjusted Mean (SE)	7.374 (0.069)	7.333 (0.027)	7.141 (0.044)	7.564 (0.035)	7.190 (0.028)	7.591 (0.040)
Coefficient (SE)	1.03 (0.89)	-1.33 (0.45)	-0.48 (0.44)	-1.80 (0.71)	-0.97 (0.33)	-0.28 (0.65)
95% CI	(-0.71, 2.77)	(-2.22, -0.44)	(-1.34, 0.37)	(-3.19, -0.40)	(-1.62, -0.33)	(-1.56, 1.00)
Acute care hospital						
Unadjusted Mean (SE)	2.493 (0.023)	2.5133 (0.0098)	2.452 (0.015)	2.540 (0.013)	2.669 (0.011)	2.2391 (0.0103)
Coefficient (SE)	0.06 (0.28)	-0.16 (0.14)	-0.06 (0.14)	-0.01 (0.22)	-0.24 (0.11)	0.13 (0.17)
95% CI	(-0.49, 0.62)	(-0.43, 0.12)	(-0.34, 0.21)	(-0.44, 0.41)	(-0.45, -0.02)	(-0.21, 0.47)
Long-term care hospital						
Unadjusted Mean (SE)	0.1696 (0.0090)	0.1626 (0.0030)	0.1228 (0.0044)	0.1832 (0.0045)	0.1551 (0.0030)	0.1775 (0.0044)
Coefficient (SE)	-0.060 (0.081)	0.05 (0.05)	-0.01 (0.05)	0.03 (0.08)	-0.025 (0.036)	0.153 (0.068)
95% CI	(-0.22, 0.10)	(-0.05, 0.14)	(-0.11, 0.08)	(-0.12, 0.18)	(-0.09, 0.05)	(0.02, 0.29)
Inpatient rehabilitation facility						
Unadjusted Mean (SE)	0.1138 (0.0073)	0.1517 (0.0029)	0.1409 (0.0046)	0.1546 (0.0038)	0.1430 (0.0028)	0.1563 (0.0036)
Coefficient (SE)	-0.044 (0.063)	0.01 (0.04)	0.00 (0.04)	0.043 (0.055)	-0.020 (0.026)	0.027 (0.050)
95% CI	(-0.17, 0.08)	(-0.06, 0.08)	(-0.07, 0.08)	(-0.06, 0.15)	(-0.07, 0.03)	(-0.07, 0.12)
Nursing home						
Unadjusted Mean (SE)	4.598 (0.058)	4.505 (0.022)	4.425 (0.036)	4.686 (0.029)	4.223 (0.022)	5.018 (0.035)
Coefficient (SE)	1.07 (0.72)	-1.23 (0.38)	-0.41 (0.37)	-1.86 (0.61)	-0.69 (0.27)	-0.59 (0.56)
95% CI	(-0.34, 2.48)	(-1.98, -0.49)	(-1.13, 0.31)	(-3.04, -0.67)	(-1.22, -0.17)	(-1.70, 0.51)
Days in home health						
Unadjusted Mean (SE)	14.69 (0.23)	16.29 (0.12)	13.15 (0.11)	15.90 (0.12)	11.015 (0.048)	24.94 (0.20)
Coefficient (SE)	-1.45 (1.29)	-3.93 (0.77)	-1.13 (0.68)	-5.20 (1.18)	-2.23 (0.45)	-4.94 (1.15)
95% CI	(-3.98, 1.07)	(-5.45, -2.41)	(-2.46, 0.20)	(-7.51, -2.89)	(-3.11, -1.36)	(-7.20, -2.67)
Days deceased						
Unadjusted Mean (SE)	13.20 (0.12)	11.145 (0.045)	11.47 (0.07)	11.904 (0.061)	10.759 (0.045)	12.382 (0.081)
Coefficient (SE)	1.17 (1.65)	-1.25 (0.83)	1.11 (0.85)	-0.80 (1.32)	-0.19 (0.61)	-0.38 (1.18)
95% CI	(-2.06, 4.40)	(-2.88, 0.38)	(-0.54, 2.77)	(-3.39, 1.79)	(-1.38, 1.00)	(-2.70, 1.94)

Notes: Unadjusted means reflect mean outcomes among patients that are far/equidistant to the top agency. All regressions are covariate adjusted for patient covariates (age, sex, race/ethnicity, full and partial Medicare and Medicaid eligibility status, end-stage renal disease status, whether their home health care was preceded by an inpatient stay, functional impairment at the start of home health care, and number of days spent in acute care, long-term care hospitals, inpatient rehabilitation, nursing homes, and home health in the 180 days before the start of home health) and ZIP code fixed effects.

Finally, because the results apply to the marginal patient for whom closer proximity led to their treatment in a top agency, I compared the characteristics of these patients against the rest of the patient population in the sample to better understand the generalizability of the results (Table 3.4). I find that the compliers, consisting of approximately 24 percent of the total sample, were very similar to the full sample in terms of observable health characteristics. Any differences were slight. For instance, the compliers were more likely to be black (10.98% vs. 10.91%), more likely to be partially Medicare and Medicaid eligible (5.85% vs. 5.89%), less likely to have more severe bathing limitations (28.39% vs. 28.62%), and less likely to have spent more than 0.46 days in inpatient rehabilitation facilities in the 180 days before starting home health care. These profiles suggest that the complier population was similar to the rest of the population.

Table 3.4: Characteristics of complier population compared with patients in the full study sample.

	Full Sample (percent)	**Compliers (percent)**
Demographics		
Age > mean (76.89)	55.37	55.34
Female	61.97	61.98
White	80.08	79.95
Black	10.91	10.98
Hispanic	5.72	5.73
Other race/ethnicity	3.29	3.29
Medicaid coverage		
Full dual-eligible	18.03	18.06
Partial dual-eligible	5.85	5.89
End-stage renal disease	3.07	3.08
Referred from inpatient institution	64.22	64.21
Health care use before start of care (days)		
Any institutional care > mean (15.32)	27.96	27.87
Acute care hospital > mean (5.27)	30.68	30.60
Long-term care hospital > mean (0.27)	1.01	1.01
Inpatient rehabilitation > mean (0.46)	3.33	3.30
Nursing home > mean (9.33)	21.91	21.83
Home health > mean (15.43)	17.52	17.51
Health status at start of care		
Pressure ulcer	3.60	3.60
Severe shortness of breath	15.47	15.50
Urinary incontinence or catheter	55.48	55.60
Any bowel Incontinence	15.89	15.88
Low ability to dress upper body	50.44	50.45
Low ability to dress lower body	75.44	75.39
Low ability to bathe	28.62	28.39
Low toilet transferring ability	21.96	21.91
Low bed transfer ability	43.06	42.86
Low ambulation ability	62.52	62.26

Notes: Each continuous patient characteristic is dichotomized to compare against the mean value or above, or middle category or above of that characteristic. Severe shortness of breath: patients who are dyspneic with minimal exertion or at rest. Any bowel incontinence: frequency of less than once weekly or more. Low ability to dress upper body and lower body: patients who must receive help from others to put on clothing or entirely depend on others to dress. Low ability to bathe: patients who are unable to use the shower or tub or completely require bathing by another person. Low toilet transferring ability: patients who are unable to get to and from the toilet. Low bed transfer ability: unable to transfer self. Low ambulation ability: requires at supervision from another person at all times to walk or unable to ambulate.

3.4 Discussion

Among Medicare home health patients, receiving treatment by home health agencies with the highest star ratings in each ZIP code was associated with a 2.6 percent increase or 3.75 more days alive without use of health care in the 180 days following the end of the initial home health episode. These gains were the result of decreased use of home health and nursing home care. The benefit of being treated by a top agency as opposed any other agency was most pronounced for patients residing in ZIP codes where the top agencies had above average ratings, i.e. 4 to 5 stars, (6.51 more days alive and without care), and where there was a greater difference in ratings between the top and non-top agencies (7.80 more days alive and without care). Furthermore, both post-acute and community-entry home health patients appeared to benefit from treatment by a top agency, alleviating concerns that the ratings may be only relevant to healthier patients. Together, these results suggest that the Home Health Star Ratings program contain meaningful information for patients choosing home health care across a variety of contexts.

Other similarly structured star rating systems have been found to contain useful information for patients choosing nursing homes (Cornell et al. 2019) and hospitals (D. E. Wang et al. 2016). To my knowledge, this study provides the first large-scale estimates of the effects of being treated by higher-rated versus lower-rated home health agencies. Furthermore, this study examines outcomes while accounting for the choice constraints faced by patients in the real-world. Approximately 15 percent of patients reside in ZIP codes where the best home health option is only rated average when compared to the rest of the nation. Therefore, it is important to understand, from the perspective of the patient, what they can expect in terms of outcomes, if they were to choose the best option possible within their choice set. These results indicate that for most patients choosing among home health care agencies, there is value in selecting agencies with higher ratings and that future policies efforts should focus on how to encourage patients' use of this information.

These results have important policy implications. First, the Star Ratings program is intended to guide patients to better home health agencies. For that task, it is imperative to that the program provides good advice to most patients. Given that the star ratings contain meaningful information and the benefits are applicable to many patients in a variety of contexts, current practices in incorporating star ratings to guide patient choice in discharge planning is appropriate. Second, this paper raises the issue that there is heterogenous distribution of agencies of varying quality across the US. I find that rural markets are less likely to have access to agencies with more than 3.5 stars. Urban and rural disparities in access to health care and health outcomes is a well-documented issue (Loomer et al. 2020; NC Rural Health Research Program 2017; Singh and Siahpush 2014), and this study finds similar patterns in access. Policymakers should continue to focus on implementing policy interventions that not only make star ratings information accessible to all patients, but also consider ways to achieve more equitable access to high quality care.

APPENDICES

Appendix A

Appendix to Value-Based Payments in Health Care: Evidence from a Nationwide Randomized Experiment in the Home Health Sector

Table A.1: Marginal incentive to improve on agency-reported measures targeted by year 1 of the Home Health Value-Based Purchasing program.

	Agencies with Measure (1)	2015 Performance Rate, out of 100 (2)	Change in Performance Rate with decile improvement, out of 100 (3)	Change in Expected Total Performance Points, out of 100 (4)
Panel A. Agency-reported outcome measures				
Improvement in Ambulation-Locomotion				
N (Percent of sample with 0 TPS change)	8,383 (4%)			
Mean (SE)		63.7 (0.1)	5.19 (0.05)	1.03 (0.01)
Improvement in Bed Transferring				
N	8,254 (4%)			
Mean (SE)		58.9 (0.2)	6.0 (0.1)	1.00 (0.01)
Improvement in Bathing				
N	8,412 (4%)			
Mean (SE)		67.5 (0.2)	5.36 (0.05)	1.09 (0.01)
Improvement in Dyspnea				
N	8,128 (4%)			
Mean (SE)		64.2 (0.2)	6.3 (0.1)	1.11 (0.01)
Improvement in Pain Interfering with Activity				
N	8,321 (5%)			
Mean (SE)		67.5 (0.2)	5.62 (0.05)	1.06 (0.01)
Improvement in Management of Oral Medications				
N	8,121 (4%)			
Mean (SE)		52.4 (0.2)	6.4 (0.1)	1.02 (0.01)
Panel B. Agency-reported process measures				
Pneumococcal Polysaccharide Vaccine Ever Received				
N (% with no changes in TPS even with decile improvement)	8,578 (4%)			
Mean (SE)		68.5 (0.2)	7.4 (0.1)	1.33 (0.01)
Influenza Immunization Received for Current Flu Season				
N	8,476 (4%)			
Mean (SE)		65.5 (0.2)	7.0 (0.1)	1.23 (0.01)
Drug Education on All Medications Provided to Patient/Caregiver				
N	8,584 (13%)			
Mean (SE)		94.8 (0.1)	2.3 (0.1)	1.79 (0.01)

Notes: There are 50 states represented for each measure.

Table A.2: Expected marginal gains from improvement for agency-reported measures targeted by year 1 of the Home Health Value-Based Purchasing program.

	Agencies with Measure (1)	2015 Performance Rate, out of 100 (2)	Change in Performance Rate with decile improvement, out of 100 (3)	Change in Expected Total Performance Points, out of 100 (4)
Panel A. Patient-survey measures				
Professional care				
N (% with no change in TPS with decile improvement)	5,131 (15%)			
Mean (SE)		88.7 (0.1)	1.53 (0.02)	0.81 (0.01)
Communication with patients				
N (% with no change in TPS with decile improvement)	5,131 (10%)			
Mean (SE)		85.9 (0.1)	1.84 (0.02)	0.86 (0.01)
Specific care issues				
N (% with no change in TPS with decile improvement)	5,131 (6%)			
Mean (SE)		83.3 (0.1)	2.22 (0.03)	0.84 (0.01)
Overall care rating				
N (% with no change in TPS with decile improvement)	5,131 (5%)			
Mean (SE)		84.5 (0.1)	2.41 (0.03)	0.83 (0.01)
Would recommend agency				
N (% with no change in TPS with decile improvement)	5,131 (4%)			
Mean (SE)		79.6 (0.1)	3.03 (0.04)	0.85 (0.01)
Panel B. Administrative-claims measures				
No unplanned Hospitalization				
N (% with no change in TPS with decile improvement)	7,791 (4%)			
Mean (SE)		84.25 (0.05)	1.86 (0.02)	1.01 (0.01)
No Emergency Department Use				
N (% with no change in TPS with decile improvement)	7,791 (4%)			
Mean (SE)		87.57 (0.05)	1.70 (0.02)	1.04 (0.01)

Notes: There are 50 states represented for each measure.

Table A.3: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on agency-reported measures using a difference-in-differences design.

Agency-reported measures	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value	Adj. *p*-value
Agency-reported outcome measures					
Index	8,454		0.078 (0.022)	0.000	
Ambulation-Locomotion	8,383	65.48 (13.37)	1.10 (0.38)	0.006	0.096
Bed Transferring	8,254	60.88 (15.03)	1.49 (0.39)	0.000	0.062
Bathing	8,412	68.54 (14.86)	1.36 (0.42)	0.002	0.080
Dyspnea	8,128	65.13 (18.64)	1.11 (0.39)	0.007	0.096
Pain Interfering with Activity	8,321	68.57 (17.74)	1.17 (0.55)	0.039	0.098
Management of Oral Medications	8,121	54.47 (15.05)	1.03 (0.30)	0.001	0.080
Agency-reported process measures					
Index	8,587		0.075 (0.020)	0.000	
Pneumococcal Polysaccharide Vaccine	8,578	71.93 (22.35)	1.07 (0.86)	0.221	0.452
Influenza Immunization	8,476	68.50 (19.34)	1.24 (0.70)	0.086	0.377
Drug Education	8,584	95.49 (7.86)	0.90 (0.29)	0.004	0.181

Notes: p-values reflect standard errors that are clustered at state level. Adjusted *p*-values are family-wise adjusted.

Table A.4: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on non-agency-reported measures using a difference-in-differences design.

Non-agency-reported measures	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value	Adj. *p*-value
Patient-survey measures					
Index	5,131		-0.011 (0.021)	0.616	
Professional care	5,131	88.60 (4.11)	-0.03 (0.12)	0.777	0.983
Communication with patients	5,131	85.87 (4.85)	-0.01 (0.11)	0.960	0.983
Specific care issues	5,131	83.51 (5.73)	-0.38 (0.24)	0.125	0.531
Overall care rating	5,131	84.54 (6.52)	0.10 (0.11)	0.354	0.761
Would recommend agency	5,131	79.47 (8.26)	0.05 (0.21)	0.814	0.983
Administrative-claims measures					
Index	7,791		0.015 (0.019)	0.417	
No unplanned Hospitalization	7,791	84.08 (4.09)	0.24 (0.15)	0.106	0.223
No Emergency Department Use	7,791	87.42 (4.14)	-0.12 (0.08)	0.137	0.223

Notes: p-values reflect standard errors that are clustered at state level. Adjusted *p*-values are family-wise adjusted.

Table A.5: Effect of marginal incentive size on agency-reported measures in year 1 of the Home Health Value-Based Purchasing Program using a difference-in-difference-in-differences design.

Agency-reported measures	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value	Adj. *p*-value	HHVBP × Post × Incentive (SE)	*p*-value	Adj. *p*-value
Agency-reported outcome measures								
Index	8,454		0.013	0.731		0.062	0.002	
			(0.037)			(0.020)		
Ambulation-Locomotion	8,383	65.48	-0.13	0.821	0.973	1.19	0.011	0.326
		(13.37)	(0.57)			(0.45)		
Bed Transferring	8254	60.88	0.44	0.286	0.763	1.06	0.006	0.304
		(15.03)	(0.41)			(0.37)		
Bathing	8,412	68.54	0.46	0.492	0.844	0.83	0.016	0.326
		(14.86)	(0.67)			(0.33)		
Dyspnea	8,128	65.13	-1.15	0.439	0.844	1.95	0.080	0.413
		(18.64)	(1.47)			(1.09)		
Pain Interfering with Activity	8,321	68.57	0.04	0.970	0.974	1.07	0.154	0.426
		(17.74)	(1.00)			(0.74)		
Management of Oral Medications	8,121	54.47	1.28	0.020	0.323	-0.13	0.785	0.801
		(15.05)	(0.53)			(0.48)		
Agency-reported process measures								
Index	8,587		-0.002	0.962		0.049	0.017	
			(0.032)			(0.021)		
Pneumococcal Polysaccharide Vaccine	8,578	71.93	-0.74	0.596	0.639	1.55	0.051	0.199
		(22.35)	(1.40)			(0.78)		
Influenza Immunization	8,476	68.50	2.04	0.162	0.561	-0.62	0.465	0.520
		(19.34)	(1.44)			(0.84)		
Drug Education	8,584	95.49	-0.60	0.192	0.561	0.86	0.000	0.008
		(7.86)	(0.46)			(0.16)		

Notes: p-values reflect standard errors that are clustered at state level. Adjusted *p*-values are family-wise adjusted.

Table A.6: Effect of marginal incentive size on non-agency-reported measures in year 1 of the Home Health Value-Based Purchasing Program using a difference-in-difference-in-differences design.

Non-agency-reported measures	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value	Adj. *p*-value	HHVBP × Post × Incentive (SE)	*p*-value	Adj. *p*-value
Patient-survey measures								
Index	5,131		-0.004	0.895		-0.009	0.844	
			(0.029)			(0.045)		
Professional care	5,131	88.60	0.08	0.593	0.933	-0.16	0.394	0.884
		(4.11)	(0.16)			(0.18)		
Communication with patients	5,131	85.87	-0.01	0.974	0.976	0.00	0.997	0.997
		(4.85)	(0.25)			(0.34)		
Specific care issues	5,131	83.51	-0.49	0.093	0.541	0.14	0.547	0.939
		(5.73)	(0.29)			(0.23)		
Overall care rating	5,131	84.54	0.08	0.795	0.960	0.03	0.940	0.997
		(6.52)	(0.32)			(0.39)		
Would recommend agency	5,131	79.47	0.29	0.344	0.801	-0.28	0.578	0.939
		(8.26)	(0.31)			(0.51)		
Administrative-claims measures								
Index	7,791		0.073	0.045		-0.059	0.039	
			(0.036)			(0.028)		
No unplanned Hospitalization	7,791	84.08	0.58	0.011	0.071	-0.35	0.063	0.180
		(4.09)	(0.22)			(0.18)		
No Emergency Department Use	7,791	87.42	0.01	0.931	0.935	-0.13	0.433	0.478
		(4.14)	(0.16)			(0.17)		

Notes: *p*-values reflect standard errors that are clustered at state level. Adjusted *p*-values are family-wise adjusted.

Table A.7: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on coding manipulation measures using a difference-in-differences design.

Coding manipulation measures	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value
Functional deficit level at start of care				
Index	8,241		0.068 (0.022)	0.002
Ambulation-Locomotion	8,209	2.53 (0.38)	0.026 (0.015)	0.076
Bed Transferring	8,137	1.47 (0.31)	0.036 (0.012)	0.003
Bathing	8,213	2.97 (0.49)	0.037 (0.018)	0.043
Dyspnea	8,050	1.91 (0.34)	0.008 (0.013)	0.529
Pain Interfering with Activity	8,154	2.84 (0.26)	0.010 (0.010)	0.305
Management of Oral Medications	8,046	1.91 (0.37)	0.032 (0.014)	0.028
Change in functional deficit level, discharge vs. readmission				
Index	636		0.162 (0.039)	0.000
Ambulation-Locomotion	636	0.70 (0.42)	0.148 (0.050)	0.003
Bed Transferring	636	0.51 (0.36)	0.037 (0.043)	0.390
Bathing	636	0.85 (0.53)	0.043 (0.064)	0.504
Dyspnea	636	0.76 (0.43)	0.030 (0.052)	0.568
Pain Interfering with Activity	636	1.06 (0.49)	0.132 (0.060)	0.028
Management of Oral Medications	620	0.58 (0.46)	0.046 (0.054)	0.400

Notes: p-values reflect standard errors that are clustered at state level. Functional deficit measures examine only agencies with at least 10 cases per calendar year in the denominator. For functional deficit level at start of care, this means only agencies with at least 10 complete episodes of care that meet CMS' self-reported outcome measure denominator inclusion and exclusion criteria are included. For change in functional deficit level between discharge and readmission, only agencies with at least 10 admissions following 1 day of discharge from a prior home health episode are included.

Table A.8: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on health care utilization quality measures not a part of the program using a difference-in-differences design.

Non-HHVBP quality measures	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value
Percent of discharges to inpatient institutions	8,382	6.55 (7.75)	-0.05 (0.18)	0.802
Percent of episodes followed by additional home health care	8,585	40.66 (25.56)	0.32 (0.80)	0.691

Notes: p-values reflect standard errors that are clustered at state level. Covariates include total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.9: Sensitivity analyses on the aggregate treatment effects of the Home Health Value-Based Purchasing Program in year 1 using a difference-in-differences design.

	Number of Agencies	HHVBP × Post (SE)	*p*-value
Agency-reported outcomes measures index			
Main Specification	8,454	0.078 (0.022)	0.000
Controls for covariates	8,454	0.088 (0.023)	0.000
Controls for expected marginal gains from improvement on other measures	8,454	0.078 (0.022)	0.000
Excludes chains that serve both treatment and control states	7,152	0.067 (0.023)	0.003
Agency-reported process measures index			
Main Specification	8,587	0.075 (0.020)	0.000
Controls for covariates	8,587	0.086 (0.023)	0.000
Controls for expected marginal gains from improvement on other measures	8,587	0.075 (0.020)	0.000
Excludes chains that serve both treatment and control states	7,283	0.070 (0.034)	0.040
Patient-survey measures index			
Main Specification	5,131	-0.011 (0.021)	0.616
Controls for covariates	5,131	-0.005 (0.026)	0.842
Controls for expected marginal gains from improvement on other measures	5,131	-0.011 (0.021)	0.616
Excludes chains that serve both treatment and control states	3,936	0.011 (0.020)	0.603
Administrative-claims measures index			
Main Specification	7,791	0.015 (0.019)	0.417
Controls for covariates	7,791	0.019 (0.017)	0.278
Controls for expected marginal gains from improvement on other measures	7,791	0.015 (0.019)	0.417
Excludes chains that serve both treatment and control states	6,500	0.000 (0.021)	0.999

Notes: p-values reflect standard errors that are clustered at state level. Covariates include total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.10: Sensitivity analyses of effects of marginal incentive size on performance in the Home Health Value-Based Purchasing Program in year 1 using a difference-in-difference-in-differences design.

	Number of agencies	HHVBP × Post (SE)	*p*-value	HHVBP × Post × Incentive (SE)	*p*-value
Agency-reported outcomes measures index					
Main Specification	8,454	0.013 (0.037)	0.731	0.062 (0.020)	0.002
Controls for covariates	8,454	0.015 (0.037)	0.682	0.069 (0.020)	0.000
Controls for expected marginal gains from improvement on other measures	8,454	0.013 (0.037)	0.731	0.062 (0.020)	0.002
Excludes chains that serve both treatment and control states	7,152	-0.025 (0.040)	0.526	0.085 (0.024)	0.000
Agency-reported process measures index					
Main Specification	8,587	-0.002 (0.032)	0.962	0.049 (0.021)	0.017
Controls for covariates	8,587	0.014 (0.034)	0.689	0.043 (0.021)	0.038
Controls for expected marginal gains from improvement on other measures	8,587	-0.002 (0.032)	0.962	0.049 (0.021)	0.017
Excludes chains that serve both treatment and control states	7,283	0.011 (0.048)	0.823	0.037 (0.021)	0.080
Patient-survey measures index					
Main Specification	5,131	-0.004 (0.029)	0.895	-0.009 (0.045)	0.844
Controls for covariates	5,131	0.003 (0.026)	0.901	-0.011 (0.043)	0.802
Controls for expected marginal gains from improvement on other measures	5,131	-0.004 (0.029)	0.895	-0.009 (0.045)	0.844
Excludes chains that serve both treatment and control states	3,936	-0.004 (0.030)	0.887	0.017 (0.039)	0.654
Administrative-claims measures index					
Main Specification	7,791	0.073 (0.036)	0.045	-0.059 (0.028)	0.039
Controls for covariates	7,791	0.079 (0.037)	0.032	-0.061 (0.030)	0.040
Controls for expected marginal gains from improvement on other measures	7,791	0.073 (0.036)	0.045	-0.059 (0.028)	0.039
Excludes chains that serve both treatment and control states	6,500	0.042 (0.039)	0.281	-0.040 (0.030)	0.182

Notes: p-values reflect standard errors that are clustered at state level. Covariate adjusted regresssions include adjustment for total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.11: Sensitivity analysis of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on coding manipulation using a difference-in-differences design.

Coding manipulation measures	Control for covariates	Number of Agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value
Functional deficit level at start of care					
Main specification (10+ episodes)					
Index	No	8,241		0.068 (0.022)	0.002
	Yes	8,241		0.060 (0.016)	0.000
All observations, weighted					
Index	No	8,576		0.093 (0.026)	0.000
	Yes	8,576		0.088 (0.023)	0.000
Change in functional deficit level, discharge vs. readmission					
Main specification (10+ episodes)					
Index	No	636		0.162 (0.039)	0.000
	Yes	636		0.158 (0.041)	0.000
All observations, weighted					
Index	No	5,750		0.060 (0.027)	0.024
	Yes	5,750		0.060 (0.026)	0.022

Notes: p-values reflect standard errors that are clustered at state level. In the main specifications, only agencies with at least 10 admissions following 1 day of discharge from a prior home health episode are included. For change in functional deficit level between discharge and readmission, only agencies with at least 10 admissions following 1 day of discharge from a prior home health episode are included. Covariates include total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.12: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on percentage of discharges to inpatient institutions (non-targeted measure) using a difference-in-differences design, stratified by patient type.

Percent of discharges to inpatient institutions	Control for covariates	Number of agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value
All admissions					
	No	8,382	6.55 (7.75)	-0.05 (0.18)	0.802
	Yes	8,382	6.55 (7.75)	-0.16 (0.15)	0.311
Post-acute care admissions					
	No	8,335	9.06 (12.13)	-0.18 (0.29)	0.533
	Yes	8,335	9.06 (12.13)	-0.35 (0.24)	0.156
Community admissions					
	No	8,308	4.68 (7.03)	0.26 (0.18)	0.142
	Yes	8,308	4.68 (7.03)	0.21 (0.17)	0.207

Notes: p-values reflect standard errors that are clustered at state level. Post-acute admissions refer to home health episodes where the patients were discharged from an inpatient institution in the last 14 days. Covariates include total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.13: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on percentage of episodes followed by additional home health care (non-targeted measure) using a difference-in-differences design, stratified by patient type.

Percent of episodes followed by additional home health care	Control for covariates	Number of Agencies	Mean (SD) of control group	HHVBP × Post (SE)	*p*-value
All admissions					
	No	8,585	40.66 (25.56)	0.32 (0.80)	0.691
	Yes	8,585	40.66 (25.56)	-0.13 (0.65)	0.846
Post-acute care admissions					
	No	8,537	41.03 (26.24)	0.10 (1.05)	0.928
	Yes	8,537	41.03 (26.24)	-0.32 (0.76)	0.676
Community admissions					
	No	8,563	41.87 (26.60)	0.33 (0.68)	0.632
	Yes	8,563	41.87 (26.60)	-0.17 (0.63)	0.792

Notes: p-values reflect standard errors that are clustered at state level. Post-acute admissions refer to home health episodes where the patients were discharged from an inpatient institution in the last 14 days. Covariates include total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.14: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on agency-reported outcome and agency-reported process measures using a cross-sectional design across total sample.

			Estimates using Unrestricted Sample					
			No covariates			Covariate adjusted		
	Number of agencies	Mean (SD) of control group	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value
Agency-reported outcome measures								
Index	8,454		0.38 (0.17)	0.025		0.141 (0.029)	0.000	
Ambulation-Locomotion	8,383	67.73 (13.36)	4.64 (2.25)	0.044	0.660	1.94 (0.47)	0.000	0.157
Bed Transferring	8,254	63.46 (14.92)	6.22 (2.24)	0.008	0.625	2.59 (0.41)	0.000	0.044
Bathing	8,412	70.12 (15.04)	5.60 (2.57)	0.034	0.655	2.17 (0.44)	0.000	0.104
Dyspnea	8,128	66.76 (18.60)	8.46 (3.46)	0.018	0.638	2.58 (0.68)	0.000	0.157
Pain Interfering with Activity	8,321	70.40 (17.83)	8.19 (4.14)	0.054	0.663	2.58 (0.62)	0.000	0.157
Management of Oral Medications	8,121	56.62 (15.27)	3.06 (1.77)	0.090	0.663	1.49 (0.32)	0.000	0.119
Agency-reported process measures								
Index	8,587		0.035 (0.063)	0.580		0.050 (0.011)	0.000	
Pneumococcal Polysaccharide Vaccine	8,578	75.14 (21.40)	-0.59 (2.67)	0.825	0.894	0.40 (0.35)	0.268	0.362
Influenza Immunization	8,476	71.54 (18.96)	1.54 (1.50)	0.308	0.668	0.92 (0.37)	0.016	0.092
Drug Education	8,584	96.06 (7.67)	0.40 (0.26)	0.131	0.640	0.65 (0.16)	0.000	0.028

Notes: p-values reflect standard errors that are clustered at state level. Adj. p-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.15: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on patient-survey and administrative-claims measures using a cross-sectional design across total sample.

			Estimates using Unrestricted Sample					
			No covariates			Covariate adjusted		
	Number of agencies	Mean (SD) of control group	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value
Patient-survey measures								
Index	5,131		-0.10	0.425		-0.02	0.572	
			(0.13)			(0.03)		
Professional care	5,131	88.55	-0.27	0.604	0.780	-0.04	0.781	0.982
		(4.12)	(0.52)			(0.16)		
Communication with patients	5,131	85.86	-0.33	0.556	0.780	0.00	0.972	0.982
		(4.87)	(0.56)			(0.13)		
Specific care issues	5,131	83.54	-1.42	0.167	0.583	-0.55	0.086	0.528
		(5.72)	(1.01)			(0.31)		
Overall care rating	5,131	84.58	-0.49	0.520	0.777	0.07	0.736	0.982
		(6.61)	(0.76)			(0.20)		
Would recommend agency	5,131	79.35	-0.53	0.621	0.780	0.04	0.842	0.982
		(8.31)	(1.07)			(0.22)		
Administrative-claims measures								
Index	7,791		0.10	0.286		0.043	0.194	
			(0.10)			(0.033)		
No unplanned Hospitalization	7,791	83.87	0.19	0.391	0.818	0.18	0.310	0.756
		(4.09)	(0.22)			(0.18)		
No Emergency Department Use	7,791	87.37	0.66	0.382	0.818	0.17	0.617	0.771
		(4.11)	(0.75)			(0.34)		

Notes: p-values reflect standard errors that are clustered at state level. Adj. p-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.16: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on agency-reported outcome and process measures using a cross-sectional design across propensity-score matched samples.

			Estimates using Propensity-Score Matched Samples					
			No covariates			Covariate adjusted		
	Number of agencies	Mean (SD) of control group	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value
Agency-reported outcome measures								
Index	5,509		0.26 (0.13)	0.048		0.159 (0.024)	0.000	
Ambulation-Locomotion	2,860	69.32 (11.20)	2.61 (1.44)	0.077	0.685	1.48 (0.25)	0.000	0.022
Bed Transferring	2,802	66.28 (12.70)	3.30 (1.41)	0.023	0.675	2.19 (0.36)	0.000	0.021
Bathing	2,862	72.30 (12.31)	3.27 (1.61)	0.048	0.685	2.23 (0.28)	0.000	0.006
Dyspnea	2,716	69.87 (15.69)	5.10 (2.36)	0.036	0.685	2.45 (0.64)	0.000	0.095
Pain Interfering with Activity	2,816	73.45 (14.41)	3.84 (2.77)	0.172	0.685	2.82 (0.53)	0.000	0.032
Management of Oral Medications	2,744	57.08 (13.92)	3.22 (1.55)	0.044	0.685	1.60 (0.27)	0.000	0.022
Agency-reported process measures								
Index	4,556		0.091 (0.044)	0.039		0.065 (0.015)	0.000	
Pneumococcal Polysaccharide Vaccine	2,968	76.89 (19.95)	0.63 (1.88)	0.737	0.875	0.68 (0.40)	0.096	0.292
Influenza Immunization	2,904	72.99 (17.94)	2.23 (1.10)	0.047	0.458	1.17 (0.59)	0.053	0.292
Drug Education	2,970	95.62 (8.56)	0.99 (0.34)	0.005	0.280	0.83 (0.20)	0.000	0.057

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.17: Estimates of the aggregate effect of year 1 of the Home Health Value-Based Purchasing Program on patient-survey and administrative-claims measures using a cross-sectional design across propensity-score matched samples.

			Estimates using Propensity-Score Matched Samples					
			No covariates			Covariate adjusted		
	Number of agencies	Mean (SD) of control group	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value	HHVBP vs. no HHVBP (SE)	*p*-value	Adj. *p*-value
Patient-survey measures								
Index	3,581		-0.077 (0.090)	0.393		-0.030 (0.026)	0.249	
Professional care	1,928	88.67 (3.98)	-0.33 (0.35)	0.350	0.768	-0.19 (0.14)	0.187	0.735
Communication with patients	1,928	85.83 (4.88)	-0.17 (0.41)	0.685	0.819	0.05 (0.14)	0.751	0.822
Specific care issues	1,928	83.09 (5.91)	-0.55 (0.65)	0.400	0.780	-0.32 (0.22)	0.153	0.735
Overall care rating	1,928	84.79 (6.31)	-0.57 (0.55)	0.307	0.763	-0.14 (0.17)	0.426	0.822
Would recommend agency	1,928	79.92 (7.93)	-0.66 (0.79)	0.413	0.780	-0.26 (0.27)	0.355	0.822
Administrative-claims measures								
Index	3,654		0.054 (0.081)	0.504		0.039 (0.038)	0.304	
No unplanned Hospitalization	2,754	83.62 (4.05)	0.39 (0.15)	0.010	0.253	0.41 (0.14)	0.006	0.166
No Emergency Department Use	2,754	87.52 (3.88)	0.04 (0.58)	0.942	0.967	-0.09 (0.28)	0.749	0.840

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.18: Effect of marginal incentive size on agency-reported outcome measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples.

Agency-reported outcome measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	5,509		-0.190 (0.054)	0.000	
	Yes	5,509		0.005 (0.020)	0.809	
Ambulation-Locomotion	No	2,860	69.32 (11.20)	-2.31 (0.74)	0.003	0.248
	Yes	2,860	69.32 (11.20)	0.12 (0.41)	0.781	0.977
Bed Transferring	No	2,802	66.28 (12.70)	-3.08 (1.15)	0.010	0.322
	Yes	2,802	66.28 (12.70)	-0.77 (0.56)	0.178	0.755
Bathing	No	2,862	72.30 (12.31)	-2.35 (1.12)	0.041	0.430
	Yes	2,862	72.30 (12.31)	0.03 (0.48)	0.950	0.977
Dyspnea	No	2,716	69.87 (15.69)	-5.17 (2.55)	0.048	0.430
	Yes	2,716	69.87 (15.69)	0.43 (0.72)	0.557	0.964
Pain Interfering with Activity	No	2,816	73.45 (14.41)	-0.66 (0.84)	0.435	0.529
	Yes	2,816	73.45 (14.41)	0.95 (0.48)	0.054	0.524
Management of Oral Medications	No	2,744	57.08 (13.92)	-1.77 (0.78)	0.029	0.407
	Yes	2,744	57.08 (13.92)	-0.23 (0.60)	0.701	0.977

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.19: Effect of marginal incentive size on agency-reported process measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples.

Agency-reported process measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	4,556		0.122 (0.045)	0.007	
	Yes	4,556		0.100 (0.017)	0.000	
Pneumococcal Polysaccharide Vaccine	No	2,968	76.89 (19.95)	4.67 (1.82)	0.014	0.226
	Yes	2,968	76.89 (19.95)	2.66 (0.73)	0.001	0.033
Influenza Immunization	No	2,904	72.99 (17.94)	1.86 (1.79)	0.303	0.608
	Yes	2,904	72.99 (17.94)	0.80 (0.73)	0.284	0.371
Drug Education	No	2,970	95.62 (8.56)	0.25 (0.56)	0.655	0.701
	Yes	2,970	95.62 (8.56)	1.06 (0.36)	0.005	0.062

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.20: Effect of marginal incentive size on patient-survey measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples.

Patient-survey measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	3,581		-0.016 (0.069)	0.820	
	Yes	3,581		-0.017 (0.019)	0.367	
Professional care	No	1,928	88.67 (3.98)	0.30 (0.37)	0.422	0.779
	Yes	1,928	88.67 (3.98)	-0.10 (0.15)	0.535	0.910
Communication with patients	No	1,928	85.83 (4.88)	-0.05 (0.52)	0.926	0.937
	Yes	1,928	85.83 (4.88)	0.04 (0.21)	0.851	0.980
Specific care issues	No	1,928	83.09 (5.91)	-0.53 (0.42)	0.218	0.732
	Yes	1,928	83.09 (5.91)	-0.21 (0.21)	0.312	0.864
Overall care rating	No	1,928	84.79 (6.31)	-0.58 (0.58)	0.322	0.766
	Yes	1,928	84.79 (6.31)	-0.23 (0.30)	0.435	0.908
Would recommend agency	No	1,928	79.92 (7.93)	0.30 (0.70)	0.673	0.881
	Yes	1,928	79.92 (7.93)	0.02 (0.23)	0.947	0.980

Notes: *p*-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.21: Effect of marginal incentive size on administrative-claims measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across propensity-score matched samples.

Administrative-claims measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	3,654		0.017 (0.061)	0.781	
	Yes	3,654		-0.008 (0.030)	0.784	
No unplanned Hospitalization	No	2,754	83.62 (4.05)	0.29 (0.20)	0.160	0.457
	Yes	2,754	83.62 (4.05)	0.14 (0.16)	0.383	0.644
No Emergency Department Use	No	2,754	87.52 (3.88)	-0.15 (0.37)	0.696	0.760
	Yes	2,754	87.52 (3.88)	-0.20 (0.22)	0.371	0.644

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.22: Effect of marginal incentive size on agency-reported outcome measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample.

Agency-reported outcome measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	8,454		-0.087 (0.046)	0.058	
	Yes	8,454		0.025 (0.019)	0.190	
Ambulation-Locomotion	No	8,383	67.73 (13.36)	-0.52 (0.56)	0.362	0.714
	Yes	8,383	67.73 (13.36)	0.75 (0.31)	0.019	0.400
Bed Transferring	No	8,254	63.46 (14.92)	-1.27 (0.83)	0.131	0.631
	Yes	8,254	63.46 (14.92)	0.45 (0.35)	0.197	0.782
Bathing	No	8,412	70.12 (15.04)	-1.21 (1.01)	0.236	0.714
	Yes	8,412	70.12 (15.04)	0.35 (0.40)	0.385	0.782
Dyspnea	No	8,128	66.76 (18.60)	-2.59 (2.06)	0.215	0.714
	Yes	8,128	66.76 (18.60)	0.60 (0.51)	0.246	0.782
Pain Interfering with Activity	No	8,321	70.40 (17.83)	-0.45 (1.39)	0.747	0.793
	Yes	8,321	70.40 (17.83)	0.89 (0.74)	0.236	0.782
Management of Oral Medications	No	8,121	56.62 (15.27)	-2.37 (1.32)	0.079	0.601
	Yes	8,121	56.62 (15.27)	-0.67 (0.56)	0.237	0.782

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.23: Effect of marginal incentive size on agency-reported process measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample.

Agency-reported process measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	8,587		0.063 (0.048)	0.185	
	Yes	8,587		0.047 (0.019)	0.015	
Pneumococcal Polysaccharide Vaccine	No	8,578	75.14 (21.40)	3.45 (2.14)	0.114	0.484
	Yes	8,578	75.14 (21.40)	2.00 (0.78)	0.013	0.163
Influenza Immunization	No	8,476	71.54 (18.96)	1.08 (1.58)	0.498	0.810
	Yes	8,476	71.54 (18.96)	-0.11 (0.73)	0.878	0.905
Drug Education	No	8,584	96.06 (7.67)	-0.22 (0.45)	0.629	0.810
	Yes	8,584	96.06 (7.67)	0.42 (0.21)	0.052	0.213

Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.24: Effect of marginal incentive size on patient-survey measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample.

Patient-survey measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	5,131		-0.029 (0.064)	0.649	
	Yes	5,131		-0.011 (0.019)	0.563	
Professional care	No	5,131	88.55 (4.12)	0.27 (0.31)	0.393	0.727
	Yes	5,131	88.55 (4.12)	0.00 (0.10)	0.972	0.974
Communication with patients	No	5,131	85.86 (4.87)	0.03 (0.49)	0.943	0.992
	Yes	5,131	85.86 (4.87)	0.10 (0.16)	0.547	0.940
Specific care issues	No	5,131	83.54 (5.72)	-0.61 (0.38)	0.115	0.574
	Yes	5,131	83.54 (5.72)	-0.13 (0.19)	0.492	0.940
Overall care rating	No	5,131	84.58 (6.61)	-0.67 (0.42)	0.118	0.574
	Yes	5,131	84.58 (6.61)	-0.22 (0.22)	0.320	0.856
Would recommend agency	No	5,131	79.35 (8.31)	-0.07 (0.74)	0.920	0.992
	Yes	5,131	79.35 (8.31)	-0.15 (0.22)	0.491	0.940

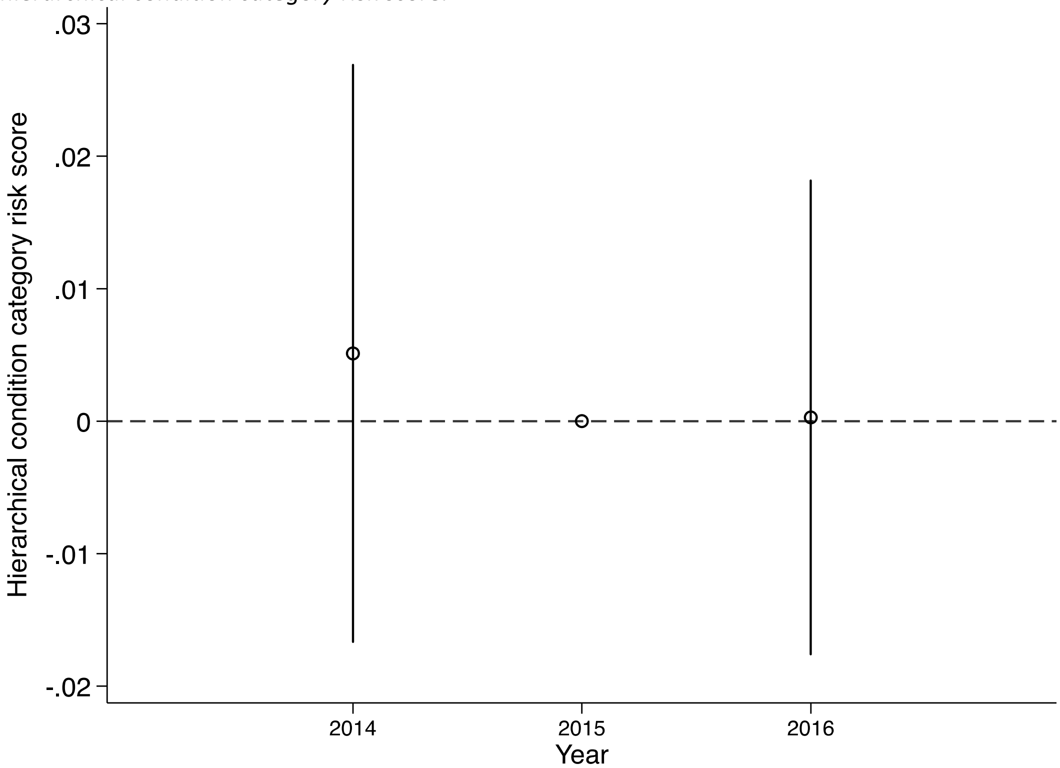
Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Table A.25: Effect of marginal incentive size on administrative-claims measures in year 1 of the Home Health Value-Based Purchasing Program using a cross-sectional design across total sample.

Administrative-claims measures	Control for covariates	Number of agencies	Mean (SD) of control group	Incentive × HHVBP (SE)	*p*-value	Adj. *p*-value
Index	No	7,791		-0.005 (0.030)	0.863	
	Yes	7,791		-0.021 (0.016)	0.180	
No unplanned Hospitalization	No	7,791	83.87 (4.09)	-0.05 (0.12)	0.713	0.943
	Yes	7,791	83.87 (4.09)	-0.13 (0.13)	0.297	0.601
No Emergency Department Use	No	7,791	87.37 (4.11)	0.00 (0.20)	0.984	0.990
	Yes	7,791	87.37 (4.11)	-0.04 (0.11)	0.690	0.740

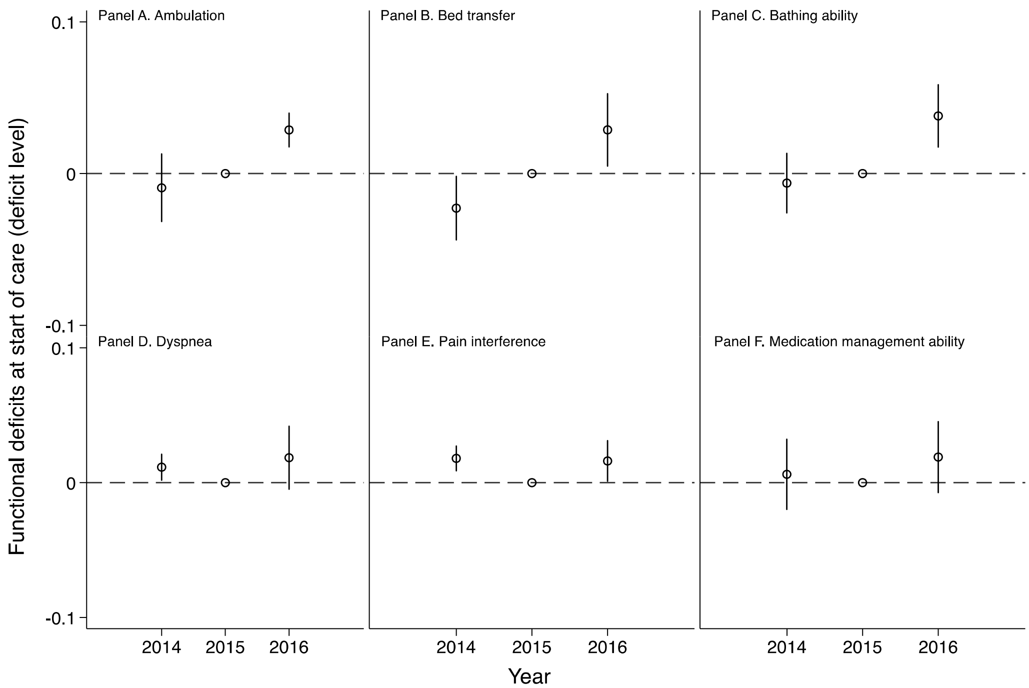
Notes: p-values reflect standard errors that are clustered at state level. Adj. *p*-value is family-wise adjusted and also clustered at state level. All regressions include adjustment for region. Covariate adjusted regresssions also include adjustment for lagged performance from 2015, rural status, ownership status, freestanding status, total number of Medicare admissions, percent of admissions that were discharged from acute care, percent of admissions from the bottom income quartile of each state, and percent of admission that are associated with 10 to 20% lower profit margins for home health agencies.

Figure A.1: Event studies of the effect of the Home Health Value-Based Purchasing model on the average hierarchical condition category risk score.



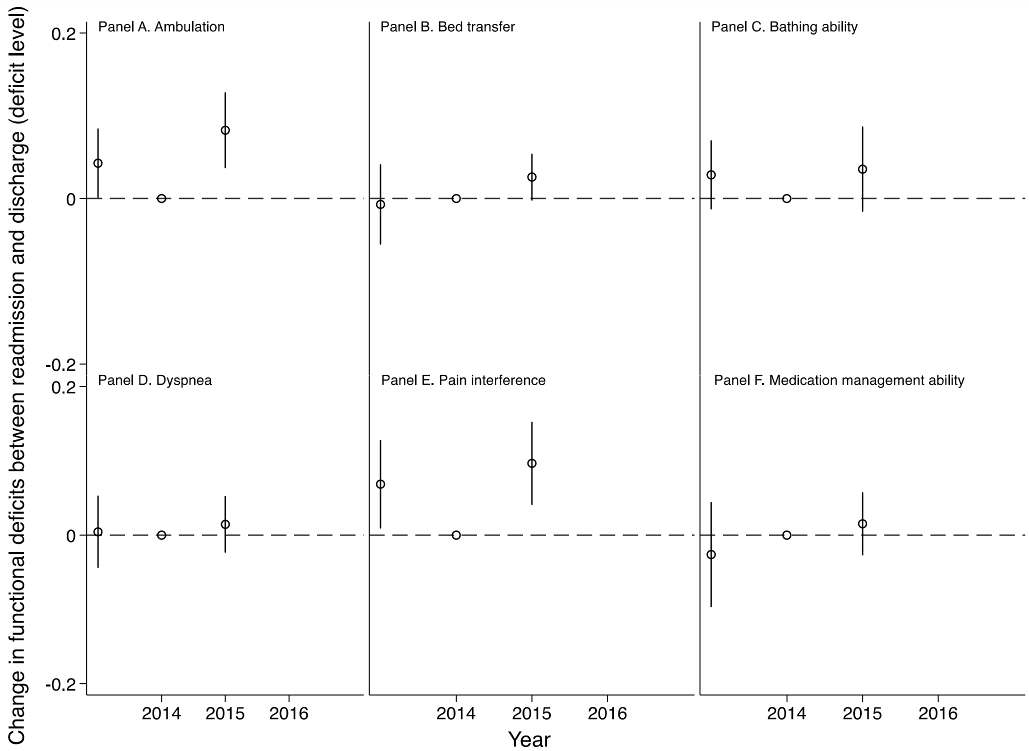
Notes: Treatment effects and 95 percent confidence intervals are estimated from a generalized difference-in-differences model plus dummy variables for 2014 and 2016.

Figure A.2: Event studies of the functional deficit level at start of care across agencies using a weighted regression.



Notes: Treatment effects and 95 percent confidence intervals are estimated from a generalized difference-in-differences model plus dummy variables for 2014 and 2016 using a weighted regression that gives more weight to agencies with more patients.

Figure A.3: Event studies of the change in functional deficit level between initial discharge and start of care at readmission across agencies using a weighted regression.



Notes: Treatment effects and 95 percent confidence intervals are estimated from a generalized difference-in-differences model plus dummy variables for 2014 and 2016 using a weighted regression that gives more weight to agencies with more patients.

Appendix B

Appendix to Medicare's Home Health Star Ratings Program Has Not Shifted Patients into Higher-Rated Agencies

Specifications

One more half-star:

1. Linear (preferred) $$Pats_{jq} = \beta_0 + \beta_1 Star_{jq} + \beta_2 running_{jq} + X\beta_{jq} + \epsilon_{jq}$$
2. Linear interaction $$Pats_{jq} = \beta_0 + \beta_1 Star_{jq} + \beta_2 running_{jq} + \beta_3 running_{jq} \times Star_{jq} + X\beta_{jq} + \epsilon_{jq}$$
3. Quadratic $$Pats_{jq} = \beta_0 + \beta_1 Star_{jq} + \beta_2 running_{jq} + \beta_3 running_{jq}^2 + X\beta_{jq} + \epsilon_{jq}$$
4. Quadratic interaction $$Pats_{jq} = \beta_0 + \beta_1 Star_{jq} + \beta_2 running_{jq} + \beta_3 running_{jq}^2 + \beta_4 running_{jq} \times Star_{jq} + \beta_5 running_{jq}^2 \times Star_{jq} + X\beta_{jq} + \epsilon_{jq}$$

Where:

$Pats_{jq}$ is the number of new patients per agency j in quarter q; β_0 is the mean number of new patients after conditioning on the unrounded star ratings and covariates; $Star_{jq}$ is equal to 1 if the observation received a higher star rating (right of the threshold) or 0 otherwise; and $running_{jq}$ is the unrounded star rating for agency j in quarter q, centered at the rounding threshold; $X\beta_{jq}$ is a vector of agency-level covariates including number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency age, chain affiliation, for-profit status, prior quarter agency Medicare patient characteristics including average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease, and star rating release fixed effects. Standard errors were clustered at the home health agency level. In all specifications, β_1 provides the marginal effect of having one more half-star at the rounding threshold.

Highest-ranked option in ZIP code:

$$Pats_{jzq} = \alpha_0 + \alpha_1 Quarter2_q + \alpha_2 Highest_{jz} + \alpha_3 Quarter2_q \times Highest_{jz} + X\beta_{jzq} + \epsilon_{jq}$$

Where:

$Pats_{jzq}$ is the number of new patients per agency j in ZIP code z in quarter q; α_0 is the number of new patients among agencies that did not become the highest-ranked option per ZIP code in quarter 1 after controlling for covariates; $Quarter2_q$ takes the value of 1 for quarter 2 and 0 for quarter 1; $Highest_{jz}$ takes the value of 1 if an agency became the highest-ranked option within the ZIP code and 0 otherwise; $X\beta_{jq}$ is a vector of time varying agency and agency-ZIP code level covariates, including unrounded star ratings, number of home health competitors in each ZIP code, percent of full Medicare-Medicaid dually eligible patients in each ZIP-agency pair, percent of partial Medicare-Medicaid dually eligible patients in each ZIP-agency pair, percent of patients discharged from an inpatient source in the past 14 days in each ZIP-agency pair, percent of patients with end stage renal disease in each ZIP-agency pair, percent female in each ZIP-agency pair, percent white in each ZIP-agency pair, percent black in each ZIP-agency pair, percent Hispanic in each ZIP-agency pair, mean age of Medicare patients in each ZIP-agency pair, and percent of Medicare fee-for-service patients in each ZIP-agency pair. I cluster standard errors at the home health agency level. In this specification, α_3 provides the marginal effect of becoming the highest-ranked option within a ZIP code.

Table B.1: Star rating release dates and days included in analysis.

Star ratings examined in study	Days in pre-release period	Days in post-release period
Quarter 1 (July 16, 2015)	83	83
Quarter 2 (October 8, 2015)	83	111
Quarter 3 (January 28, 2016)	111	82
Quarter 4 (April 20, 2016)	82	83
Quarter 5 (July 13, 2016)	83	97
Quarter 6 (October 19, 2016)	97	73

Table B.2: Measures included in the first six quarters of the home health star ratings program.

Measure Type	Measure and brief description
Process of Care	***Timely initiation of care*** Percentage of home health quality episodes in which the start or resumption of care date was either on the physician specified date or within 2 days of the referral date or inpatient discharge date, whichever is later.
Process of Care	***Drug education on all medications provided to patient/caregiver*** Percentage of home health 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.
Process of Care	***Influenza immunization received for the current flu season*** Percentage of home health quality episodes during which patients were offered and refused influenza immunization for the current flu season.
Health Outcome	***Improvement in ambulation*** Percentage of home health quality episodes during which the patient improved in ability to ambulate.
Health Outcome	***Improvement in bed transferring*** Percentage of home health quality episodes during which the patient improved in ability to get in and out of bed.
Health Outcome	***Improvement in bathing*** Percentage of home health quality episodes during which the patient got better at bathing self.
Health Outcome	***Improvement in pain*** Percentage of home health quality episodes during which the patient's frequency of pain when moving around improved.
Health Outcome	***Improvement in shortness of breath*** Percentage of home health quality episodes during which the patient became less short of breath or dyspneic.
Health Outcome	***Acute care hospitalization*** Percentage of home health stays in which patients were admitted to an acute care hospital during the 60 days following the start of the home health stay.

Table B.3: Functional form assessment of regression discontinuity design for all thresholds, 1 vs. 1.5 stars, and 1.5 vs. 2 stars.

		Heap			Not heaped		
	Covariates	Treatment Estimate	(SE)	AIC	Treatment Estimate	(SE)	AIC
All thresholds							
Linear	No	2.59	(3.18)	553238	-1.16	(1.00)	35951
	Yes	-0.57	(0.80)	438845	0.34	(0.66)	31545
Linear Interaction	No	2.60	(3.18)	553239	-0.49	(1.06)	35951
	Yes	-0.57	(0.80)	438847	0.85	(0.76)	31544
Quadratic	No	2.60	(3.18)	553239	-0.66	(1.01)	35950
	Yes	-0.57	(0.80)	438847	0.65	(0.71)	31545
Quadratic Interaction	No	1.77	(4.90)	553243	0.45	(2.46)	35953
	Yes	-1.99	(1.31)	438849	1.37	(1.61)	31548
1 vs. 1.5 stars							
Linear	No	17.05	(11.26)	2742	0.22	(2.53)	1479
	Yes	-1.47	(5.59)	2685	-0.13	(1.79)	1409
Linear Interaction	No	10.92	(9.35)	2743	-2.08	(2.47)	1480
	Yes	0.74	(6.65)	2686	-0.88	(2.43)	1411
Quadratic	No	7.35	(7.98)	2742	-1.83	(2.54)	1479
	Yes	0.10	(6.08)	2687	-0.72	(1.99)	1411
Quadratic Interaction	No	-21.74	(13.46)	2743	12.41	(13.06)	1478
	Yes	-25.64	(8.68)	2688	9.55	(9.59)	1411
1.5 vs. 2 stars							
Linear	No	3.09	(3.63)	25758	-1.11	(1.20)	5365
	Yes	0.81	(1.20)	21364	-0.12	(0.83)	4592
Linear Interaction	No	1.58	(3.29)	25756	0.58	(1.26)	5363
	Yes	1.04	(1.14)	21366	0.92	(0.91)	4589
Quadratic	No	1.05	(3.25)	25754	0.14	(1.19)	5362
	Yes	1.12	(1.14)	21365	0.46	(0.87)	4591
Quadratic Interaction	No	0.47	(4.33)	25756	5.11	(3.72)	5364
	Yes	1.09	(1.88)	21369	2.95	(1.94)	4590

Notes: Covariates include the number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Standard errors were clustered at the home health agency level.

Table B.4: Functional form assessment of regression discontinuity design 2 vs. 2.5 stars, 2.5 vs. 3 stars, and 3 vs. 3.5 stars.

		Heap			Not heaped		
	Covariates	Treatment Estimate	(SE)	AIC	Treatment Estimate	(SE)	AIC
2 vs. 2.5 stars							
Linear	No	4.00	(2.67)	73788	0.58	(1.56)	7496
	Yes	-0.28	(0.93)	58330	-0.64	(1.03)	6735
Linear Interaction	No	3.23	(2.60)	73784	1.77	(1.73)	7495
	Yes	-0.29	(0.93)	58332	0.10	(1.40)	6735
Quadratic	No	3.13	(2.59)	73784	1.50	(1.64)	7495
	Yes	-0.28	(0.93)	58332	-0.03	(1.25)	6734
Quadratic Interaction	No	4.04	(4.22)	73788	-1.27	(4.37)	7498
	Yes	-1.10	(1.73)	58336	-1.04	(2.75)	6738
2.5 vs. 3 stars							
Linear	No	8.75	(4.86)	141210	-1.91	(1.78)	7130
	Yes	-0.99	(1.20)	111135	0.53	(1.04)	6311
Linear Interaction	No	8.34	(4.84)	141211	-1.32	(1.88)	7132
	Yes	-1.00	(1.19)	111137	1.34	(1.16)	6310
Quadratic	No	8.07	(4.81)	141209	-1.48	(1.77)	7131
	Yes	-1.01	(1.18)	111137	1.01	(1.07)	6309
Quadratic Interaction	No	-4.06	(10.10)	141206	-1.56	(4.46)	7135
	Yes	-4.02	(2.17)	111139	0.82	(2.54)	6313
3 vs. 3.5 stars							
Linear	No	6.88	(6.32)	171223	1.18	(2.48)	5515
	Yes	-0.96	(1.45)	137993	0.87	(1.93)	5018
Linear Interaction	No	6.84	(6.30)	171224	0.55	(2.45)	5517
	Yes	-0.97	(1.45)	137995	0.03	(1.89)	5018
Quadratic	No	6.83	(6.30)	171223	0.77	(2.42)	5517
	Yes	-0.97	(1.45)	137994	0.23	(1.86)	5017
Quadratic Interaction	No	3.61	(8.66)	171221	1.36	(4.43)	5521
	Yes	-3.83	(2.30)	137993	0.26	(2.65)	5020

Notes: Covariates include the number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Standard errors were clustered at the home health agency level.

Table B.5: Functional form assessment of regression discontinuity design 3.5 vs. 4 stars, 4 vs. 4.5 stars, and 4.5 vs. 5 stars.

		Heap			Not heaped		
	Covariates	Treatment Estimate	(SE)	AIC	Treatment Estimate	(SE)	AIC
3.5 vs. 4 stars							
Linear	No	3.16	(7.67)	145923	0.71	(3.93)	3811
	Yes	0.96	(2.08)	118218	3.84	(2.57)	3383
Linear Interaction	No	4.24	(7.75)	145917	1.24	(4.07)	3813
	Yes	0.98	(2.11)	118220	4.70	(2.75)	3384
Quadratic	No	4.33	(7.75)	145918	0.92	(3.92)	3813
	Yes	0.98	(2.11)	118220	4.25	(2.61)	3385
Quadratic Interaction	No	1.20	(14.41)	145921	2.25	(7.16)	3817
	Yes	0.72	(3.74)	118224	15.55	(6.62)	3383
4 vs. 4.5 stars							
Linear	No	-1.77	(8.19)	90093	-13.86	(7.19)	2578
	Yes	1.00	(1.97)	73456	-8.89	(4.49)	2212
Linear Interaction	No	-0.66	(8.40)	90092	-11.99	(8.28)	2580
	Yes	0.96	(1.94)	73458	-9.30	(5.59)	2214
Quadratic	No	-0.44	(8.42)	90092	-12.87	(7.52)	2580
	Yes	0.92	(1.92)	73458	-8.91	(4.99)	2214
Quadratic Interaction	No	2.86	(14.51)	90096	-4.92	(25.04)	2584
	Yes	0.73	(2.94)	73462	-25.54	(14.14)	2213
4.5 vs. 5 stars							
Linear	No	-5.25	(7.72)	39041	-7.35	(13.40)	1163
	Yes	-0.93	(1.81)	30816	23.45	(14.75)	1035
Linear Interaction	No	-9.69	(7.51)	39034	-5.86	(12.31)	1164
	Yes	-0.94	(1.80)	30818	22.72	(14.79)	1037
Quadratic	No	-10.39	(7.51)	39033	-8.28	(13.76)	1163
	Yes	-1.09	(1.79)	30818	23.68	(14.63)	1037
Quadratic Interaction	No	-14.92	(11.01)	39037	13.66	(16.78)	1157
	Yes	-0.38	(2.80)	30820	22.69	(14.96)	1040

Notes: Covariates include the number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Standard errors were clustered at the home health agency level.

Table B.6: Effects of removing outermost observations on regression discontinuity estimates.

		Heap			Not heaped		
	Covariates	N agencies	Prior Quarter New Patients, mean (SD)	Treatment Estimate (SE)	N agencies	Prior Quarter New Patients, mean (SD)	Treatment Estimate (SE)
Pooled threshold							
(-0.25, 0.25)		8,998	83 (157)		1,818	9 (15)	
	No			2.59 (3.18)			-1.16 (1.00)
	Yes, partial			-0.50 (0.86)			0.40 (0.67)
	Yes, full			-0.57 (0.80)			0.34 (0.66)
(-0.15, 0.15)		8,840	83 (151)		1,513	9 (14)	
	No			2.15 (3.47)			-0.10 (1.10)
	Yes, partial			-0.97 (1.01)			0.56 (0.75)
	Yes, full			-0.77 (0.93)			0.43 (0.74)
(-0.11, 0.11)		8,330	82 (152)		1,255	9 (13)	
	No			2.27 (4.64)			-0.77 (1.31)
	Yes, partial			-2.34 (1.33)			-0.16 (0.89)
	Yes, full			-2.36 (1.23)			-0.21 (0.89)

Notes: Full covariates include the number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Partial covariates include every variable in full covariate list except for star rating release fixed effects. Standard errors were clustered at the home health agency level.

Table B.7: Regression discontinuity estimates from a negative binomial 2 model.

		Heap					Not heaped				
	Covariates	N agencies	Prior Quarter New Patients, mean (SD)		Marginal Effects (SE)		N agencies	Prior Quarter New Patients, mean (SD)		Marginal Effects (SE)	
Pooled (-0.25, 0.25)											
	No	8,998	83	(157)	2.64	(3.16)	1,818	9	(15)	-1.19	(1.04)
	Yes	8,971	84	(157)	-1.37	(0.96)	1,773	10	(15)	0.35	(0.80)
By threshold [-0.25, 0.25]											
1 vs. 1.5											
	No	174	17	(18)	12.04	(5.93)	112	8	(10)	0.47	(2.17)
	Yes	167	18	(18)	-20.74	(12.47)	108	8	(10)	1.33	(2.49)
1.5 vs. 2											
	No	1,086	27	(38)	2.70	(3.41)	413	7	(10)	-1.26	(1.39)
	Yes	1,065	27	(38)	1.56	(1.51)	398	8	(10)	0.96	(1.07)
2 vs. 2.5											
	No	2,592	39	(61)	3.48	(2.51)	602	7	(10)	0.57	(1.61)
	Yes	2,568	39	(62)	0.44	(1.23)	576	8	(10)	-2.57	(1.22)
2.5 vs. 3											
	No	4,053	69	(179)	6.48	(4.68)	575	9	(12)	-1.75	(1.79)
	Yes	4,019	69	(179)	-1.25	(1.23)	554	9	(12)	1.38	(1.13)
3 vs. 3.5											
	No	4,713	94	(165)	6.19	(6.27)	436	10	(14)	1.09	(2.35)
	Yes	4,687	95	(165)	-1.78	(1.67)	421	11	(14)	-0.12	(2.04)
3.5 vs. 4											
	No	4,097	108	(182)	3.03	(7.41)	297	12	(18)	0.77	(4.11)
	Yes	4,071	109	(183)	-0.04	(1.77)	284	13	(19)	4.63	(2.89)
4 vs. 4.5											
	No	2,582	99	(147)	-1.43	(8.09)	172	14	(26)	-19.24	(12.59)
	Yes	2,573	100	(147)	1.83	(1.81)	169	14	(26)	-8.01	(5.36)
4.5 vs. 5											
	No	1,183	71	(109)	-6.03	(8.10)	76	16	(36)	-5.23	(10.78)
	Yes	1,180	71	(109)	-2.61	(2.91)	71	18	(38)	19.88	(14.30)

Notes: Covariates include linear and quadratic terms for number of measures used to calculate the star ratings, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Standard errors were clustered at the home health agency level. Models also include offsets for the log of the number of new patients in the prior quarter.

Table B.8: Assessment of evidence of violation of parallel pre-trends.

	Estimate (SE)
Highest rated	0.28 (0.20)
Pre-rating quarter	-0.1185 (0.0099)
Quarter 2	0.947 (0.022)
Highest Rated x Pre-ratings	-0.009 (0.045)
Highest Rated x Quarter 2	0.104 (0.087)
N	
HHA-ZIP pairs	185,454
HHAs	8,610

Notes: The model controls for linear terms for unrounded star ratings, number of home health competitors in each ZIP code, percent of full Medicare-Medicaid dually eligible patients in each ZIP-agency pair, percent of partial Medicare-Medicaid dually eligible patients in each ZIP-agency pair, percent of patients discharged from an inpatient source in the past 14 days in each ZIP-agency pair, percent of patients with end stage renal disease in each ZIP-agency pair, percent female in each ZIP-agency pair, percent white in each ZIP-agency pair, percent black in each ZIP-agency pair, percent Hispanic in each ZIP-agency pair, mean age of Medicare patients in each ZIP-agency pair, and percent of Medicare fee-for-service patients in each ZIP-agency pair. Standard errors are clustered at the home health agency level.

Table B.9: Effects of becoming highest-ranked option within a ZIP code using a negative binomial-2 model.

	N HHA-ZIP pairs (agencies)	Marginal Effect (SE)
Overall	189,503 (8,610)	0.101 (0.055)
Quarter 1 rating		
2	10,948 (730)	-0.09 (0.42)
2.5	26,413 (1,454)	0.35 (0.18)
3	45,880 (1,975)	0.28 (0.18)
3.5	50,317 (1,997)	0.23 (0.13)
4	33,943 (1,345)	0.26 (0.13)
4.5	16,491 (719)	-0.37 (0.17)
No rating change	133,897 (5,764)	0.127 (0.077)
Rating change	5,784 (224)	-0.09 (0.18)
7 or fewer agencies	54,513 (4,827)	0.13 (0.08)
More than 28 agencies	46,746 (4,078)	-0.05 (0.12)

Notes: All models control for linear and quadratic terms for unrounded star ratings, number of home health competitors in each ZIP code, percent of full Medicare-Medicaid dually eligible patients, percent of partial Medicare-Medicaid dually eligible patients, percent of patients discharged from an inpatient source in the past 14 days, percent of patients with end stage renal disease, percent female, percent white, percent black, percent Hispanic, mean age of Medicare patients, percent of Medicare fee-for-service patients, total number of Medicare home health patients in each ZIP code. Models also include offsets for the log of the number of new patients in the prior quarter. Standard errors are clustered at the home health agency level. 7+ agencies and >28 agencies refer to the bottom and top quartiles of number of agencies serving each ZIP code.

Table B.10: Association of star ratings and number of patients using naïve linear regression.

Number of new patients per quarter, mean (SE)	
Overall	75.77 (0.62)
1 star	8.75 (1.20)
1.5 stars	16.82 (0.69)
2 stars	26.57 (0.60)
2.5 stars	43.38 (0.75)
3 stars	77.55 (1.61)
3.5 stars	99.49 (1.51)
4 stars	99.62 (1.66)
4.5 stars	80.18 (1.77)
5 stars	62.77 (3.06)
Naive linear regression, estimate (SE)	
No Covariates	25.96 (1.30)
Covariates	0.91 (0.14)

Notes: The naïve linear regression refers to a simple model evaluating the association between the star rating and number of patients, conditional on covariates across the entire sample and over the first six quarterly star rating releases. The covariates include the number of measures used to calculate the star ratings, number of new patients in the prior quarter, number of Medicare patients in the prior quarter, agency organizational characteristics (agency age, chain affiliation, for-profit status), prior quarter agency Medicare patient characteristics (average age, percent female, percent white, percent black, percent Hispanic, percent that were fee-for-service enrollees, percent that were fully enrolled in Medicaid, percent that were partially enrolled in Medicaid, percent that were discharged from an inpatient institution, percent that had end stage renal disease), and star rating release fixed effects. Standard errors are clustered at the agency level.

Figure B.1: Relationship between unrounded star ratings and star ratings in the first six quarters of the Home Health Star Ratings program.

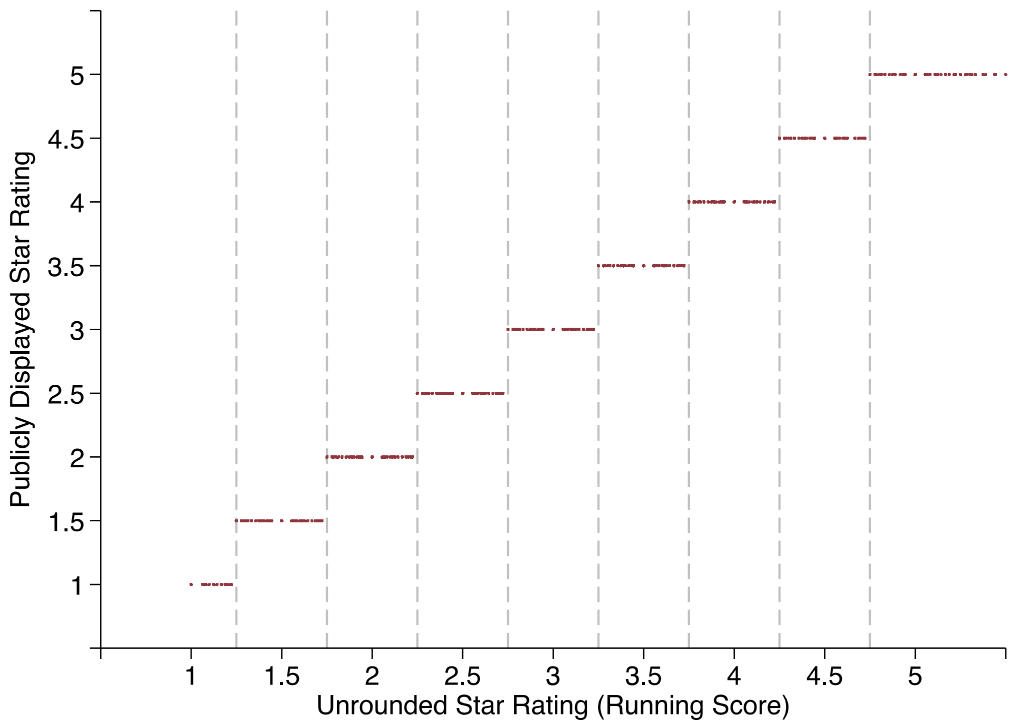
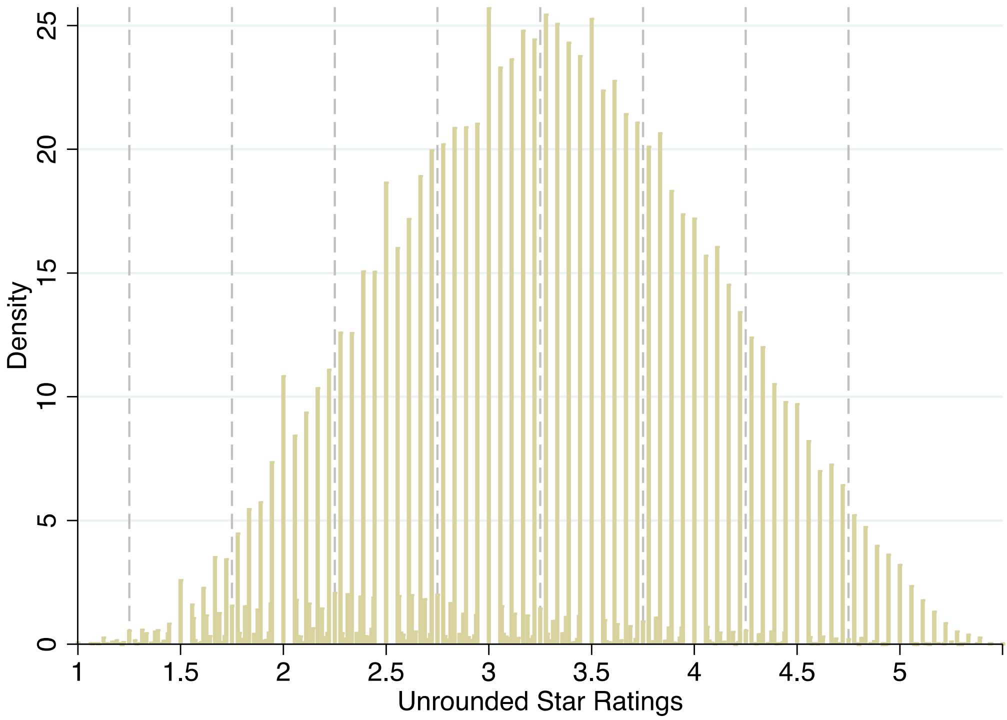


Figure B.2: Distribution of unrounded star ratings in the first six quarters of the Home Health Star Ratings program.



Appendix C

Appendix to Choosing Home Health Care on the Basis of Medicare Star Ratings: Is There Value in Picking the Best?

Table C.1: Characteristics of ZIP codes by instrumental variable values.

	Overall Sample	Instrumental Variable Values	
		Closer to top agency	Far/equidistance to top agency
Number of observations	1,870,080	320,313 (17%)	1,549,767 (83%)
Number of ZIP-quarter pairs			
Overall	67,094	17,021 (25%)	50,073 (75%)
<=3.5 optimal stars	15,272	5,636 (37%)	9,636 (63%)
4 optimal stars	22,484	6,706 (30%)	15,778 (70%)
4.5 optimal stars	20,124	3,800 (19%)	16,324 (81%)
5 optimal stars	9,214	879 (10%)	8,335 (90%)
Percentage of patients treated by agencies located in same ZIP	10.7	10.7	10.8
Percentage of agencies serving a ZIP code per quarter located in same ZIP	4.3	4.2	3.2
Distance to nearest optimal Option			
Mean	14.6	6.9	16.2
Median	9.5	4.2	10.8
Range	(0, 304.6)	(0, 196.2)	(0, 304.6)
Percentage in same ZIP	10.2	28.1	6.4
Distance to nearest suboptimal Option			
Mean	6.9	18.2	4.6
Median	3.5	11.9	2.6
Range	(0, 359.8)	(0.1, 359.8)	(0, 118.9)
Percentage in same ZIP	31.2	0.0	37.6
Differential Distance			
Mean	7.7	-11.3	11.6
Median	4.2	-4.9	6.0
Range	(-355.6, 287.2)	(-355.6, -0.1)	(0.0, 287.2)

Table C.2: Characteristics of home health agencies serving ZIP codes in the sample.

	Overall sample	By rating of top agencies in ZIP code			
		1 to 3.5	4	4.5	5
ZIP-quarter pairs, number	67,094	15,272	22,484	20,124	9,214
No. of total agencies per ZIP, mean (SE)	7.21	3.62	5.30	7.98	16.10
	(0.03)	(0.02)	(0.02)	(0.05)	(0.17)
No. of top agencies per ZIP, mean (SE)	1.489	1.542	1.599	1.389	1.346
	(0.004)	(0.007)	(0.007)	(0.006)	(0.010)
Average star of non-top agencies, mean (SE)	3.112	2.677	3.111	3.283	3.463
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Top star minus average star of suboptimal, mean (SE)	1.035	0.707	0.889	1.217	1.537
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)

Table C.3: First-stage results using dichotomized instrumental variable.

	No. of observations (%)	No. of ZIP codes	Coefficient (SE)	*F*-statistic
All observations	1,870,080	22,333	0.2372 (0.0054)	1942
Referral source				
Inpatient	1,199,955 (64)	21,249	0.2490 (0.0058)	1850
Not from inpatient	667,378 (36)	18,890	0.2141 (0.0061)	1224
Difference in ratings (top vs. others)				
<1 star difference	581,378 (42)	13,223	0.2630 (0.0104)	644
1+ star difference	794,257 (58)	14,517	0.1972 (0.0091)	472
Top star in ZIP code				
1 to 3.5 stars	244,515 (13)	8,775	0.2588 (0.016)	272
4 to 5 stars	1,625,565 (87)	18,215	0.1861 (0.0058)	1014

Notes: Singleton observations (ZIP codes with one observation) are not analyzed so inpatient referral subgroups do not add up to total. Difference in ratings subgroups focus on ZIP codes with top ratings between 3.5 and 4.5 stars where there is common support. In contrast, a falsification test using this IV to predict drug coverage yields coefficient and standard error of 0.0078 (0.0071), with F-statistic of 0.27.

Table C.4: Population characteristics across ZIP codes in the sample.

	Overall sample	By rating of top agencies in ZIP code			
		1 to 3.5	4	4.5	5
ZIP-quarter pairs, number	67,094	15,272	22,484	20,124	9,214
Persons per 10,000 sq. meters of land, mean (SE)	14.11	6.28	8.20	14.97	20.84
	(0.04)	(0.10)	(0.07)	(0.09)	(0.06)
Percent of households in urban residence, mean (SE)	78.28	53.03	67.29	81.17	93.39
	(0.05)	(0.18)	(0.11)	(0.08)	(0.05)
Race					
Percent white, mean (SE)	73.86	79.90	77.95	75.32	66.75
	(0.03)	(0.09)	(0.06)	(0.05)	(0.06)
Percent black, mean (SE)	13.66	11.26	13.26	12.22	16.43
	(0.03)	(0.08)	(0.06)	(0.05)	(0.06)
Percent other race, mean (SE)	12.48	8.84	8.79	12.45	16.82
	(0.02)	(0.04)	(0.03)	(0.03)	(0.04)
Language spoken at home					
Percent English only, mean (SE)	78.81	87.91	87.12	79.39	68.13
	(0.03)	(0.06)	(0.04)	(0.05)	(0.06)
Percent non-English language, mean (SE)	21.19	12.09	12.88	20.61	31.87
	(0.03)	(0.06)	(0.04)	(0.05)	(0.06)
Percent Spanish language, mean (SE)	13.35	7.39	7.45	13.11	20.58
	(0.03)	(0.06)	(0.04)	(0.04)	(0.06)
Percent Indo-European language, mean (SE)	3.80	2.42	2.87	3.61	5.27
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Percent Asian language, mean (SE)	3.08	1.67	1.82	2.98	4.73
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Education					
Percent with high school degree or higher, mean (SE)	86.39	85.99	87.36	86.89	85.24
	(0.01)	(0.04)	(0.02)	(0.02)	(0.03)
Percent with bachelor’s degree or higher, mean (SE)	29.16	24.41	27.71	30.35	30.81
	(0.02)	(0.07)	(0.05)	(0.04)	(0.04)
Labor Force and Unemployment					
Labor force participation rate, mean (SE)	62.31	60.46	62.13	62.75	62.65
	(0.01)	(0.04)	(0.02)	(0.02)	(0.02)
Unemployment rate, mean (SE)	7.90	7.70	7.49	7.52	8.72
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Household income in 2016 dollars, mean (SE)	76,776	70,775	75,076	79,035	77,926
	(49)	(137)	(94)	(86)	(90)
Household income Gini index (scaled to 100), mean (SE)	43.93	43.32	43.38	43.74	44.79
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)

Table C.5: Alternative specifications of the first-stage regression using linear and linear spline differential distance as instruments.

	Linear		Linear spline	
	(1,870,080 patients 22,333 ZIP codes)		(1,870,080 patients 22,333 ZIP codes)	
Days in initial home health care				
Unadjusted Mean (SE)	50.42	(0.19)	50.42	(0.19)
Coefficient (SE)	-7.41	(1.12)	-7.50	(1.04)
95% CI	(-9.60,	-5.22)	(-9.54,	-5.45)
Days alive without health care				
Unadjusted Mean (SE)	145.18	(0.11)	145.18	(0.11)
Coefficient (SE)	2.75	(0.76)	2.83	(0.72)
95% CI	(1.26,	4.24)	(1.43,	4.24)
Days in an institutional setting after end of care				
Unadjusted Mean (SE)	7.34	(0.02)	7.34	(0.02)
Coefficient (SE)	-0.32	(0.29)	-0.45	(0.29)
95% CI	(-0.90,	0.25)	(-1.01,	0.11)
Acute care hospital				
Unadjusted Mean (SE)	2.511	(0.009)	2.511	(0.009)
Coefficient (SE)	-0.094	(0.087)	-0.086	(0.085)
95% CI	(-0.26,	0.08)	(-0.25,	0.08)
Long-term care hospital				
Unadjusted Mean (SE)	0.1633	(0.0029)	0.1633	(0.0029)
Coefficient (SE)	-0.013	(0.033)	-0.009	(0.031)
95% CI	(-0.078,	0.052)	(-0.070,	0.051)
Inpatient rehabilitation				
Unadjusted Mean (SE)	0.1479	(0.0027)	0.1479	(0.0027)
Coefficient (SE)	-0.024	(0.022)	-0.023	(0.022)
95% CI	(-0.067,	0.020)	(-0.066,	0.020)
Nursing home				
Unadjusted Mean (SE)	4.515	(0.021)	4.515	(0.021)
Coefficient (SE)	-0.19	(0.25)	-0.33	(0.25)
95% CI	(-0.68,	0.30)	(-0.82,	0.15)
Days in home health after end of care				
Unadjusted Mean (SE)	16.13	(0.11)	16.13	(0.11)
Coefficient (SE)	-2.05	(0.51)	-2.08	(0.47)
95% CI	(-3.04,	-1.05)	(-2.99,	-1.16)
Days deceased after end of care				
Unadjusted Mean (SE)	11.36	(0.04)	11.36	(0.04)
Coefficient (SE)	-0.38	(0.56)	-0.30	(0.53)
95% CI	(-1.47,	0.71)	(-1.34,	0.73)

BIBLIOGRAPHY

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. “When Should You Adjust Standard Errors for Clustering?” w24003. NBER Working Paper.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princton University Press.

Arbor Research Collaborative for Health, and L&M Policy Research. 2018. “Evaluation of the Home Health Value-Based Purchasing (HHVBP) Model: First Annual Report.” https://innovation.cms.gov/initiatives/home-health-value-based-purchasing-model.

Avalere. 2015. “Home Health Chartbook 2015.” http://ahhqi.org/images/uploads/AHHQI_2015_Chartbook_FINAL_October.pdf.

Baier, Rosa R., Andrea Wysocki, Stefan Gravenstein, Emily Cooper, Vincent Mor, and Melissa Clark. 2015. “A Qualitative Study of Choosing Home Health Care After Hospitalization: The Unintended Consequences of ‘Patient Choice’ Requirements.” *Journal of General Internal Medicine* 30 (5): 634–40. https://doi.org/10.1007/s11606-014-3164-7.

Baiocchi, Michael, Jing Cheng, and Dylan S Small. 2019. “Instrumental Variable Analysis BT - Health Services Evaluation.” In , edited by Adrian Levy, Sarah Goring, Constantine Gatsonis, Boris Sobolev, Ewout van Ginneken, and Reinhard Busse, 479–522. New York, NY: Springer US. https://doi.org/10.1007/978-1-4939-8715-3_32.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2016. “Heaping-Induced Bias in Regression-Discontinuity Designs.” *Economic Inquiry* 54: 268–93. https://doi.org/10.1111/ecin.12225.

Bastani, Hamsa, Joel Goh, and Mohsen Bayati. 2019. “Evidence of Upcoding in Pay-for-Performance Programs.” *Management Science* 65 (3): 1042–60. https://doi.org/https://doi.org/10.1287/mnsc.2017.2996.

Bell, Max, Lars I Eriksson, Tobias Svensson, Linn Hallqvist, Fredrik Granath, Jennifer Reilly, and Paul S Myles. 2019. “Days at Home after Surgery: An Integrated and Efficient Outcome Measure for Clinical Trials and Quality Assurance.” *EClinicalMedicine* 11 (May): 18–26. https://doi.org/10.1016/j.eclinm.2019.04.011.

Black, Dan A., Joonhwi Joo, Robert LaLonde, Jeffrey A. Smith, and Evan J. Taylor. 2017. “Simple Tests for Selection: Learning More from Instrumental Variables.” *CESifo Working Paper Series No. 6392*.

Bruhn, Miriam, and David McKenzie. 2009. “In Persuit of Balance: Randomization in Practice in Development Field Experiments.” *American Economic Journal: Applied Economics* 1 (4): 200–232. https://doi.org/10.1257/app.1.4.200.

Burke, Laura G, E John Orav, Jie Zheng, and Ashish K Jha. 2019. “Healthy Days at Home: A Novel Population-Based Outcome Measure.” *Healthcare*, 100378. https://doi.org/https://doi.org/10.1016/j.hjdsi.2019.100378.

Cattaneo, Matias D., Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare. 2016. “Interpreting Regression Discontinuity Designs with Multiple Cutoffs.” *The Journal of Politics* 78 (4): 1229–48. https://doi.org/10.1086/686802.

Centers for Medicare & Medicaid Services. n.d. “Report to Congress Medicare Home Health Study: An Investigation on Access to Care and Payment for Vulnerable Patient Populations.” https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HomeHealthPPS/Downloads/HH-Report-to-Congress.pdf.

———. 2015a. “Frequently Asked Questions (FAQs) about the Quality of Patient Care Star Ratings.”

———. 2015b. “OASIS-C1/ICD-10 Guidance Manual.”

———. 2015c. “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 (214): 68623–719. http://www.ncbi.nlm.nih.gov/pubmed/26552111.

———. 2016. “Home Health Agency Quality Measures: Technical Documentation of OASIS-Based Patient Outcome Measures Revision 5.”

———. 2018a. “Home Health Agencies Data 2016.” CMS.Gov. 2018.

———. 2018b. “Home Health Star Ratings.” 2018. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/HHQIHomeHealthStarRatings.html.

———. 2019a. “Chapter 10 - Home Health Agency Billing.” *Medicare Claims Processing Manual* 2977.

———. 2019b. “Home Health Compare Datasets.” 2019.

Centers for Medicare and Medicaid Services (CMS). 2019. “Medicare and Medicaid Programs; Revisions to Requirements for Discharge PLanning for Hospitals, Critical Access Hospitals, and Home Health Agencies, and Hospital and Critical Access Hospital Changes to Promote Innovation, Flexibility, and Improvement in P.” *Federal Register* 84: 51836–84.

Christianson, Jon B., Sheila Leatherman, and Kim Sutherland. 2008. “Lessons From Evaluations of Purchaser Pay-for-Performance Programs.” *Medical Care Research and Review* 65 (6_suppl): 5S-35S. https://doi.org/10.1177/1077558708324236.

Cornell, Portia Y, David C Grabowski, Edward C Norton, and Momotazur Rahman. 2019. “Do Report Cards Predict Future Quality? The Case of Skilled Nursing Facilities.” *Journal of Health Economics* 66: 208–21. https://doi.org/https://doi.org/10.1016/j.jhealeco.2019.05.008.

Cuckler, Gigi A., Andrea M. Sisko, John A. Poisal, Sean P. Keehan, Sheila D. Smith, Andrew J.

Madison, Christian J. Wolfe, and James C. Hardesty. 2018. "National Health Expenditure Projections, 2017-26: Despite Uncertainty, Fundamentals Primarily Drive Spending Growth." *Health Affairs* 37 (3): 482–92. https://doi.org/10.1377/hlthaff.2017.1655.

Darden, M., and I. M. McCarthy. 2015. "The Star Treatment: Estimating the Impact of Star Ratings on Medicare Advantage Enrollments." *Journal of Human Resources* 50 (4): 980–1008. https://doi.org/10.3368/jhr.50.4.980.

David, Guy, Evan Rawley, and Daniel Polsky. 2013. "Integration and Task Allocation: Evidence from Patient Care." *Journal of Economics & Management Strategy* 22: 617–39. https://doi.org/10.1111/jems.12023.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210: 2–21. https://doi.org/https://doi.org/10.1016/j.socscimed.2017.12.005.

Doran, Tim, Kristin A. Maurer, and Andrew M. Ryan. 2017. "Impact of Provider Incentives on Quality and Value of Health Care." *Annual Review of Public Health* 38 (1): 449–65. https://doi.org/10.1146/annurev-publhealth-032315-021457.

Ellenbecker, Carol Hall, Linda Samia, Margaret J. Cushman, and Kristine Alster. 2008. *Patient Safety and Quality in Home Health Care*. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. http://www.ncbi.nlm.nih.gov/pubmed/21328733.

Faber, Marjan, Marije Bosch, Hub Wollersheim, Sheila Leatherman, and Richard Grol. 2009. "Public Reporting in Health Care: How Do Consumers Use Quality-of-Care Information?" *Medical Care* 47 (1): 1–8. https://doi.org/10.1097/MLR.0b013e3181808bb5.

Figueroa, Jose F, Yusuke Tsugawa, Jie Zheng, E John Orav, and Ashish K Jha. 2016. "Association between the Value-Based Purchasing Pay for Performance Program and Patient Mortality in US Hospitals: Observational Study." *BMJ*, May, i2214. https://doi.org/10.1136/bmj.i2214.

Frandsen, B. 2017. "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable Is Discrete." In *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, Vol. 38)*, 281–315. Emerald Publishing Limited. https://doi.org/https://doi.org/10.1108/S0731-905320170000038012.

Fuloria, Prashant C., and Stefanos A. Zenios. 2001. "Outcomes-Adjusted Reimbursement in a Health-Care Delivery System." *Management Science* 47 (6): 735–51.

Golden, Brian R., and Frank A. Sloan. 2008. "Physician Pay for Performance: Alternative Perspectives." In *Incentives and Choice in Health Care*, edited by Frank A. Sloan and Hirschel Kasper, 289–300. Cambridge, MA: MIT Press.

Greene, William H. 2018. *Econometric Analysis*. New York: Pearson.

Groff, Adam C, Carrie H Colla, and Thomas H Lee. 2016. "Days Spent at Home — A Patient-Centered Goal and Outcome." *New England Journal of Medicine* 375 (17): 1610–12. https://doi.org/10.1056/NEJMp1607206.

Health Care Financing Administration. 2000. "Medicare Program; Prospective Payment System for Home Health Agencies." *Federal Register* 65 (128): 41128–214. https://www.govinfo.gov/content/pkg/FR-2000-07-03/pdf/00-16432.pdf.

Hibbard, Judith H., and Ellen Peters. 2003. "Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches That Facilitate the Use of Information in Choice." *Annual Review of Public Health* 24 (1): 413–33. https://doi.org/10.1146/annurev.publhealth.24.100901.141005.

Institute of Medicine. 2007. *Rewarding Provider Performance*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/11723.

Keohane, Laura M., Salama Freed, David G. Stevenson, Sunita Thapa, Lucas Stewart, and Melinda B. Buntin. 2018. "Trends in Postacute Care Spending Growth During the Medicare Spending Slowdown."

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119. https://doi.org/10.1111/j.1468-0262.2007.00733.x.

Konetzka, R. Tamara, Elizabeth A. Stuart, and Rachel M. Werner. 2018. "The Effect of Integration of Hospitals and Post-Acute Care Providers on Medicare Payment and Patient Outcomes." *Journal of Health Economics*, February. https://doi.org/10.1016/j.jhealeco.2018.01.005.

Kowalski, Amanda E. 2016. "Doing More When You're Running Late: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments." *NBER Working Paper Series No. 22362*.

———. 2018. "How to Examine External Validity Within an Experiment." *NBER Working Paper Series No. 24834*.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics* 142: 655–74. https://doi.org/10.1016/j.jeconom.2007.05.003.

Leuven, Edwin, and Barbara Sianesi. 2018. "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing." http://ideas.repec.org/c/boc/bocode/s432001.html.

Li, Chia-Lin, Hsing-Yi Chang, Hui-Hsuan Wang, and Yuh-Bin Bai. 2011. "Diabetes, Functional Ability, and Self-Rated Health Independently Predict Hospital Admission within One Year among Older Adults: A Population Based Cohort Study." *Archives of Gerontology and Geriatrics* 52 (2): 147–52.

Li, Jun. 2019. "Pay-for-Performance in Health Care: Evidence from a Nationwide Randomized Experiment in the Home Health Sector." *AEA RCT Registry*. https://doi.org/10.1257/rct.4159-2.0.

Li, Jun, and Edward C. Norton. 2019. "Pay-for-Performance and Long-Term Care." *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. https://doi.org/http://dx.doi.org/10.1093/acrefore/9780190625979.013.407.

Loomer, Lacy, Momotazur Rahman, Tracy M. Mroz, Pedro L. Gozalo, and Vincent Mor. 2020. "Do Higher Payments Increase Access to Post-Acute Home Health Care for Rural Medicare Beneficiaries?" *Journal of the American Geriatrics Society* 68: 663–64. https://doi.org/10.1111/jgs.16332.

Luppa, M., T. Luck, S. Weyerer, H.-H. Konig, E. Brahler, and S. G. Riedel-Heller. 2010. "Prediction of Institutionalization in the Elderly. A Systematic Review." *Age and Ageing* 39 (1): 31–38. https://doi.org/10.1093/ageing/afp202.

Lynn, Joanne, and Jane Brock. 2019. "Time at Home as a Quality Measure." *EClinicalMedicine* 11 (May): 5–6. https://doi.org/10.1016/j.eclinm.2019.05.006.

McCarthy, Ian M., and Michael Darden. 2017. "Supply-Side Responses to Public Quality Ratings: Evidence from Medicare Advantage." *American Journal of Health Economics* 3 (2): 140–64. https://doi.org/10.1162/AJHE_a_00070.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714. https://doi.org/10.1016/j.jeconom.2007.05.005.

Medicare Payment Advisory Commission. 2018. "Chapter 5: Encouraging Medicare Beneficiaries to Use Higher Quality Post-Acute Care Providers."

———. 2019a. "Report to the Congress: Medicare and the Health Care Delivery System."

———. 2019b. "Report to the Congress: Medicare Payment Policy."

MedPac. 2018. "A Data Book: Health Care Spending and the Medicare Program." http://medpac.gov/docs/default-source/data-book/jun18_databookentirereport_sec.pdf?sfvrsn=0.

Murtaugh, Christopher M., Timothy R. Peng, Stanley Moore, and Gill A. Maduro. 2008. "Assessing Home Health Care Quality for Post-Acute and Chronically Ill Patients: Final Report."

National Center for Health Statistics. 2019. "Long-Term Care Providers and Services Users in the United States, 2015-2016." *Vital & Health Statistics. Series 3, Number 43*. https://www.cdc.gov/nchs/data/series/sr_03/sr03_43-508.pdf.

NC Rural Health Research Program. 2017. "Rural Health Snapshot."

Norton, Edward C., Jun Li, Anup Das, and Lena M. Chen. 2018. “Moneyball in Medicare.” *Journal of Health Economics*, August. https://doi.org/10.1016/j.jhealeco.2017.07.006.

O’Connor, Melissa, and Joan K. Davitt. 2012. “The Outcome and Assessment Information Set (OASIS): A Review of Validity and Reliability.” *Home Health Care Services Quarterly* 31 (4): 267–301. https://doi.org/10.1080/01621424.2012.703908.

Perraillon, Marcelo Coca, R. Tamara Konetzka, Daifeng He, and Rachel M. Werner. 2017. “Consumer Response to Composite Ratings of Nursing Home Quality.” *American Journal of Health Economics*, December, 1–36. https://doi.org/10.1162/ajhe_a_00115.

Peters, Ellen, Nathan Dieckmann, Anna Dixon, Judith H. Hibbard, and C. K. Mertz. 2007. “Less Is More in Presenting Quality Information to Consumers.” *Medical Care Research and Review* 64 (2): 169–90. https://doi.org/10.1177/10775587070640020301.

Pope, Gregory, John Kautter, Melvin Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. 2011. “Evaluation of the CMS-HCC Risk Adjustment Model.” https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/Evaluation_Risk_Adj_Model_2011.pdf.

Porell, Frank W., Korbin Liu, and David P. Brungo. 2006. “Agency and Market Area Factors Affecting Home Health Agency Supply Changes.” *Health Services Research* 41 (5): 1847–75. https://doi.org/10.1111/j.1475-6773.2006.00561.x.

Probst, Janice C., Samuel Towne, Jordan Mitchell, Kevin J. Bennett, and Robert Chen. 2014. “Home Health Care Agency Availability in Rural Counties.”

Reid, Rachel O., Partha Deb, Benjamin L. Howell, and William H. Shrank. 2013. “Association Between Medicare Advantage Plan Star Ratings and Enrollment.” *JAMA* 309 (3): 267. https://doi.org/10.1001/jama.2012.173925.

Richardson, Samuel S. 2012. “Integrating Pay-for-Performance into Health Care Payment Systems.”

Rosenthal, Meredith B., and Richard G. Frank. 2006. “What Is the Empirical Basis for Paying for Quality in Health Care?” *Medical Care Research and Review* 63 (2): 135–57. https://doi.org/10.1177/1077558705285291.

Ryan, Andrew M., Sam Krinsky, Kristin A. Maurer, and Justin B. Dimick. 2017. “Changes in Hospital Quality Associated with Hospital Value-Based Purchasing.” *New England Journal of Medicine* 376 (24): 2358–66. https://doi.org/10.1056/NEJMsa1613412.

Ryan, Andrew M., Christopher P. Tompkins, Adam A. Markovitz, and Helen R. Burstin. 2016. “Linking Spending and Quality Indicators to Measure Value and Efficiency in Health Care.” *Medical Care Research and Review*, May. https://doi.org/10.1177/1077558716650089.

Scott, Anthony, Miao Liu, and Jongsay Yong. 2018. “Financial Incentives to Encourage Value-

Based Health Care." *Medical Care Research and Review* 75 (1): 3–32. https://doi.org/10.1177/1077558716676594.

Singh, Gopal K., and Mohammad Siahpush. 2014. "Widening Rural-Urban Disparities in Life Expectancy, US, 1969-2009." *American Journal of Preventive Medicine* 46 (2): 19–29. https://doi.org/doi.org/10.1016/j.amepre.2013.10.017.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557. https://doi.org/10.2307/2171753.

Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25 (1): 1–21.

Sukul, Devraj, Geoffrey J Hoffman, Ushapoorna Nuliyalu, Julia R Adler-Milstein, Bill Zhang, Justin B Dimick, and Andrew M Ryan. 2019. "Association Between Medicare Policy Reforms and Changes in Hospitalized Medicare Beneficiaries' Severity of Illness." *JAMA Network Open* 2 (5): e193290–e193290. https://doi.org/10.1001/jamanetworkopen.2019.3290.

Tao, Hong, and Carol H. Ellenbecker. 2013. "Is OASIS Effective in Predicting Rehospitalization for Home Health Care Elderly Patients?" *Home Health Care Management & Practice* 25 (6).

US Department of Health and Human Services. n.d. "Report to Congress: Plan to Implement a Medicare Home Health Agency Value-Based Purchasing Program."

Wang, David E., Yusuke Tsugawa, Jose F. Figueroa, and Ashish K. Jha. 2016. "Association Between the Centers for Medicare and Medicaid Services Hospital Star Rating and Patient Outcomes." *JAMA Internal Medicine* 176 (6): 848–50. https://doi.org/10.1001/jamainternmed.2016.0784.

Wang, Yun, Erica S. Spatz, Maliha Tariq, Suveen Angraal, and Harlan M. Krumholz. 2017. "Home Health Agency Performance in the United States: 2011-15." *Journal of the American Geriatrics Society* 65 (12): 2572–79. https://doi.org/10.1111/jgs.14987.

Weisbrod, Burton A. 1991. "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment." *Journal of Economic Literature* 29 (2): 523–52.

Werner, Rachel M., Norma B. Coe, Mingyu Qi, and R. Tamara Konetzka. 2019. "Patient Outcomes After Hospital Discharge to Home with Home Health Care vs to a Skilled Nursing Facility." *JAMA Internal Medicine* 179: 617–23. https://doi.org/doi.org/10.1001/jamainternmed.2018.7998.

Werner, Rachel M., R. Tamara Konetzka, and Daniel Polsky. 2013. "The Effect of Pay-for-Performance in Nursing Homes: Evidence from State Medicaid Programs." *Health Services Research* 48 (4): 1393–1414. https://doi.org/10.1111/1475-6773.12035.

———. 2016. “Changes in Consumer Demand Following Public Reporting of Summary Quality Ratings: An Evaluation in Nursing Homes.” *Health Services Research* 51 (June): 1291–1309. https://doi.org/10.1111/1475-6773.12459.

Werner, Rachel M, Edward C Norton, R Tamara Konetzka, and Daniel Polsky. 2012. “Do Consumers Respond to Publicly Reported Quality Information? Evidence from Nursing Homes.” *Journal of Health Economics* 31 (1): 50–61. https://doi.org/10.1016/j.jhealeco.2012.01.001.

Westfall, Peter H., and Stanley S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. New York: Wiley.

Wolff, Jennifer L., Ann Meadow, Carlos O. Weiss, Cynthia M. Boyd, and Bruce Leff. 2008. “Medicare Home Health Patients’ Transitions Through Acute and Post-Acute Care Settings.” *Medical Care* 46 (11): 1188–93.

WWAMI Rural Health Research Center. 2018. “Different Populations Serviced by the Medicare Home Health Benefit: Comparison of Post-Acute versus Community-Entry Home Health in Rural Areas.” https://depts.washington.edu/fammed/rhrc/wp-content/uploads/sites/4/2018/07/RHRC_PB165_Mroz.pdf.

Wysocki, Andrea, and Valerie Cheh. 2019. “Patterns of Care and Home Health Utilization for Community-Admitted Medicare Patients.”