

SEPTEMBER 2026

The Effects of Hospice on Medicare Spending

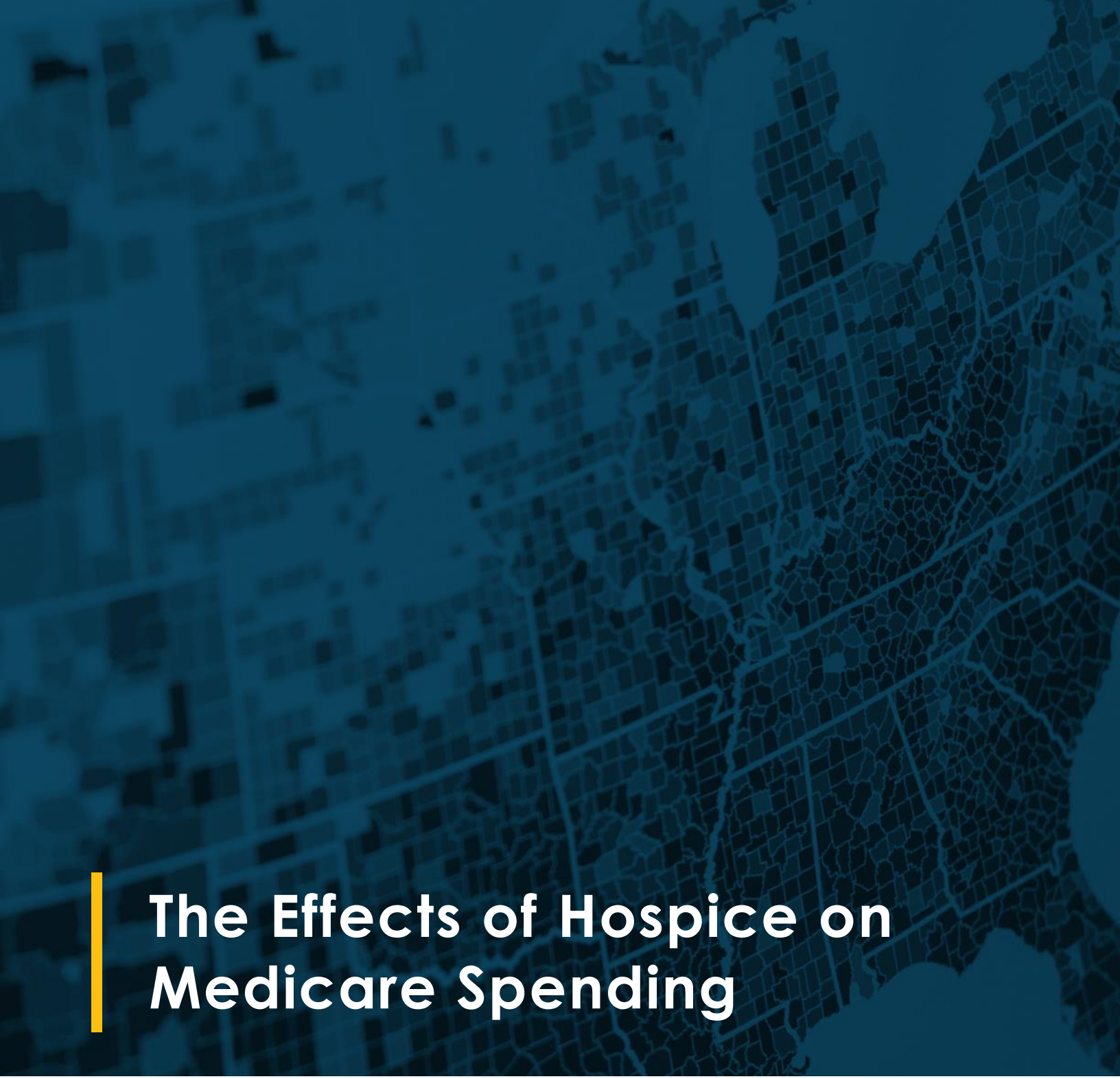
A report by the Urban Institute for the Medicare Payment Advisory Commission



The views expressed in this report are those of the authors. No endorsement by MedPAC is intended or should be inferred.

MEDPAC
Medicare Payment Advisory Commission

425 I Street, NW | Suite 701 | Washington, DC 20001 | (202) 220-3700 | www.medpac.gov



The Effects of Hospice on Medicare Spending

Bowen Garrett **Doug Wissoker** **Laura Skopec**

URBAN INSTITUTE

RESEARCH REPORT

September 2026



ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization founded on one simple idea: To improve lives and strengthen communities, we need practices and policies that work. For more than 50 years, that has been our charge. By equipping changemakers with evidence and solutions, together we can create a future where every person and community has the opportunity and power to thrive.

Copyright © September 2026. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.

Contents

Acknowledgments iv

Executive Summary v

Key Findings vii

Discussion xi

The Effects of Hospice on Medicare Spending 1

Data and Methods 4

Changes in Decedent-to-Survivor Spending Ratios 5

Individual-Level Analyses Estimating the Effect of Hospice Use on Medicare Spending 7

Zip-Code-Year-Level Analysis 18

Results 21

Changes in Patterns of Hospice Use and Decedent-to-Survivor Spending 21

Individual-Level Analysis 24

Zip-Code-Level Analysis 55

Findings in the Context of the Literature on Effects of Hospice on Spending 64

Appendix A: Aggregate Analyses by Core Based Statistical Areas 71

Appendix B: Descriptive Statistics for Predictors Used in Individual-Level Propensity Models 77

Appendix C: Modeling Potential Heterogeneous Effects of Hospice Across Three Condition Subgroups 81

Appendix D. Exploratory Analyses Using a Leave-One-Out IV Approach 88

Notes 91

References 93

About the Authors 96

Statement of Independence 97

Acknowledgments

This report was funded by the Medicare Payment Advisory Commission. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

The authors thank Kim Neuman and Karen Stockley for collaboration and feedback throughout the project and John Engberg, Sherry Glied, Tisamarie Sherry, and Jeff Stensland for their insights and comments. The authors also thank Sarah LaCorte for editorial assistance.

Executive Summary

Under the Medicare hospice benefit, qualified beneficiaries with a life expectancy of six months or less can receive in-home or institutional end-of-life care, including palliative care, intensive comfort care, and other supports that are consistent with beneficiaries' preferences. To access the benefit, beneficiaries agree to forgo conventional care for their terminal illness provided outside the hospice benefit. The primary purpose of hospice is to provide comprehensive palliative care and support to patients with a terminal illness that aligns with their goals.[1] The hospice benefit provides Medicare beneficiaries with important end-of-life care options, including pain relief, symptom management, and psychosocial supports (MedPAC 2026), and may provide significant value to individuals regardless of costs or effects on Medicare spending. This report focuses on the net effects of hospice use on total Medicare spending. The value of the hospice benefit to enrollees and their families is also important, but it is not covered here. When the hospice benefit was added to Medicare in 1982, the Congressional Budget Office estimated it would result in net lower spending in the Medicare program.[2] However, research findings in the literature have reached conflicting conclusions on whether hospice reduces Medicare spending, reflecting, in part, the substantial heterogeneity of patients using hospice.

Past studies have consistently found lower Medicare spending associated with hospice use for cancer using a variety of methods over varying time periods (Fiala et al. 2020; Hung, Hsu, and Wang 2020; Kalidindi et al. 2020; NORC 2023; Zuckerman, Stearns, and Sheingold 2016). In contrast, some studies find that hospice use is associated with higher Medicare spending for patients with Alzheimer's disease and related dementias (ADRD) (Zuckerman, Stearns, and Sheingold 2016) and for those with very long hospice stays (Hogan 2015; Hung, Hsu, and Wang 2020; Obermeyer et al. 2014). Aldridge et al. 2023, however, found hospice use to be associated with lower or similar Medicare spending for beneficiaries with ADRD, depending on subgroup. A recent quasi-experimental study found substantially lower spending associated with hospice use for Medicare beneficiaries with ADRD, driven by higher mortality for hospice users (Gruber et al. 2025). Much of the prior literature had implicitly or explicitly assumed hospice has no effect on mortality by, for example, limiting the sample to decedents and/or retrospectively comparing spending of hospice users to nonusers over a fixed time period before

death (Hogan 2015; Hung, Hsu, and Wang 2020; Leibowitz, Tan, and Gildner 2020; NORC 2023; Zuckerman, Stearns, and Sheingold 2016).

Estimating the effect of hospice on Medicare spending is difficult, first because beneficiaries who use hospice are highly heterogeneous. They enter hospice with very different health conditions and treatment needs and obtain hospice care in both home and institutional settings. Second, beneficiaries who choose hospice can differ from those who do not in ways that are not fully captured in claims data, such as frailty, prognosis, and care preferences. Comparing spending between hospice users and nonusers can therefore reflect these underlying differences rather than the effect of hospice itself. This study addresses that challenge using methods intended to isolate the causal effect of hospice, including comparing beneficiaries using hospice with nonhospice comparison groups that are otherwise similar on observable characteristics and using instrumental variable methods to better estimate causal effects in the presence of unobservable differences in beneficiary characteristics.

This study addresses several limitations in the prior literature by employing multiple methods at both the aggregate and individual levels to assess the overall effect of hospice on Medicare spending. First, we examine national-level trends in hospice use and end-of-life spending from 2007 to 2018, extending trends reported by Hogan (2015). We then use quasi-experimental instrumental variable (IV) methods using a distance-to-hospice instrument, as in Gruber et al. (2025) and previous studies that used distance-to-provider instruments, to explore the causal effects of hospice use on Medicare spending at the individual level. Our individual-level analyses include 15–18 months of follow-up from the time a beneficiary enters hospice to partially address the criticism that many hospice studies did not include adequate follow-up (Hogan 2015). Given our focus on overall effects, we include nearly all hospice entrants ages 66 and older, not just those with specific diagnoses or those who die within a specified time frame. All analyses incorporate Part D spending, which prior studies often omitted.

We further explore how assumptions about potential effects of hospice use on mortality and timing of death within the 18-month observation window affect estimates of the effect of hospice use on Medicare spending. Our base model uses a nonhospice comparison group that matches hospice users on an extensive set of demographic, utilization, and diagnosis characteristics captured before the observation window, thereby making no assumption about any effects of hospice on mortality or timing of death. We estimate alternative model specifications that (1) match hospice users and nonusers based additionally on whether they die within the observation window and (2) further match hospice users and nonusers based on the number of days they are alive within the observation window. Finally, we conduct aggregate analyses at the zip code level to examine the effect of hospice use rates in the zip code on total per capita Medicare spending and mortality.

Key Findings

Across both our aggregate- and individual-level analyses, we find the following:

- At the national level, hospice use increased from 40 percent of Medicare decedents in 2007 to 54 percent in 2018.
- Over this same period, the share of Medicare spending occurring in the last year of life fell slightly, from 26.3 percent to 26.0 percent (including Part D spending). Prior analyses conducted using data from 2002 to 2012 found that the share of Medicare spending occurring in the last year of life was increasing, whereas we find the share of such spending has since leveled off as hospice use continued to increase. This suggests that increased hospice use was not leading to overall higher Medicare spending in the last year of life during the period we examine.
- Our individual-level analyses use zip-code-level distance to hospice as an instrument for hospice use with the idea that closer proximity to a hospice provider could increase the likelihood that a beneficiary uses hospice, but is not expected to affect Medicare spending except through hospice use after other observable factors are held fixed. The IV approach treats changes in distance to a hospice provider as exogenous shocks that create natural experiments, allowing us to estimate how Medicare spending changes for individuals who would not otherwise enroll in hospice but are induced to do so when a hospice provider becomes closer to them.
- Using two-stage least squares (2SLS) to estimate IV regression models of the effect of hospice entry, we find hospice entry reduces Medicare spending by $25,878 over the 18-month observation window (table ES.1, base model). The effect is statistically significant but imprecisely estimated with a standard error of $8,186; thus, implied confidence intervals are fairly wide. The 2SLS estimate is larger in magnitude than the least squares (LS) estimate of -$7,435, which does not account for potential endogeneity of hospice entry. The exogeneity of hospice entry is rejected in the base model (p-value = 0.022). The 2SLS and LS estimates could also differ because they are estimating effects of hospice for different groups of beneficiaries. The 2SLS result applies to so-called "compliers," that is, those beneficiaries who were induced into hospice use by the IV, that is, a hospice provider becoming closer to them. With treatment effects of hospice differing across types of patients, the effects of hospice entry on compliers will generally differ from the average treatment effect across all hospice users.

TABLE ES.1

Summary of Estimates of Effects of Hospice Entry on Total Medicare Spending Under Alternative Assumptions for the Effect of Hospice on Timing of Death

	Prospective model (base): Conditions on covariates known at the start of the index quarter	Alternative model 1: Conditions additionally on death in the observation window	Alternative model 2: Conditions additionally on days alive in the observation window
LS estimates	-$7,435***	-$15,333***	$4,059***
SE	($228)	($290)	($136)
2SLS estimates	-$25,878***	-$23,682***	-$6,435
SE	($8,186)	($7,627)	($6,433)
AR test (p-value)	0.009	0.027	0.505
F-statistic for instruments	17.2	24.6	15.2
Endogeneity test (p-value)	0.022	0.277	0.118
Overidentification test (p-value)	0.417	0.961	0.459
Dependent variable means			
Hospice	$32,995	$32,995	$32,995
Nonhospice, weighted	$40,034	$48,002	$29,228
N	18,924,700	18,924,700	18,924,697

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: AR = Anderson-Rubin; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending is measured over an 18-month observation window including the index quarter and is adjusted to 2018 dollars. Model includes covariates as described in the Data and Methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by geographic fixed-effect level, are in parentheses. Regression models estimate the effect of hospice entry on total Medicare spending using LS and 2SLS. The 2SLS models use distance to hospice as an instrument for hospice entry. For methodological details, including descriptions of the tests performed, see the Data and Methods section.
*** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

- Our analysis of these compliers indicates that they are more likely to be institutionalized, have Alzheimer's disease and related dementias (ADRD), and be older than 85 than the average individual in the overall weighted sample. The compliers are less likely than average to be non-Hispanic Black, have conditions involving five or more body systems, and have cancer. Thus, compliers of the distance IV have characteristics that are associated with longer hospice stays. Among hospice decedents in 2018, average lifetime length of stay was 151 days for beneficiaries whose principal diagnosis was a neurological condition, a category that includes ADRD, compared with 53 days for beneficiaries with cancer (MedPAC 2020). The 90th percentile of lifetime length of stay was 445 days for patients with neurological conditions, compared with 128 days for patients with cancer, and 253 days overall.

- Also using the base model, 2SLS estimates indicate that hospice entrants were alive for 213 fewer days in the 18-month observation window (approximately 548 days). The corresponding LS estimate shows a difference of 330 days (day estimates not shown in the table). The large difference estimated by LS may be expected. Even after matching on a large set of important demographic, diagnosis, and utilization characteristics measured before the observation window, we expect the hospice and comparison group cases to differ on unobserved factors that affect both hospice entry and timing of death, such as frailty and care preferences of the patient and family. Also, people who enter hospice after the start of our observation window should have conditions that would lead a physician to attest that the patient is expected to live six months or less. These include factors only observed after the start of the observation window that the prospective base model intentionally does not control for (because the decision to use hospice could affect these factors) and factors like lab results that we cannot observe in Medicare administrative data.
- That a large difference of 213 days alive remains in the 2SLS result is more surprising. The distance IV severs the mechanical link between hospice entry and unobserved individual-level factors that could affect both hospice entry and days alive in the observation window. The 2SLS estimate will still consistently estimate the effects of hospice entry on Medicare spending and days alive if the distance-to-hospice instrument is valid. Analyses we conducted increased our confidence in the instrument. We found no indication of an effect running in the direction opposite to the one expected (i.e., mortality in the current year did not predict distance the following year), and a placebo test showed no indication of distance-to-hospice affecting accidental deaths. However, the 2SLS result suggests a large effect of hospice on days alive of 213 fewer days compared with a mean of 473 days for matched nonhospice cases in the analysis sample. The estimate pertains to compliers with characteristics associated with long hospice stays and who may also be expected to survive for extended periods if they had not used hospice. The estimated effect of hospice on days alive is of similar size to that found by Gruber et al. (2025). In many other areas of health care, a reduction in survival time is unambiguously a poor outcome. In the case of hospice, terminally ill beneficiaries who elect hospice choose a model of end-of-life care focused on quality of life so an estimated reduction in survival time does not necessarily indicate a poor outcome if it reflects the beneficiary's own preference for hospice care over conventional care.
- Readers who consider a mortality effect of that magnitude implausible, however, may question the instrument's validity on that basis. Therefore, we examine alternative specifications that

impose assumptions about how hospice use affects mortality and timing of death. As a first alternative, we re-estimated the regression models that control for whether the beneficiary died within the 18-month observation window. In a second alternative, we further control for days alive in the observation window. These models effectively impose the assumptions that hospice use has no effect on survival or survival time, consistent with assumptions made in much of the hospice literature and with findings of a small number of studies that found hospice or palliative care has little or no effect on mortality (Connor et al. 2007; Pyenson et al. 2004; Temel et al. 2010).

- In the first alternative model (table ES.1, column 2), conditioning on having died within the observation window, 2SLS estimates a similar effect of hospice entry on Medicare spending as in the base model (-$23,682), and the LS effect is larger than in the base model (-$15,333). Exogeneity of hospice entry is not rejected in this model. In this alternative model, the effect of hospice entry on days alive in the observation window is 84 fewer days and statistically significant, and the mean number of days alive for nonhospice cases is 313 (not shown in table).
- In the second alternative model that fully conditions on days alive in follow-up (table ES.1, column 3), using 2SLS we find that hospice entry has no statistically significant effect on overall Medicare spending (point estimate = -$6,435), whereas the corresponding LS estimate is positive ($4,059) and statistically significant. Exogeneity of hospice in this model is not rejected. For both alternative model specifications, the reported estimates can be interpreted as causal effects of hospice only if the maintained assumptions hold; that is, hospice entry has no effect on death or timing of death, respectively, in the observation window.
- In exploratory analysis, we use an alternative "leave-one-out" IV method that uses the mean zip code hospice rate by time period (i.e., quarter) for each case, excluding each individual's own hospice use from the calculation. Preliminary estimates show effects that are broadly similar to those reported here for the main analyses, but with a much stronger instrument and effects that are more precisely estimated.
- In zip-code-level analyses, also using a distance-to-hospice instrument, 2SLS estimates show negative effects of hospice use on Medicare spending that are broadly similar in magnitude to our individual-level results that do not control for days alive. Hospice use significantly increases hospice spending and significantly reduces nonhospice spending by a substantially larger amount. The overall net effects are significantly different from zero at the 10 percent level for contemporaneous year spending and at the 1 percent level for spending the following year in

models that include zip code fixed effects. Estimated effects of hospice use on mortality are mixed and depend on model specification.

Discussion

This study seeks to estimate the effects of hospice use on Medicare spending overall. To this end, we used a variety of individual-level and aggregate approaches, each designed to address selection effects inherent in studying hospice use while including nearly all hospice users and Medicare spending for beneficiaries ages 66 and older. We also estimated the effects of hospice use on factors that contribute to total spending, spending per day, and days alive in the observation window, as well as whether an individual died in the observation window.

Across datasets and model specifications, we find that estimates of the effects of hospice entry on Medicare spending differ depending on whether the models make no assumptions about the effects of hospice use on mortality and survival time or impose an assumption. In prospective 2SLS models that do not control for time until death, we find significantly lower spending associated with hospice entry, primarily driven by earlier death among hospice users. In models that control for overall mortality but allow survival time to vary between hospice users and nonusers, we similarly find significantly lower Medicare spending, including both fewer days alive and lower spending per day. However, when we assume that hospice use had no effect on survival time by fully controlling for the number of days alive in the observation window, our estimated effects change substantially, showing no statistically significant effect of hospice use on Medicare spending.

Further research using methods for causal inference is needed to disentangle what drives lower or higher costs associated with hospice entry for specific subgroups. We find that estimates using distance to hospice as an instrumental variable estimate effects for a particular subgroup of compliers with characteristics associated with longer hospice stays. Estimates for such cases are valuable because they focus on the types of patients for which the literature has not converged on the direction of spending effects, unlike for cancer. However, the approach does not appear to be an applicable identification method to estimate effects for other types of hospice users. More analysis is needed to better understand the mechanics behind any effect of hospice on mortality and/or the reasons why distance-to-provider IV or other approaches might overestimate such effects. Qualitative work could also shed light on what may be plausible effects of hospice on days alive by condition groups, allowing for better modeling of the effects of hospice on Medicare spending. The value of the hospice benefit for patients

and their families, which this work did not examine, is of prime importance regardless of whether Medicare's hospice benefit on net reduces or increases program spending.

The Effects of Hospice on Medicare Spending

Under the Medicare hospice benefit, qualified beneficiaries with a life expectancy of six months or less can receive in-home or institutional end-of-life care, including palliative care, intensive comfort care, and other supports that are consistent with beneficiaries' preferences. To access the hospice benefit, a physician must certify that the beneficiary meets this life expectancy criterion, and beneficiaries agree to receive palliative care for their terminal illness and related conditions under the hospice benefit and forgo conventional care for their terminal illness provided outside of hospice (MedPAC 2026). In theory, exchanging conventional or life-extending care for supportive care at the end of life could lead beneficiaries to receive less intensive care, including less hospital care, and may therefore be expected to lower Medicare end-of-life spending and improve end-of-life experiences for patients and their families. On the other hand, Medicare pays hospices a daily fee for enrolled patients, which could increase Medicare costs, particularly for those with long hospice stays.

Medicare instituted the hospice benefit in 1982 to improve end-of-life care and with the expectation that more appropriate care would also generate lower program costs. At the time, the Congressional Budget Office estimated that the benefit would result in net lower Medicare spending.[3] The hospice benefit provides Medicare beneficiaries with important end-of-life care options, including pain relief, symptom management, and psychosocial supports (MedPAC 2026), and may provide significant value to individuals regardless of costs or effects on Medicare spending. This report focuses on the net effects of hospice use on total Medicare spending. The value of the hospice benefit to enrollees and their families is also important, but it is not covered here.

The literature is mixed on whether hospice reduces Medicare spending overall, and the differences in results across studies appear to be related both to the conditions studied and the methodological approach (Hogan 2015). Past studies have consistently found lower Medicare spending associated with hospice use for cancer, with stays of at least one month more likely to result in lower spending than shorter stays (Fiala et al. 2020; Hung, Hsu, and Wang 2020; Kalidindi et al. 2020; NORC 2023; Zuckerman, Stearns, and Sheingold 2016). These studies used a variety of methods over varying time periods, suggesting a robust, consistent association between hospice use and reduced Medicare spending for patients with cancer. In contrast, some studies find that hospice use is associated with higher Medicare spending for patients with Alzheimer's disease and related dementias (ADRD; Zuckerman, Stearns, and Sheingold 2016) and for those with very long hospice stays (Hogan 2015;

Hung, Hsu, and Wang 2020; Obermeyer et al. 2014). Aldridge et al. (2023), however, found hospice use to be associated with lower or similar Medicare spending for beneficiaries with ADRD, depending on subgroup. A recent quasi-experimental study found substantially lower Medicare spending associated with hospice use for Medicare beneficiaries with ADRD, driven by higher mortality for hospice users (Gruber et al. 2025). Much of the prior literature had implicitly or explicitly assumed hospice has no effect on mortality by, for example, limiting the sample to decedents and/or retrospectively comparing spending of hospice users to nonusers over a fixed time period before death (Hogan 2015; Hung, Hsu, and Wang 2020; Leibowitz, Tan, and Gildner 2020; NORC 2023; Zuckerman, Stearns, and Sheingold 2016).

In general, studies of the effects of hospice on Medicare spending suffer from a few common problems. First, developing a comparison group for hospice users is very difficult as hospice users are, by definition, seriously ill and can differ from nonusers in ways that are not fully captured in claims data, such as frailty, prognosis, and care preferences. Some studies attempt to correct for this matching problem by focusing only on decedents, matching hospice decedents with nonhospice patients who died over a similar period. However, these studies fail to account for the large subset of Medicare hospice users who are discharged alive (19 percent in 2024; CMS 2025), and they assume that hospice does not shorten or lengthen life relative to conventional care (Zuckerman, Stearns, and Sheingold 2016). Also, many studies of the effect of hospice on Medicare spending focus on narrow time windows, generally one year or less, and therefore do not account for the substantial share of hospice spending that occurs outside the last year of life (Hogan 2015).

A few studies have attempted to address these problems. Most recently, a 2025 study used a quasi-experimental approach to follow patients diagnosed with ADRD forward five years from the time of diagnosis to estimate the effect of for-profit hospice on Medicare spending, finding that the expansion of for-profit hospices reduces Medicare spending for the marginal hospice patient with ADRD through lower inpatient utilization and higher mortality (Gruber et al. 2025). The quasi-experimental approach uses distance to for-profit hospice as an instrument for hospice use. This is intended to avoid the bias that results from hospice users being sicker than the comparison group in ways that cannot be controlled in the model.

Other studies have attempted to address weaknesses in the literature by employing a population-level geographic approach rather than assessing the effects of hospice at the individual level. This sidesteps the need to match hospice users to nonusers at the individual level, a key source of bias in the literature. If hospice use, on average, results in lower Medicare spending, geographic areas with increasing rates of hospice use, all else equal, should see decreases in spending on end-of-life care

relative to other spending. However, a 2020 study found that, while there is significant geographic variation in both end-of-life spending and hospice use, there does not appear to be a substantial association between the two (Iyer et al. 2020). In addition, a 2015 study found that increases in hospice use within a Core Based Statistical Area (CBSA) were associated with higher end-of-life spending (Hogan 2015).

This study explores the overall effect of hospice use on Medicare spending at both the aggregate and individual levels. Prior literature suggests that estimated relationships between hospice use and Medicare spending vary by condition, length of stay, and study methods, making it difficult to infer an overall effect of the hospice benefit on the Medicare program. We therefore use multiple methods and levels of analysis to attempt to infer the overall direction and estimate the magnitude of the effects of hospice use on Medicare spending, focusing on approaches that help overcome long-standing issues with matching hospice users to nonusers based on characteristics observable in claims data. We examine national-level trends in hospice use and end-of-life spending from 2007 to 2018, extending trends reported by Hogan (2015).

To estimate the causal effects of new hospice entry on Medicare spending at the individual level, we then turn to quasi-experimental instrumental variable (IV) methods, using zip-code-level distance to hospice as an instrument for hospice use. Previous studies have used distance-to-provider instruments as a source of variation in access to care hypothesized to be as good as random for identifying the causal effects of health care service use, and Gruber et al. (2025) used distance to for-profit hospice as an instrument to estimate the effect of hospice on Medicare spending among beneficiaries diagnosed with ADRD. Given our focus on overall effects, we include nearly all hospice entrants ages 66 and older, unlike many studies that focus on Medicare beneficiaries with specific diagnoses or those who die within a specified time frame. These analyses follow beneficiaries for 15–18 months from initial hospice entry, partially addressing the criticism that many hospice studies do not include adequate follow-up (Hogan 2015). All analyses incorporate Part D spending, which prior studies often omitted. Because hospice may itself affect when beneficiaries die, we further explore how assumptions about that possibility affect estimates of the effects of hospice entry on Medicare spending.

Finally, we conduct aggregate analyses at the zip code level using data from 2007 to 2018. We estimate instrumental variable regression models of the effect of hospice use rates in the zip code on total per capita Medicare spending, per capita spending by component, and mortality. Here we assess whether an effect of hospice on spending is detectable in overall spending patterns using zip-code-level data representative of all traditional (FFS) Medicare enrollees ages 66 and older.

Data and Methods

This study aims to estimate the effects of hospice use on Medicare spending. As noted above, we take several approaches to addressing our research question given the conflicting results in the literature. We pursued three main approaches:

- First, we updated and extended prior work by Hogan (2015). This includes national-level analyses focused on trends in decedent-to-survivor spending overall from 2007 to 2018 and CBSA-level analyses on the association between hospice use rates and decedent-to-survivor spending over the same period. The CBSA-level analyses are provided in appendix A.
- Second, we conducted individual-level analyses exploring the relationship between hospice use and Medicare spending for new hospice users compared with nonusers. These analyses use an instrumental variables approach.
- Lastly, we conducted zip-code-level analyses exploring the relationship between hospice use and Medicare spending, also employing an instrumental variables approach.

For the national-, CBSA-, and individual-level analyses, we used the Geographic Variation Database (GVDB) from 2007 to 2019 (CMS 2021), with the specific quarters and years noted in each section below. The GVDB is based primarily on information from the Centers for Medicare & Medicaid Services' Chronic Conditions Data Warehouse. These files include claims records for Part A and Part B (institutional and noninstitutional) services and supplies, as well as enrollment data. Part D spending was obtained from the Part D event file for the same period.

Our sample for each year is restricted to fee-for-service (i.e., traditional Medicare) beneficiaries ages 65 and older at the start of the year for the national and CBSA-level analyses (ages 66 and older at the start of the year for individual-level analyses). We excluded persons who had any months of Part C (Medicare Advantage) during the year or the next year, or who had no months of Part A coverage during the year. We also excluded beneficiaries with addresses in US territories and those without state information. Social Security Administration location codes for those without state information indicate that beneficiaries without state information are overwhelmingly living at non-US addresses. Finally, we excluded a relatively small number of persons who did not die in the reference year and were not present in the following year's file, meaning their death date was unknown.

Additional exclusions are described below for the individual-level analyses. In some analyses, we restricted our sample to specific years or quarters to make processing more feasible.

Changes in Decedent-to-Survivor Spending Ratios

The first analysis compares spending within the last year of life (decedents) to spending not within the last year of life (survivors) and examines how decedent spending relative to survivor spending changed during a period in which hospice use increased. Using the ratio, rather than decedent spending alone, normalizes for overall trends in Medicare spending unrelated to hospice, so that movements in the ratio reflect changes in the intensity of end-of-life spending rather than growth in Medicare spending generally. This analysis updates one previously reported by Hogan (2015), who found that, over a period when hospice use increased from 26 percent of decedents in 2002 to 47 percent in 2012, the ratio of decedent-to-survivor spending also increased. Hogan interpreted this finding as evidence that hospice was not reducing aggregate end-of-life spending as some reported estimates of lower hospice-related spending would predict. We extend that analysis through 2018 to examine whether the pattern of increasing decedent-to-survivor spending continued to hold.

The spending measures for this analysis are based on Medicare spending per person-year. For persons with any measurement-year spending that is within the last year of life ("decedents"), about half of their measurement-year spending is within the last year of life, on average. For persons with any measurement-year spending that is not within the last year of life ("survivors"), a small share of their measurement-year spending is within the last year of life, on average. Use of spending per person-year allows us to compare the intensity of spending within and not within the last year of life, adjusting for this difference. Within a given calendar year, beneficiaries who die in the following calendar year contribute to both survivor and decedent spending. Calendar year spending not within the last year of life is attributed to the survivor group, and calendar year spending within the last year of life is attributed to the decedent group.

For each calendar year from 2007 to 2018, we developed annual estimates of spending for decedents relative to survivors, allowing us to observe trends over a period with growth in hospice use.

Defining the Sample

For this national-level analysis, our sample included all beneficiaries in the GVDB beneficiary file, subject to the exclusions above. In 2010, for example, the full file included approximately 50 million beneficiaries. Of these, 0.8 percent did not have Part A coverage in 2010, 27.4 percent were dropped because of Part C enrollment in 2010 or 2011, less than 0.1 percent neither died nor had a record the following year, and 1.1 percent were either residing in a territory or missing a state code. The resulting eligible sample was 35.3 million beneficiaries in 2010. We excluded an additional 8.3 million beneficiaries who were younger than age 65, resulting in an analysis sample of 27 million beneficiaries.

Spending Measures

To calculate decedent and survivor spending per person-year, we accumulate per-person-year spending from claims, separated by whether an individual claim's end date falls within the last year of life. Spending within the last year of life is referred to as decedent spending, and spending not within the last year of life is referred to as survivor spending. The spending is based on Part A and Part B (institutional and noninstitutional) claims files and the Part D events file. Part D events were not part of previous population-level studies.

Following Hogan, we calculate separate totals for decedents and survivors for (1) direct program spending and (2) all spending, that is, the sum of payments received by providers, including direct program spending, copayments, deductibles, and Part D patient liabilities. For Part D, the direct program spending is the sum of the covered Part D plan payments and the low-income subsidy; the Part D patient liabilities include the patient payment amount plus the amounts below the true out-of-pocket maximum.[4] Throughout this report, we focus on direct program spending.

The calculation of spending per person-year for each calendar year is as follows:

- Sum the calendar year spending for each beneficiary separately by whether within the last year of life (decedents) or not within the last year of life (survivors).
- Exclude hospice spending from the total for survivor beneficiaries.
- Calculate average spending per person-year for decedents and for survivors.
    - » For decedents: Divide average spending for those beneficiaries with positive decedent spending by their average share of the calendar year within the last year of life.
    - » For survivors: Divide average spending for those beneficiaries with positive survivor spending by their average share of the calendar year not within the last year of life.

Like Hogan, we weight our decedent and survivor estimates to hold the age/sex distribution fixed for each year to avoid having the spending patterns reflect changes in the age/sex distribution over time. We weight the estimates within 12 age/sex groups by the distribution of days across those groups found in the 2002 Master Beneficiary Summary File (MBSF). Spending is reported in 2018 dollars, using the Consumer Price Index (CPI) for inflation adjustment.

Individual-Level Analyses Estimating the Effect of Hospice Use on Medicare Spending

Using individual-level data on Medicare beneficiaries, we conducted regression analyses to estimate the effects of a beneficiary entering hospice on individual-level Medicare spending. Specifically, we estimate the effect of first entry into hospice within a calendar quarter on spending in the 18 months following the start of that quarter. We used propensity score weighting to obtain a comparison sample that matches hospice entrants in their distributions of demographic characteristics and measures of medical history captured over the past two quarters and up to one year. We first report least squares (LS) estimates of effects of hospice entry relative to the propensity-weighted comparison group. Although these estimates control for a detailed set of observable factors, they do not account for unobserved factors that may be correlated with both hospice use and Medicare spending. To produce estimates that may better estimate causal effects of hospice in the presence of such unobserved factors, we also estimate hospice effects using two-stage least squares (2SLS) with distance to the nearest hospice as an instrument for hospice entry.

Defining the Sample

We estimate the effects of hospice use for Medicare beneficiaries who entered hospice for the first time in the first and third quarters of 2008, 2013, and 2018. We call these six quarters "index quarters" and the year when they occur the "index year." The follow-up window (i.e., observation window) for each index quarter is shown in table 1.

TABLE 1

Index Quarters, Index Years, and Follow-up Windows

Index year	Index quarter	Follow-up window
2008	January–March	January 2008–June 2009
2008	July–September	July 2008–December 2009
2013	January–March	January 2013–June 2014
2013	July–September	July 2013–December 2014
2018	January–March	January 2018–June 2019
2018	July–September	July 2018–December 2019

Source: Urban Institute analysis and definitions used for construction of analytic file.

Because our goal was to estimate overall hospice effects, we generally sought to retain the broadest possible sample of hospice enrollees, though with a few exceptions. We restricted the analysis to beneficiaries ages 66 and older throughout the index year who had Part A and Part B coverage and no Part C coverage in the index year, the previous year, and the following year. We made these exclusions

to narrow the sample to beneficiaries with complete traditional Medicare spending and risk adjustment data.[5] We further excluded beneficiaries with no Medicare claims in the two quarters before the index quarters to narrow the comparison group to beneficiaries with adequate diagnosis and utilization data to assess comparability with those entering hospice. We also excluded beneficiaries who had organ transplants or received dialysis treatment in the index quarter or the two preceding quarters because risk scores for this small population are defined differently from those of other beneficiaries. These restrictions resulted in our excluding 16 percent of new hospice cases in our base file (with Part A in the index year and no Part C in the index or following year) for the first quarter of 2008:

- 8.1 percent who were younger than age 66 at the end of the year;
- 3.4 percent without qualifying enrollment in the previous year;
- 1.1 percent without Part B in at least one of the three years;
- 1.6 percent without claims in the two preceding quarters;
- 1.7 percent with either dialysis or a recent transplant in the index quarter or preceding two quarters; and
- 0.2 percent without consistent zip code assignment.[6]

Similar shares of new hospice users were excluded for other index quarters.

All other Medicare beneficiaries who met the criteria above but were not hospice entrants were potentially included as comparison cases with one exception: we excluded any potential comparison cases with a first hospice entry in the quarter following the index quarter to avoid having the comparison group include persons receiving hospice treatment over a very similar follow-up period. To reduce file-processing time and enable high-dimensional statistical models to run within server memory constraints, we limited the comparison group for each index quarter to a random 15 percent sample of the beneficiary population. This reduction in sample size made it feasible to process files and run computationally intensive estimation procedures for analyses that pool data across the six index quarters.

Constructing a Propensity-Score-Weighted Comparison Group for Hospice Entrants

Our propensity-score weighting approach aimed to create a sample of Medicare beneficiaries who did not enter hospice with demographic characteristics, service use, diagnoses, baseline spending patterns, and other characteristics as close as possible to those of hospice entrants. To do this, we use a

propensity-score weighting approach such that, when weighted, hospice users and nonusers are matched on a detailed set of observable characteristics (Hirano, Imbens, and Ridder 2003; Rosenbaum and Rubin 1985). For each index quarter, we estimated propensity scores calculated as predicted probabilities from a logit model predicting whether a beneficiary was a new entrant to hospice in the quarter. Table 2 shows the predictors used in the propensity score models. All predictors were defined using data and measures determined before the start of the index quarter using a lookback period of up to one year. Measures based on diagnoses (e.g., risk scores, Hierarchical Condition Categories [HCCs], cancer indicators) were chosen to ensure comparable measures that could be constructed using both International Classification of Diseases (ICD) codes: ICD-9 (before October 2015) and ICD-10 (after October 2015). The references to distribution in the definitions of spending and risk scores refer to the inclusion of seven indicators for <25th percentile; 25th–50th percentile; 50th–75th percentile; 75th–90th percentile; 90th–95th percentile; 95th–99th percentile; 99th and above. To adjust for geographic differences at a reasonably detailed level without overburdening the logit models with finer levels, we included CBSA-group fixed effects as predictors along with the other person-level measures.

TABLE 2

Predictors Used in Propensity Score Models of Hospice Entry

Measure	Categories	Source
Age	66–69, 70–74, 75–79, 80–84, 85–89, 90–94, 95–99, 100+	Beneficiary file, age at end of index year
Female	Female	Beneficiary file
Race/ethnicity	Unknown, non-Hispanic white, non-Hispanic Black, other, AAPI, Hispanic, AIAN	Beneficiary file using Research Triangle Institute race categorization
Risk score	Distribution among risk scores (<p25, p25–p50, p50–p75, p75–p90, p90–p95, p95–p99, ≥p99) for hospice entrants	Based on inpatient, outpatient, and physician claims in the past year
Risk score type	Community/full dual, community/nondual, community/partial dual, institutional, new enrollee	Medicare enrollment and other data files
Original reason for entitlement	Disability	Medicare enrollment files
Severity of illness (highest observed)	No acute claims, and severity indicators from 1 (least severe) to 4 (most severe)	APR-DRG severity indicators from acute stays over the three previous quarters
Risk of mortality (highest observed)	No acute claims, and risk indicators from 1 (least severe) to 4 (most severe)	APR-DRG risk of mortality indicators from acute stays over the three previous quarters

Measure	Categories	Source
Number of emergency department visits in prior year	None, 1–2, 3–7, 8+	ED visits (either Part A or B) in the prior year, from MBSF_CU (Master Beneficiary Summary File—Cost and Use) for index quarters in the first calendar quarter of the year and revenue center files for Parts A and B (institutional) for index quarters in the third calendar quarter of the year
Number of ICU/CCU days	None, 1–2, 3–5, 6–10, 11–20, 21+	Part A revenue claims
Part A payments (prior quarter)	None and indicators for distribution among the positive payments (<p25, p25–p50, p50–p75, p75–p90, p90–p95, p95–p99, ≥p99) for hospice entrants	Part A claims in the quarter before the index quarter
Part A payments (two quarters prior)	None and indicators for distribution among the positive payments (<p25, p25–p50, p50–p75, p75–p90, p90–p95, p95–p99, ≥p99) for hospice entrants	Part A claims two quarters before the index quarter
Part B institutional payments	None and indicators for distribution among the positive payments (<p25, p25–p50, p50–p75, p75–p90, p90–p95, p95–p99, ≥p99) for hospice entrants	Part B institutional claims, summed over two quarters before the index quarter
Part B noninstitutional payments	None and indicators for distribution among the positive payments (<p25, p25–p50, p50–p75, p75–p90, p90–p95, p95–p99, ≥p99) for hospice entrants	Part B noninstitutional claims, summed over two quarters before the index quarter
Part D provider payments	None and indicators for distribution among the positive payments (<p25, p25–p50, p50–p75, p75–p90, p90–p95, p95–p99, ≥p99) for hospice entrants	Part D event records, summed over two quarters before the index quarter
Alzheimer's	MBSF indicates Alzheimer's	Master Beneficiary Summary File chronic condition indicators, based on most recent six-month report from MBSF; includes those identified whether or not they consistently met the FFS criteria.
Alzheimer's and dementia	MBSF indicates Alzheimer's or dementia	Master Beneficiary Summary File chronic condition indicators, based on most recent six-month report from MBSF; includes those identified whether or not they consistently met the FFS criteria.

Measure	Categories	Source
Cancer, grouped CCS and CCSR categories	Indicators for secondary malignancy, pancreatic, nervous system, or brain, malignant with an unknown site, esophagus, stomach, liver/bile duct, ovary, and other	Based on hospitalization, outpatient, and physician claims diagnoses over previous four quarters. Groups are based on CCS for 2008 and 2013 and CCSR for 2018, because of change from ICD-9 to ICD-10. Category assignments were made to obtain reasonably consistent definitions over time. "Other" includes categories that do not fall neatly into a given system, those that cannot be consistently assigned across years, and those that had relatively weak effects on hospice entry. Benign cancers were excluded. Note: The "other" category accounts for a substantial share of hospice cases: in 2008 Q1, the large subgroups were neoplasms of unspecified nature (24%), maintenance chemotherapy (26%), prostate cancers (18%), skin cancers (19%), and breast cancers (14%).
Nonhierarchical HCC groupings	21 broad groups: alcohol or drug disease, cancer, cardiac and vascular, complications of device or graft, eye disorders, GI and liver, hematologic and immunologic diseases, HIV/AIDS, mental illness, metabolic endocrine, neuro excluding stroke, obesity, orthopedic, renal, respirator dependence, respiratory septicemia and other systemic infections, skin disorders, stroke, transplant	Nonhierarchical HCC groupings based on inpatient, outpatient, physician claims from four quarters prior to index quarter (PACE/ESRD groupings)
Five or more body systems	Broad HCC categories cover at least five body systems	Nonhierarchical HCC groupings based on inpatient, outpatient, physician claims from four quarters prior to index quarter (PACE/ESRD groupings)
Index quarter	Indicator of month and year of index quarter	Geographic Variation Database
CBSA group	Indicators of 430 CBSA groups	Core Based Statistical Areas (CBSAs) for metropolitan areas, grouping other areas by state

Source: Urban Institute analysis of the Geographic Variation Database, 2007–09, 2012–14, and 2017–19.
Notes: AAPI = Asian American and Pacific Islander; AIAN = American Indian and Alaska Native; APR-DRG = all patient refined diagnosis related group; CBSA = Core Based Statistical Area; CCS = Clinical Classifications Software; CCSR = Clinical Classifications Software Refined; CCU = coronary care unit; ED = emergency department; ESRD = end-stage renal disease; FFS = fee-for-service; GI = gastrointestinal; HCC = Hierarchical Condition Category; ICU = intensive care unit; MBSF = Master Beneficiary Summary File; PACE = Program of All-Inclusive Care for the Elderly; Q1 = first quarter. The references to distribution in the definitions of risk scores and spending refer to the inclusion of indicators for <25th percentile; 25th–50th percentile; 50th–75th percentile; 75th–90th percentile; 90th–95th percentile; 95th–99th percentile; 99th and above.

We then used the predicted probability of hospice entry for a beneficiary, p_i, to generate a weight for comparison cases. Hospice cases were given a weight of one, representing themselves. Nonhospice cases were first given a weight of $w_i = p_i / (1-p_i)$. This is the standard inverse probability weighting approach formulated to estimate the average effect of treatment on the treated (i.e., hospice) cases (Hirano and Imbens 2001). We found, however, that with a weight calibration step we could somewhat improve the balance of covariates in treated and (weighted) untreated samples (Gutman, Karavani, and Shimoni 2024). The calibration also meaningfully reduced the variability of the weights for the nonhospice cases which, as expected, improved the precision of regression model estimates. We produced calibrated weights as follows. First, using quantiles of the estimated propensity score, we grouped observations into 50 bins, increasing in propensity score. Then, within each bin, we estimated the empirical share of cases that are hospice entrants. We then created weights using this empirical estimate of probability within the bin rather than the propensity score. This is the calibration aspect of the exercise. The last step is producing the weights as described above but using the empirical probability of hospice in each case's bin as a calibrated propensity rather than the original propensity score.

With the calibrated weights applied, the distributions of beneficiary characteristics used as predictors are very similar in the hospice group and comparison group. In table 3, we show how selected characteristics of hospice and nonhospice cases compare before and after weighting. The first column shows the distribution of age, risk scores, and a subset of health conditions for hospice cases; the second column shows the unweighted distribution for comparison cases; and the third column shows the weighted distribution for comparison cases. Pooled over all six quarters, there are 901,837 hospice cases and 18,022,863 nonhospice cases (total N = 18,924,700). The weighted sample size (N) of the nonhospice cases equals the N of the hospice cases (all with a weight of 1). With weights applied, the distributions of characteristics for nonhospice cases are quite similar to those for hospice cases. For example, before weighting, 10.7 percent of nonhospice cases are diagnosed with Alzheimer's or dementia. After weighting, the share of nonhospice comparison cases with Alzheimer's or dementia is 52.3 percent and 49.7 percent for hospice cases. Appendix table B.1 shows these distributions for all individual-level predictors used in the propensity model of hospice use. The weights lead to similarly close covariate balance for other characteristics as well. Nearly all standardized mean differences (SMDs) between hospice users and nonusers are smaller than the commonly employed threshold of 10 percent (Austin 2009). Exceptions are risk scores in the 25th to 50th percentile (SMD = 0.126) and secondary malignancy (SMD = -0.144). As an additional measure of control, we include these and other characteristics as controls in the regression analyses.

TABLE 3

Distributions of Selected Beneficiary Characteristics Before and After Reweighting Comparison Group Cases

	Unweighted percent among nonhospice cases	Weighted percent among nonhospice comparison cases	Percent among hospice entrants (unweighted and weighted)
Age at end of index year			
66–69	23.1	5.1	6.7
70–74	25.7	9.3	11.0
75–79	20.5	14.3	15.0
80–84	15.5	21.0	20.2
85–89	9.9	24.8	22.9
90–94	4.2	18.0	16.9
95–99	1.0	6.4	6.3
100+	0.1	1.0	1.0
Risk score percentile			
≤25th percentile	75.3	22.3	25.0
25th–50th percentile	15.3	29.0	25.0
50th–75th percentile	6.8	27.5	25.0
75th–90th percentile	1.9	13.8	15.0
90th–95th percentile	0.4	3.9	5.0
95th–99th percentile	0.3	2.8	4.0
>99th percentile	0.1	0.7	1.0
Alzheimer's or dementia	10.7	52.3	49.7
Cancer			
Secondary malignancy	1.1	9.3	12.5
Respiratory	0.9	5.7	7.2
Pancreatic	0.1	0.9	1.5
Nervous system or brain	0.1	0.7	1.0
Malignant, unknown site	0.3	2.2	3.3
Liver or bile duct	0.1	0.8	1.2
Esophagus	0.1	0.5	0.7
Stomach	0.1	0.6	0.8
Ovary	0.2	0.8	0.9
Other	26.2	35.3	37.1
N	18,022,863	18,022,863	901,837
Weighted N	—	901,837	901,837

Source: Urban Institute analysis of the Geographic Variation Database, 2007–09, 2012–14, and 2017–19.
Note: See appendix table B.1 for a complete comparison of beneficiary characteristics before and after reweighting.

Propensity-Score-Weighted Models of the Effect of Hospice Entry

We estimated least squares regression models of the effect of hospice entry on Medicare spending using our sample of hospice entrants and the weighted comparison group. The key outcomes for this analysis, measured over the index quarter and subsequent five quarters, are as follows:

- Medicare spending
- Spending per day alive
- Days alive in the 18-month observation window

- Whether died
- Spending by source and setting
  - » Part A Medicare spending
    - Acute care
    - Other inpatient
    - Post-acute care (PAC)
      - Home health
      - Skilled nursing facility (SNF)
      - Long-term care facility
      - Inpatient rehabilitation facility (IRF)
    - Hospice
  - » Part B Medicare spending
    - Part B institutional
      - Outpatient
    - Part B noninstitutional
      - Evaluation and management
  - » Part D Medicare spending

Data for hospice entrants and comparison beneficiaries from each of the six index quarters were pooled, together with their propensity weights. We estimated weighted regression models of each outcome on whether a beneficiary first entered hospice in the index quarter, controlling for age, sex, risk score percentile group, broad condition category (cancer, ADRD/dementia [no cancer], other condition), secondary malignancy, index quarter, and geographic area. Relatively few regression controls were used in this model because the propensity weights already balanced observable differences between hospice and comparison beneficiaries. The controls we retained account for any minor imbalances that remain. Including the regression controls may also increase the precision of the estimated effects of hospice to the degree that they reduce the variance of the error terms. The coefficient on the hospice entry indicator measures the average difference in outcome for hospice entrants in comparison with the weighted nonhospice cases.

2SLS Estimates

To better estimate the causal effects of hospice entry on the outcomes of interest, we used 2SLS, with distance to hospice as an instrument for hospice entry. Because we expect that hospice entry and Medicare spending may both be correlated with variables we cannot measure, like care preferences, we instrumented for hospice entry with a variable we expect is related to hospice entry but affects our outcomes only through its effect on hospice entry—distance to hospice. 2SLS treats variation in distance to hospice over time within a geographic area as though, after controlling for other factors, it were

randomly assigned, isolating the portion of hospice entry driven by that variation and allowing us to better observe the causal effect of hospice entry on Medicare spending.

For the 2SLS results to be valid, the instrument must fulfill two primary conditions: (1) it must be related to hospice entry, and (2) it must affect the outcome variables only through the treatment variable (i.e., it must be exogenous). The first condition is tested in the first stage of the model. If the instrumental variable is weak, it is not closely related enough to the potentially endogenous hospice entry measure to produce reliable results. A first-stage F-statistic below about 10 is the conventional signal of a weak instrument, but that threshold is a rule of thumb rather than a formal test. A weak first stage leads to an imprecise 2SLS estimate of the hospice effect. But it also tends to bias 2SLS toward the least squares estimate, pulling the result back toward the association and away from the true causal effect (Andrews, Stock, and Sun 2019). To help identify cases where lack of instrument strength could be driving our 2SLS results, we report weak-instrument-robust test statistics alongside conventional significance levels, as described below. The stronger the relationship between the instrument and the variable of interest (hospice entry), all else equal, the more reliable IV estimates become.

The second condition cannot be tested directly. Examining how the instrument relates to other measures can offer suggestive evidence, and where we use more than one instrument, the overidentifying restrictions can be tested, as described below, but the exclusion restriction itself remains a maintained assumption. A concern about the distance instrument is that distance could pick up uncontrolled differences in patient characteristics that affect spending separate from hospice, meaning the instrument may not be exogenous. For example, hospices may choose to locate in areas with high or increasing mortality. If current year's morbidity predicted next year's distance to hospice, it would cause concern. In analyses at the zip code level, we found no relationship between next year's distance to hospice and current year's mortality. Further, as a placebo test, with county-level data on accidental mortality from the Centers for Disease Control and Prevention, we found no relationship between accidental mortality and distance to hospice. We further assessed the relationship between distance to hospice and overall death rates at the county level and found no relationship. Overall, these findings increase our confidence that the distance instrument is valid.

If the conditions for 2SLS are met, it provides consistent estimates that can be interpreted as a causal effect of hospice entry on Medicare spending, whereas ordinary least squares is biased whenever hospice entry is correlated with unmeasured determinants of spending. However, while estimates from the least squares model provide an overall average difference in Medicare spending between hospice entrants and the comparison group holding observed control variables fixed, estimates from 2SLS

provide a local average treatment effect (LATE): the average effect of hospice entry on Medicare spending for those who were induced into hospice by a reduction in distance to hospice. We refer to these beneficiaries as "compliers" in our analysis. Therefore, our 2SLS results should be interpreted as the effects of hospice for compliers whose hospice use is sensitive to changes in distance, which will generally differ from the average treatment effect (ATE) of hospice on all hospice users if treatment effects are heterogeneous. We describe the characteristics of this complier group below.

The 2SLS model setup involves two equations. In the first-stage equation, hospice is modeled as a function of distance to hospice, regression controls, and fixed effects for index quarter and geographic area; in the second-stage equation, spending or other outcomes over the observation window are modeled as a function of hospice (instrumented with distance), the same regression controls as in the first-stage model, and fixed effects for index quarter and geographic area. Both equations are weighted using the propensity score-based weights.[7]

The 2SLS coefficient on hospice use provides an estimate of the causal effect of hospice use on spending. That is, the 2SLS estimate removes the portion of the association between hospice use and spending that results from unmeasured factors that are related to both hospice use and Medicare spending. Such factors include elements of demand such as patient preferences and knowledge of their specific condition. To test the endogeneity of hospice, we report the p-values from Durbin-Wu-Hausman endogeneity tests. A small p-value indicates rejection of the exogeneity of hospice; that is, it implies that hospice is endogenous and 2SLS is preferred over the least squares estimate.

We examined the profile of how distance to hospice is associated with hospice use and considered both a linear effect of distance (capped at 50 miles) and indicators by distance (5–10 miles, 10–20, 20–30, 30–40, and greater than 40). The linear model indicated a decreasing overall relationship, but the indicator variables suggested the pattern was not linear. At distance levels under 20 miles, the model with indicators showed a nonmonotonic pattern, with some closer distances showing a lower rate of hospice use than some farther distances. Because the idea of the distance instrument is that hospice use is expected to fall with distance, and the data show it does after a point, we do not want to use the increasing portion of the curve at nearer distances to help identify the hospice effect. Therefore, we forgo using predictive variation in that part of the curve so that the relationship we do capture is a declining pattern with distance. In the reported models, we included three indicators for distance (relative to a base category of 0–20 miles) for which hospice use declines monotonically with distance: 20–30 miles, 30–40 miles, and greater than 40 miles.

With distance to hospice expressed with multiple indicator variables, we can conduct overidentification tests.[8] With multiple instruments—in the present case through two additional distance indicators—the hospice effect is overidentified, meaning additional identifying information is being used beyond the minimum necessary for identification. The null hypothesis is that all the instruments satisfy the exclusion restriction, and rejection indicates that at least one violates it (that the instrument is uncorrelated with the second-stage error term). It could also indicate, however, that treatment effects are heterogeneous and the different instruments are identifying different LATEs for different complier populations, and the estimated coefficient represents an averaging over these effects (Angrist and Pischke 2009). Thus, the test does not determine the validity of the instruments either way. The just-identifying restriction is maintained as valid, and the test can reject because the instruments are estimating different LATEs for different subgroups. Rejection is informative in that it can suggest heterogeneous effects or invalidity of instruments in a well-specified model when true effects are constant across groups.

The strength of the instrument varied, depending on the level of geographic control we used. Gruber et al. (2025) controlled for zip code. However, we found little relationship between distance to hospice and entry into hospice when we controlled for zip code groups (individual five-digit zip codes grouped so that areas that split apart or merged over the study years are defined consistently over time; we refer to these simply as zip codes hereafter). We also found relatively weak relationships when we controlled for county or health care service area. However, with controls for three-digit zip code or CBSA group, our first-stage results met the usual standard for avoiding weak instruments (a first-stage F-statistic of at least 10). We primarily rely on the estimates using controls for three-digit zip codes in our results because they better describe contiguous geographic areas and provide finer control for potentially omitted factors than the CBSA groups.

The weak relationship between distance to hospice and hospice entry, controlling for zip code, appears to be partly driven by a lack of variation in distance to hospice over time at the zip code level over our study period. By coarsening our controls to control for larger geographic areas, we are better able to exploit variation in distance to hospice across zip codes within a period as well as over time. However, using broader geographic areas also reduces our ability to control for the effect of permanent local area factors (such as presence of other health facilities, etc.) that may be correlated with hospice use, Medicare fee-for-service spending, and distance to hospice.

To interpret the 2SLS estimates and their implications, it is important to understand to whom the results apply. To this end, we assessed the characteristics of the complier population for our 2SLS models to identify the types of beneficiaries driving our results.

Under the standard instrumental variable assumptions of independence and monotonicity (Angrist, Imbens, and Rubin 1996), the complier share of the population can be calculated directly as the difference in the probability of hospice use when distance is near and when distance is far: Pr(Complier) = Pr(Hospice | Near) − Pr(Hospice | Far). Although the population share of the complier group can be estimated, complier status cannot be observed at the individual level. Mean characteristics of the complier group can nonetheless be estimated.[9] Because this calculation requires a binary instrument, we used an indicator for whether the nearest hospice is 35 miles or more away. We chose 35 miles as the cutoff because it emphasizes the longer distances most strongly related to hospice use and provides an adequate sample for complier estimation.

Finally, to assess how different assumptions about the mortality effects of hospice affect our estimates, we re-estimated 2SLS models using weights from two alternative specifications of the propensity score model that: (1) include death during the follow-up period in the propensity score model; and (2) include days until death in the propensity score model.

Subgroup Analyses

To examine results by subgroup, we followed the overall approach within mutually exclusive subgroups: cancer, Alzheimer's/dementia without cancer, and the balance of cases (i.e., other conditions). We measured cancer from claims-based diagnoses using the Clinical Classifications Software Refined measures. We used ADRD indicators from the MBSF. The propensity score-based weights were created by index quarter, limited to beneficiaries in each group. These results are reported in appendix C.

Zip-Code-Year-Level Analysis

In this analysis, we estimate the effect of hospice use on Medicare spending at the area level, relating spending per enrollee in a zip code and year to the share of enrollees in that zip code and year who used hospice, and using distance to the nearest hospice as an instrument for hospice use. Using data from 2007 to 2018, we examine whether an effect of hospice on spending is detectable in overall spending patterns using zip-code-level data representative of all traditional (FFS) Medicare enrollees ages 66 and older. The zip-code-year file was constructed by Acumen as specified by MedPAC staff. Tabulated data by zip code and year were compiled from the MBSF Base, MBSF Cost and Use, and Geographic

Variation Database files for Medicare enrollees and linked with census zip code data. The data are for Medicare beneficiaries with Part A in a given year who were alive for some portion or all of the year. The data include summary information for demographic characteristics, types of Medicare and other coverage, Medicare spending by type of spending, and selected health conditions.

Distance to hospice, as described above, was matched to each zip code for each year. Because zip codes can change, we grouped together some individual five-digit zip codes that split apart or merged over the study years to construct zip code groups that are consistently defined with the least amount of grouping necessary to achieve consistency. We use these zip code groups in the findings section. For expositional ease, we refer to these as (five-digit) zip codes. Summary statistics for the constructed variables used in the analysis are shown in table 15 in the findings section.

The zip-code-level data allow us to estimate IV models using distance to hospice as an instrument, complementing the individual-level analysis. Like the individual-level analysis, the zip-code-level analysis uses variation in distance to hospice measured at the three- and five-digit zip code level, the latter being the most granular level at which the instrument is defined, to identify effects of hospice use on Medicare spending. The zip-code-level analysis uses data averaged over individuals in each five-digit zip code and year, with zip-code-level covariates as controls rather than a matched comparison group. This analysis complements the individual-level analysis in three ways. First, aggregation to the zip-code-year level made it feasible to use data for all sample individuals and all years from 2007 to 2018 and to estimate a broad range of model specifications. Second, the unit of observation is different, and the sample represents the broad population of enrollees rather than hospice entrants, so the effects measure how changes in area-level hospice use are associated with average Medicare spending across all fee-for-service beneficiaries in a zip code, rather than the effect of hospice among individuals who enter hospice. Third, the 2SLS findings estimate a LATE for complier geographies rather than complier individuals, thus capturing spending effects in places where hospice use is most responsive to changes in distance to hospice, rather than among individual beneficiaries whose hospice decisions are shifted by proximity to a provider.

We estimate the effect of hospice use on total Medicare program spending per capita, excluding beneficiary cost sharing. We also estimate analogous regression models on the following components of per capita spending:

- Hospice
- Inpatient
- Outpatient hospital

- Part B, excluding outpatient hospital
- Skilled nursing facility
- Home health
- Part D

The regression models specify expected medical spending per enrollee in each zip code and year as a function of the share of enrollees using hospice in the zip code and year, zip-code-year demographic composition, shares of different coverage types, shares for selected health conditions, year fixed effects, and zip code fixed effects, with separate specifications using five- and three-digit zip code fixed effects. We limit the sample to zip codes with more than 25 enrollees in each year, resulting in a strongly balanced panel of zip-code-year observations.

Because we are interested in national average effects, we weight the regressions by the number of enrollees in each zip code averaged over the sample years.[10] We use the average to reduce year-to-year random variation in zip code weights.

We estimate effects of hospice use rates on measures of spending per capita using linear regression and IV models using 2SLS. As in the individual-level analysis, we estimate IV models to allow for the possibility that hospice use is endogenous to per capita spending even after controlling for zip code fixed effects and other covariates. Hospice use could be endogenous if, for example, from one year to the next, hospice use in a zip code increased and those users were more costly in unobserved ways that increased total spending per capita in the zip code. In this case, we would expect to estimate a positive LS effect of the hospice use rate on spending even if there is no true effect of hospice. IV estimates with the distance instrument replace that assumption with a different identifying assumption: that within-zip-code changes in distance to hospice shift hospice use but are otherwise uncorrelated with omitted determinants of within-zip changes in spending per capita.

As in the individual-level analysis described above, the zip-code-year IV estimates can be interpreted as the effect of hospice for a marginal group of patients whose hospice use is responsive to the instrument, that is, compliers. The key difference is that instead of observing individual enrollment decisions, we observe average hospice use within each zip code and year. This means the hospice effect estimates from the zip-code estimates are best interpreted as the effect of shifting hospice penetration within geographic areas, rather than the effect for a particular patient subgroup. The IV estimate therefore reflects how changes in area-level hospice enrollment translate into changes in area-level Medicare spending, weighted toward zip codes and periods where there is a change in distance and where enrollment is most responsive to changes in distance.

In addition, we estimate a variant of the IV model that keeps predicted hospice use rates within the range that is possible for a share, between 0 and 1, at the cost of stronger assumptions about how the first stage is specified. This approach uses a nonlinear first stage that constrains predicted hospice use rates to the [0, 1] range. The first stage is specified as a fractional logit (or fractional response model) (Papke and Wooldridge 1996) with hospice use share as the dependent variable and distance-to-hospice categories as the instrumental variables along with covariates. The second stage of the IV model with the fractional logit first stage is estimated using a control function approach (Wooldridge 2015). We estimate a linear regression model with spending per capita as the dependent variable, hospice use as the potentially endogenous explanatory variable, the generalized residual from the first-stage model as the control function, and covariates. The generalized residual is intended to absorb correlation between unobserved determinants of hospice use and unobserved determinants of the outcome measure. This offers a potential additional channel through which the control-function estimate could improve on linear 2SLS (Wooldridge 2015), beyond respecting the range of the hospice use rate alone. This comes at the cost of additional identifying assumptions—the correct specification of the first-stage conditional mean and of the link between first- and second-stage unobservables—that are not required by the 2SLS approach.

Results

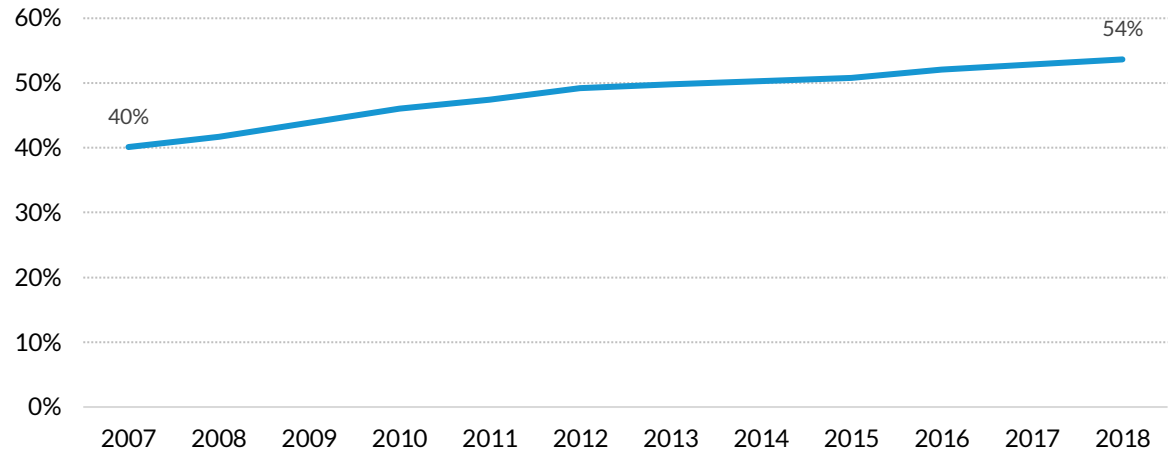
In this section, we report results of our analyses of national aggregate-level changes in patterns of hospice use and decedent-to-survivor spending ratios, followed by results from our individual analyses. Finally, we present results from our zip-code-level analyses. Analyses by CBSA, which are directly comparable to analyses by Hogan (2015), are available in appendix A.

Changes in Patterns of Hospice Use and Decedent-to-Survivor Spending

From 2007 to 2018, overall hospice use by Medicare decedents increased from 40 percent to 54 percent (figure 1). This pattern continues an upward trend in hospice use from prior years (Hogan 2015).

FIGURE 1

Share of Medicare Decedents Ages 65 and Older Using Hospice by Year of Death, 2007–18



Source: Urban Institute analysis of 2007–18 Medicare claims for Parts A and B from the Geographic Variation Database.
Notes: Sample is limited to fee-for-service beneficiaries ages 65 and older.

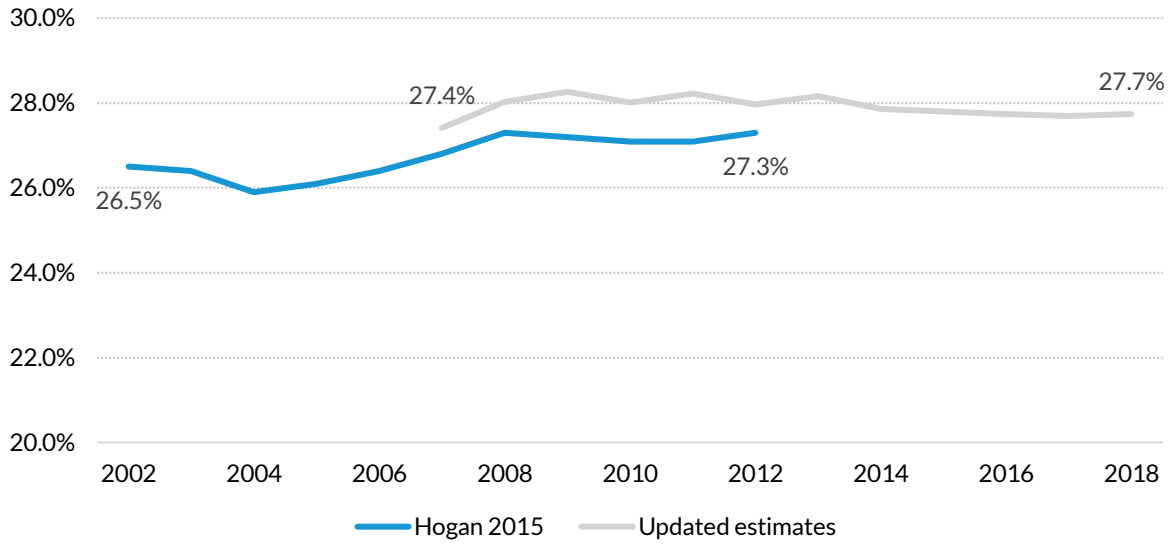
Figure 2 shows the trend in the share of Medicare spending that occurs in the last year of life in our data from 2007 to 2018, not including Part D. The overall trend is essentially flat, with decedents representing 27.4 percent of Medicare spending in 2007 and 27.7 percent in 2018. There is a slight downward trend since the share of Medicare spending in the last year of life peaked in 2009 at 28.3 percent.

Figure 2 also shows the trend in the share of Medicare spending in the last year of life reported previously (Hogan 2015). From 2002 to 2012, there was an overall upward trend, with the share increasing from 26.5 to 27.3 percent. In attempting to follow the same procedures with more recent data, our estimates show we are close to those of the earlier study for the overlapping years, but our estimates are somewhat higher by similar amounts each year.

The increased share of Medicare spending in the last year of life from the earlier period, Hogan observed, coinciding with increasing hospice use, was consistent with the idea that hospice had a net cost to Medicare and cast doubt on earlier findings of lower Medicare spending associated with hospice. We do not see a similar increase in spending in the last year of life in the more recent period we examine in figure 2, though we do see a continued increase in share of decedents using hospice in figure 1.

FIGURE 2

Share of Medicare Spending in the Last Year of Life, 2002–18



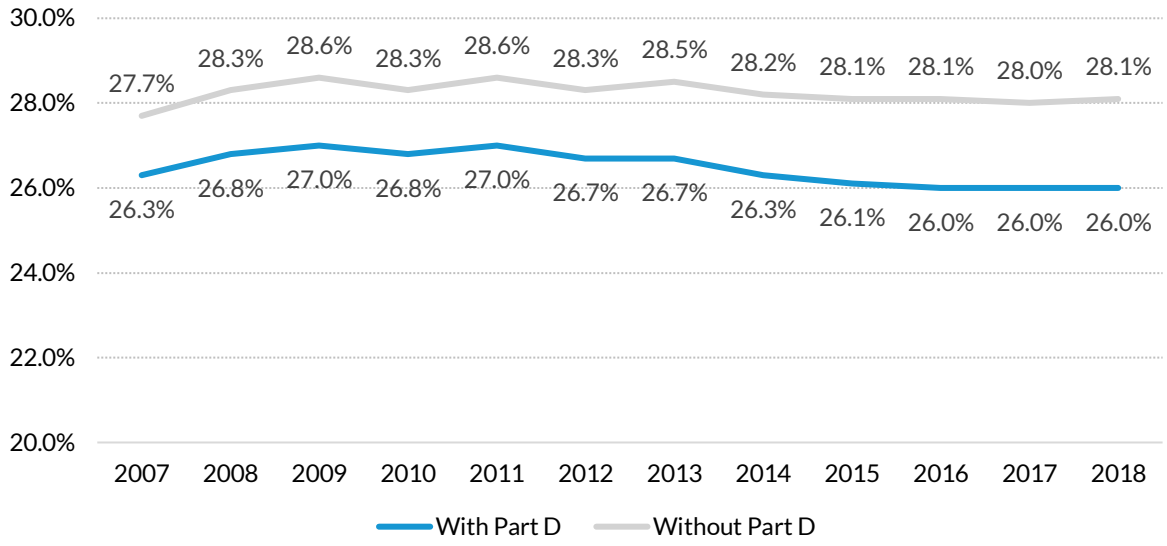
URBAN INSTITUTE

Source: Urban Institute analysis of 2007–18 Medicare claims for Parts A and B from the Geographic Variation Database. Estimates for 2002–12 from Hogan (2015).
Notes: Medicare spending excluding Part D for fee-for-service beneficiaries ages 65 and older.

As shown in figure 3, when we include Part D spending in estimates of the share of Medicare spending in the last year of life, we see a declining trend. From 2007 to 2018, the share of Medicare spending in the last year of life including Part D declined slightly from 26.3 percent to 26.0 percent. Using the more inclusive spending measure, then, we find that the share of Medicare spending in the last year of life fell slightly over the 2007–18 period as the share of decedents using hospice substantially increased (from 40 percent to 54 percent, a 35 percent increase). We do not adjust for changes in Part D coverage rates for decedents or survivors over this period. Trends in Part D coverage may vary between decedents and survivors over our study period.

FIGURE 3

Share of Medicare Spending in the Last Year of Life, 2007–18, with and Without Part D



URBAN INSTITUTE

Source: Urban Institute analysis of 2007–18 Medicare claims for Parts A and B from the Geographic Variation Database and Part D event file.
Notes: Medicare spending for fee-for-service beneficiaries ages 65 and older.

Individual-Level Analysis

Table 4 reports summary statistics for the outcome variables we examine in the individual-level analysis. We report variable means and standard deviations separately for all cases, hospice entrants, and comparison group cases that did not enter hospice. We present statistics for the comparison group both weighted to match the hospice group and unweighted.

As shown in table 4, total spending (Parts A, B, and D) over 18 months is lower for hospice users ($32,995) than for the weighted comparison group of beneficiaries not entering hospice ($40,034). Hospice users have higher Part A spending than matched nonusers, driven by higher spending on hospice care itself, but lower Part B and Part D spending.[11] This is consistent with beneficiaries forgoing conventional care outside the hospice benefit, including life-extending care, non-life-extending care, or both. Some services that would have been provided outside of hospice, such as doctor visits and certain medicines, shift to being provided by hospice. Additionally, standard deviations show substantially less variability in total spending for hospice users than in the weighted comparison group, consistent with

Medicare's hospice per diem payment policy raising the payment floor on the low side of the distribution and hospice limiting very high spending outside of hospice on the other.

Although total spending for the hospice group is lower on average, mean spending per day is higher ($413 for the hospice group and $136 for the comparison group), suggesting that earlier death may drive much of the difference in total spending between the two groups. This is further supported by much higher mortality among hospice entrants compared with weighted nonusers. Overall, 91.2 percent of hospice users died within 18 months of entry, while only 27.2 percent of weighted nonusers died. This is consistent with literature suggesting it is very difficult to predict mortality from claims data (Storick et al. 2019). Perhaps to compensate for this fact, much of the hospice literature compares only those hospice users who died with matched nonusers who also died within the specified study period (e.g., NORC 2023), which effectively assumes hospice has no effect on mortality. Table 4 also shows that, on average, hospice entrants were alive for about 142 of the approximately 548 possible days in the 18-month observation window, as compared with 473 days on average for nonhospice cases.

Thus, although the hospice and comparison groups are well balanced on many relevant characteristics, as shown in appendix table B1, we see substantial differences in outcomes in table 4. This is not surprising. First, hospice use itself is an indicator that a physician has attested that the beneficiary may be expected to live six months or less. Thus, we can expect hospice use embeds important information about risk of death that we cannot observe in claims data. Second, given the prospective nature of our primary research design, we intentionally forgo using some relevant observable information—such as timing of death and utilization and diagnoses observed after the start of the index quarter—as these measures could be affected by hospice use. Including them as controls would add potential endogeneity bias to the analysis, even though they would have high explanatory power and would proxy for unobserved differences that simultaneously affect hospice use and outcomes.

TABLE 4

Summary Statistics for Outcome Variables for Individual-Level Models for Hospice Users and Comparison Group, With and Without Weights

		Hospice use		No hospice use (comparison group, weighted)		No hospice use (comparison group, unweighted)	
		Mean	SD	Mean	SD	Mean	SD
Spending per beneficiary (2018 dollars)							
Parts A, B, D	Total	$32,995	$30,931	$40,034	$48,551	$18,662	$32,918
Part A	Total Part A	$27,985	$27,609	$22,813	$34,998	$8,795	$23,187
	Acute care	$10,048	$15,612	$11,258	$21,881	$5,308	$15,512
	Other inpatient	$251	$3,165	$357	$4,280	$121	$2,487
	Post-acute care	$4,243	$9,602	$9,439	$18,001	$3,054	$10,650
	Home health	$1,103	$2,633	$3,042	$6,671	$980	$3,716
	Long-term care facility	$485	$5,422	$673	$7,292	$185	$3,692
	Skilled nursing facility	$2,305	$6,125	$4,994	$11,764	$1,509	$6,697
	Inpatient rehabilitation facility	$349	$2,996	$729	$4,666	$380	$3,278
	Hospice	$13,443	$20,160	$1,759	$7,473	$312	$3,035
Part B	Total Part B	$4,186	$5,721	$13,090	$21,309	$7,688	$13,967
	Part B institutional	$1,074	$3,007	$4,608	$12,560	$2,523	$7,251
	Outpatient	$866	$2,809	$3,646	$12,260	$2,218	$7,024
	Part B noninstitutional	$3,112	$4,273	$8,482	$15,654	$5,165	$10,816
	Evaluation and management	$1,499	$1,937	$2,868	$3,450	$1,644	$2,192
Part D	Program spending	$824	$3,428	$4,130	$13,853	$2,178	$9,260
Other outcome measures							
	Total spending per day ($)	$413	$759	$136	$738	$45	$287
	Days alive in 6 quarters (max around 548)	142	158	473	148	532	71
	Died within observation window (6 quarters)	0.912	0.283	0.272	0.445	0.064	0.244
N		901,837		18,022,863		18,022,863	
Weighted N		901,837		901,837		—	

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: SD = standard deviation. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice use indicates beneficiary first entered hospice in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending and other outcomes are measured over an 18-month observation window including the index quarter

and are adjusted to 2018 dollars. Weighted comparison group statistics use propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates.

These considerations motivate the instrumental variables (IV) approach we use, which seeks to better isolate what differences in outcomes can be attributed to a causal effect of hospice versus differences in beneficiary condition between the hospice and comparison groups that affect the outcomes but remain uncontrolled. As discussed in the methods section, we use the zip-code distance to hospice as an instrument for estimating the effect of a beneficiary entering hospice during the index quarter on Medicare spending that occurs after the start of the index quarter and through the 18-month observation window.

Table 5 reports the main results from two first-stage model specifications that estimate the effects of distance to hospice on entry into hospice. Both models are estimated with the weighted data, so hospice and nonhospice cases are already balanced on the extensive set of covariates described in table 2. As an additional measure, we further include discrete indicators of age, sex, risk score percentile group, broad condition category (cancer, ADRD/dementia [no cancer], other condition), secondary malignancy, and index quarter as direct controls in the regression models themselves. The two model specifications differ in that the first model controls for CBSA-level fixed effects and the second model controls for three-digit zip code fixed effects. Presenting both allows us to examine the sensitivity of the estimates (in tables 5 and 6) to the granularity of the geographic area controls and, by implication, what variation in the distance instrument is used for identification. Using a zip-code-based instrument with CBSA-level or three-digit zip-code-level fixed effects leverages both the within-area changes in hospice over time as well as cross-sectional variation in distance to hospice within CBSAs or three-digit zip codes to identify hospice effects.

In both 2SLS first-stage models in table 5, greater distance is associated with a lower probability of hospice entry, and the distance effects are statistically significant. For example, compared with being within 20 miles of a hospice, being 20–30 miles away reduces the probability of using hospice by 1.1 percentage points in the first model and 0.9 percentage points in the second model. The coefficients increase in magnitude with farther distance. As a base for comparison, with weights applied, the overall probability of hospice entry is 50 percent by design because hospice and weighted comparison cases have the same weighted N.

Over each of the distance categories, the distance levels are jointly significant with F-statistics of 12.6 in the first model and 17.2 in the second model. Though these F-statistics (which are robust to heteroskedasticity and clustering at the geographic level of the fixed effects) exceed the traditional minimum threshold of F > 10 used to indicate weak instruments (Staiger and Stock 1997), they are moderate in size. Therefore, in the second-stage 2SLS results that follow, along with reporting

conventional significance levels for hospice effects of interest, we report weak-instrument-robust Anderson-Rubin (AR) test statistic p-values for the statistical significance of hospice effects. These complement the conventional significance levels indicated using asterisks in the tables.

TABLE 5

First-Stage Model Coefficients of Distance to Hospice on Hospice Entry with CBSA-Level Fixed Effects and Three-Digit Zip Code Fixed Effects

	CBSA-Level Fixed Effects				Three-Digit Zip Code-Level Fixed Effects			
	Estimate	**SE**	**t-stat**	**p-value**	**Estimate**	**SE**	**t-stat**	**p-value**
20–30 miles	−0.0111	0.0025	−4.51	0.000	−0.0091	0.0022	−4.17	0.000
30–40 miles	−0.0295	0.0060	−4.93	0.000	−0.0193	0.0048	−4.03	0.000
>40 miles	−0.0723	0.0150	−4.83	0.000	−0.0493	0.0077	−6.41	0.000
N				18,924,700				18,924,700
F-statistic of instruments				12.6				17.2
p-value				0.000				0.000

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: CBSA = Core Based Statistical Area; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Model includes covariates as described in the methods section and index quarter fixed effects and CBSA group fixed effects or three-digit zip code fixed effects, depending on specification. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Distance to nearest hospice is measured from the centroid of the beneficiary's zip code to the centroid of the nearest zip code with a hospice. Standard errors are clustered by geographic fixed-effect level. Reported F-statistic is the Kleibergen-Paap rk Wald F-statistic that accounts for heteroskedasticity and clustering.

Table 6 reports least squares and 2SLS estimates of the effect of hospice entry on Medicare spending, both with weights applied, in models that include CBSA fixed effects and three-digit zip code fixed effects. With CBSA fixed effects, hospice entry is associated with $7,603 lower total Medicare spending using least squares and $22,701 less spending using 2SLS. Both effects are statistically significant. The AR test p-value confirms the statistical significance of the 2SLS estimate even with the moderate F-statistic. As expected, the 2SLS estimate is much less precise than least squares, with a standard error of $5,128.

With three-digit zip code fixed effects, the point estimates of the respective least squares and 2SLS models are similar to those in the first model. The 2SLS estimate (-$25,878, standard error $8,186) is less precisely estimated than its CBSA counterpart, and it is significantly different from zero (conventionally and with the weak-instrument-robust AR test).

We report the p-value from a Durbin-Wu-Hausman test (Davidson and MacKinnon 1993) to evaluate the endogeneity of hospice entry. The test evaluates the statistical significance of the difference in the coefficients of hospice entry between the LS and 2SLS models. A statistically significant difference indicates that the LS estimates are inconsistent because of endogeneity and supports the use of an IV approach with the given instrument. The endogeneity tests for both specifications reject the null hypothesis that hospice entry is exogenous (p-values of 0.014 and 0.022, respectively). This leads us to focus on the 2SLS estimates.

TABLE 6

Effect of Hospice Entry on Total Medicare Spending, Least Squares and Two-Stage Least Squares Estimates

	Model with CBSA fixed effects	Model with three-digit zip code fixed effects
LS estimates	-$7,603***	-$7,435***
SE	($611)	($228)
R-squared	0.086	0.088
2SLS estimates	-$22,701***	-$25,878***
SE	($5,128)	($8,186)
AR test (p-value)	0.000	0.009
Endogeneity test (p-value)	0.014	0.022
Overidentification test (p-value)	0.112	0.417
Dependent variable means		
Hospice	$32,995	$32,995
Nonhospice, weighted	$40,034	$40,034
Number of geographic areas	430	893
N	18,924,700	18,924,700

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: AR = Anderson-Rubin; CBSA = Core Based Statistical Area; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending is measured over an 18-month observation window including the index quarter and is adjusted to 2018 dollars. Model includes covariates as described in the methods section and index quarter fixed effects and CBSA group fixed effects or three-digit zip code fixed effects, depending on specification. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by geographic fixed-effect level, are in parentheses.
*** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

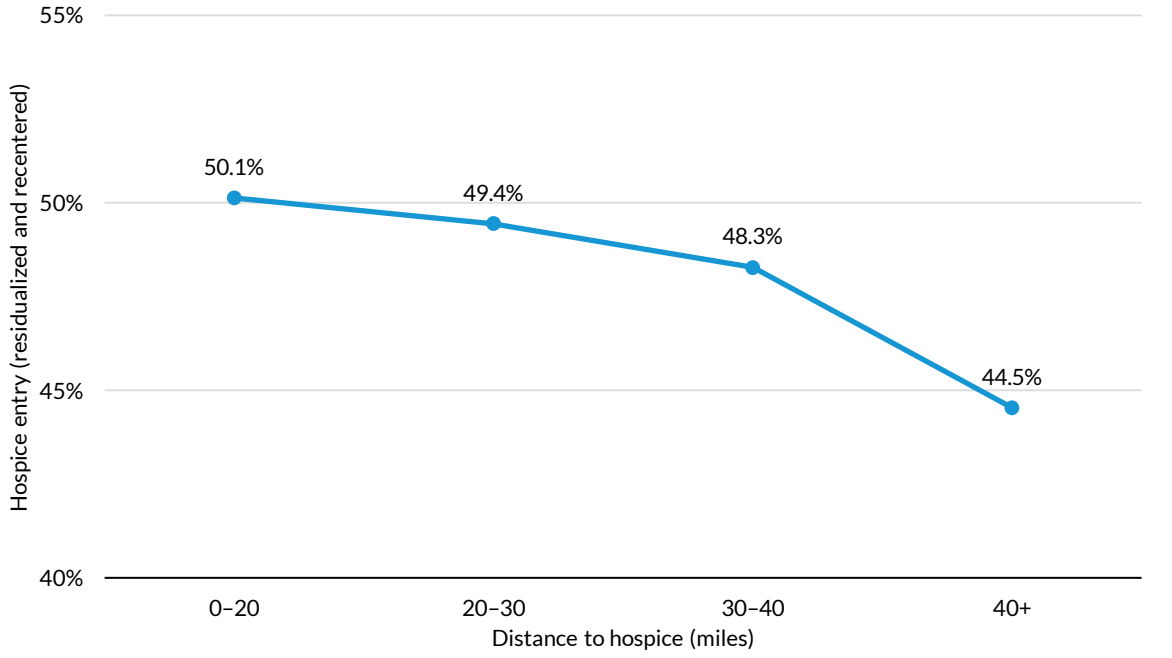
The estimates that include three-digit zip code fixed effects better control for omitted factors that could affect the outcomes and may be correlated with distance, so we would expect them to be less subject to omitted variable bias than the estimates using CBSA fixed effects. The differences in coefficients by specification in least squares and 2SLS estimates are minor, however, suggesting relatively little bias attributable to factors that are not picked up by CBSA fixed effects but are picked up by three-digit zip code fixed effects. Using three-digit zip code fixed effects, the 2SLS estimate is less

precise, but the models agree in their overall findings. Across both specifications, the overidentification tests do not reject the null hypothesis of instrument validity at conventional significance levels. Although the test approaches significance at the 10 percent level with CBSA fixed effects, we cannot reject the overidentifying restrictions for either model.

Insight into how the IV estimates are arrived at can be gained from figures 4, 5, and 6, which provide visual IV plots (Angrist and Pischke 2009). Figure 4 shows how the hospice rate varies by distance to hospice after the hospice rate has been regression-adjusted (i.e., residualized and recentered at its mean) to remove effects of covariates including CBSA fixed effects. It shows how the hospice rate falls as distance to hospice rises. Figures are based on data from the weighted analytic sample, so the weighted mean hospice rate is 0.50 by construction. Figure 5 shows how total Medicare spending, also regression-adjusted for covariates, varies by distance to hospice and shows total spending falling with increasing distance. Figure 6 brings these pieces together. The 2SLS estimate can be obtained from the slope of the weighted regression line running through the points (weighted by the total group-weighted N at each point). Thus, in figure 6 we see a negative slope similar to the point estimates in table 6. It does not align exactly because, for interpretability, the instruments were not similarly residualized.

FIGURE 4

Regression-Adjusted Hospice Use by Distance to Hospice Group



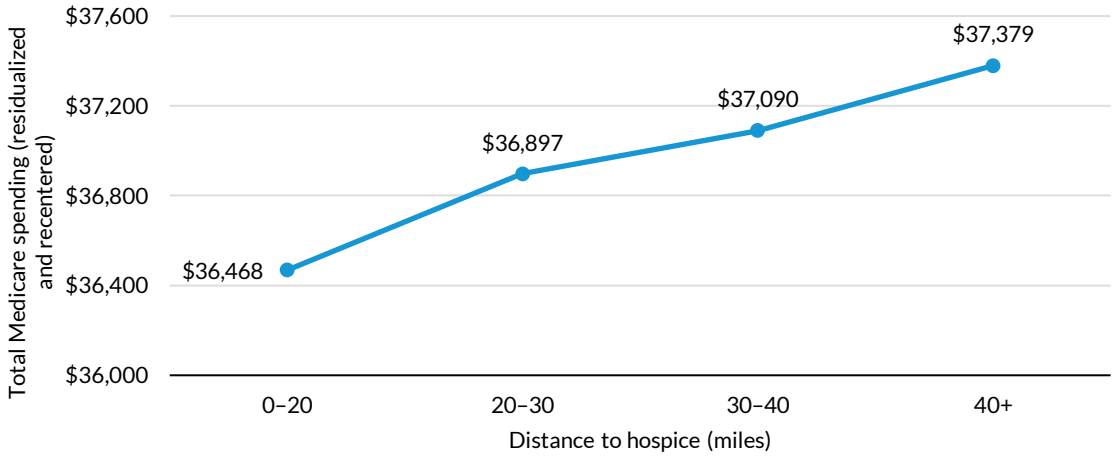
URBAN INSTITUTE

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: Data are from an analytic sample limited to fee-for-service beneficiaries ages 66 and older (N = 18,924,700) with propensity score-based weights applied.

FIGURE 5

Regression-Adjusted Medicare Spending by Distance to Hospice Group

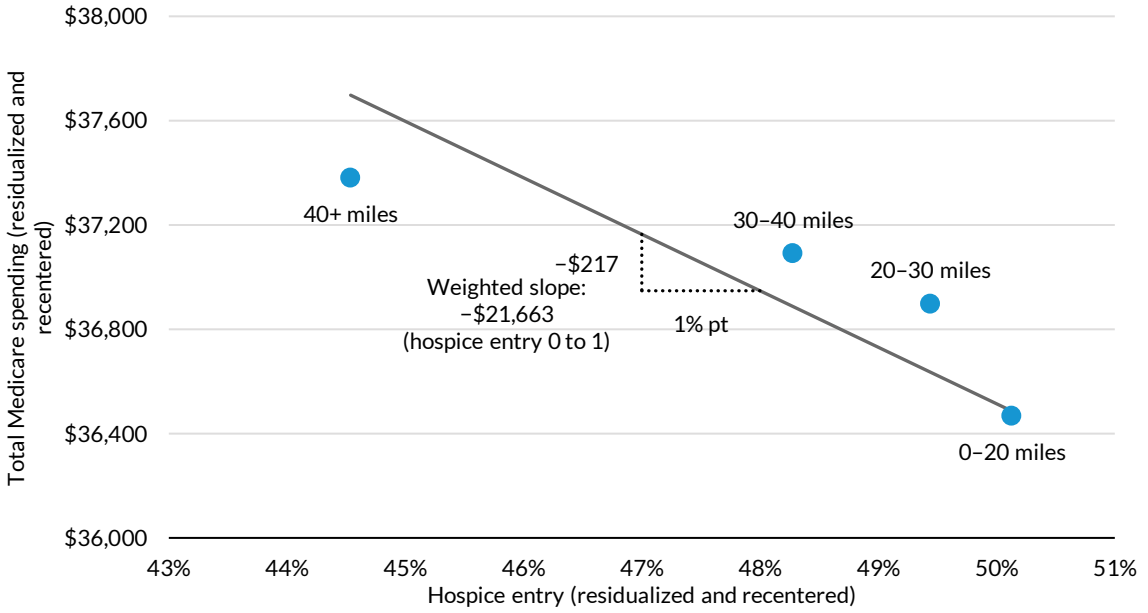


URBAN **INSTITUTE**

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: Data are from an analytic sample limited to fee-for-service beneficiaries ages 66 and older (N = 18,924,700) with propensity score-based weights applied.

FIGURE 6

Regression-Adjusted Medicare Spending and Hospice Use by Distance to Hospice Group with Weighted Linear Regression Fit Line



URBAN INSTITUTE

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Note: Data are from an analytic sample limited to fee-for-service beneficiaries ages 66 and older (N = 18,924,700) with propensity score-based weights applied.

We note that the 2SLS estimates are LATEs for the complier subpopulation—those whose decisions to opt into hospice would be changed by a change in distance to hospice. If the causal effects of hospice are different for different groups of beneficiaries, the LATE for a particular group of compliers of a particular instrument will in general differ from the average treatment effect (ATE) for the overall analysis population (Imbens 2009). In the next section, we examine the characteristics of compliers as compared with the overall weighted sample.

For Which Types of Medicare Beneficiaries Is Hospice Entry Most Affected by Distance to Hospice?

Table 7 examines how characteristics of the complier population, those whose hospice status is affected by the distance to hospice instrument, differ from those of the overall population as measured in our weighted analysis sample. We find that 4.7 percent of the weighted overall analysis sample are compliers, calculated using a single binary instrument of distance to hospice of 35 miles or more (see methods section for derivation). As half of cases in the weighted analysis sample are hospice users, we estimate that 9.4 percent of hospice users are compliers, which is a meaningful but relatively small share.

Table 7 reports mean characteristics for each group and the percentage-point difference for each characteristic, with characteristics ranked from largest positive difference to largest negative difference. The mean values for the weighted analysis sample are directly calculated. But because complier status cannot be observed at the individual level, we estimate means for the complier group based on results of Angrist, Imbens, and Rubin (1996) and Angrist and Pischke (2009) as described in the methods section.

Compliers of the distance-to-hospice instrument are 23 percentage points more likely to have Alzheimer's or dementia and 15 percentage points more likely to have Alzheimer's only. Compliers are 22 percentage points more likely to be age 85 or older, 13 percentage points more likely to be institutionalized, and about 10 percentage points more likely to be female. Compliers are also 14 percentage points less likely to have other cancer and, in each case, about 6 percentage points less likely to be originally eligible for Medicare through disability, to be non-Hispanic Black, and to have health conditions involving five or more body systems. Compliers are 3 percentage points less likely to be a full or partial Medicare-Medicaid dual enrollee. Compliers are similar to the overall population in many respects as well. They have a similar likelihood of being at high mortality risk (among those with inpatient hospital stays), having any ICU/CCU day, having a very high HCC risk score, and having any acute hospital days in the previous quarter.

TABLE 7

Characteristics of Distance-to-Hospice Instrument Compliers Relative to Full Weighted Sample Population

Characteristic	Mean characteristic estimated for complier population	Mean characteristic in full weighted sample N = 18,924,700 Weighted N = 1,803,674	Difference in mean characteristic
Population share in full weighted sample	4.7%	100%	—
Alzheimer's or dementia	0.7409	0.5102	0.2307
Age greater than 85	0.7086	0.4867	0.2219
Alzheimer's only	0.4161	0.2658	0.1503
Institutionalized	0.2952	0.1674	0.1278
Female	0.6992	0.6031	0.0961
High acute spending in previous quarter	0.0813	0.0411	0.0401
Any ICU/CCU days	0.2018	0.1876	0.0142
Risk of mortality (3 or 4)	0.2939	0.2851	0.0089
Esophagus cancer	0.0071	0.0059	0.0013
Non-Hispanic white	0.8830	0.8821	0.0009
Nervous system and brain cancer	0.0072	0.0084	-0.0012
Very high risk score	0.0804	0.0872	-0.0068
Pancreatic cancer	0.0035	0.0122	-0.0088
Hispanic	0.0226	0.0347	-0.0121
Any acute days in previous quarter	0.4663	0.4830	-0.0167
Severity of illness (3 or 4)	0.2700	0.2948	-0.0248
Dual enrollee (full or partial)	0.0937	0.1271	-0.0334
Respiratory cancer	0.0172	0.0648	-0.0476
HCCs in 5+ systems	0.1716	0.2284	-0.0568
Non-Hispanic Black	0.0028	0.0616	-0.0589
Disability (OREC)	0.0374	0.0978	-0.0604
Other cancer	0.2201	0.3618	-0.1417

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: CCU = coronary care unit; HCCs = Hierarchical Condition Categories; ICU = intensive care unit; and OREC = original reason for entitlement. See methods section for derivation of estimated mean characteristics for the complier population.

These findings indicate that, relative to the overall weighted population, distance-to-hospice compliers are more likely to have characteristics associated with long-term care needs and longer hospice stays (e.g., ADRD) and less likely to have characteristics associated with consistently high medical needs (e.g., disability and conditions affecting multiple body systems) and shorter hospice stays (e.g., cancer). The differences in length of hospice stays across health conditions are substantial. MedPAC (2020) reports that among hospice decedents in 2018, the average lifetime length of stay was 151 days for beneficiaries whose principal diagnosis was a neurological condition, a category that includes ADRD, compared with 53 days for beneficiaries with cancer, and an average of 90 days over all conditions. Half of hospice decedents had a lifetime length of stay of 18 days or less, indicating that the distribution is very skewed and concentrated in short stays with a long right tail of lengthy stays. The

90th percentile of lifetime length of stay was 445 days for patients with neurological conditions, compared with 128 days for patients with cancer, and 253 days overall.

Accordingly, we expect the estimated LATEs of hospice for compliers could be larger in magnitude than the ATE of hospice for hospice users overall. The conditions that make compliers more likely to have longer hospice stays, and possibly a larger-than-average effect of hospice, also suggest that they enter hospice earlier relative to end-of-life death than hospice entrants with shorter, more typical stays. For compliers, then, the conventional care for the terminal illness that beneficiaries would otherwise receive outside of hospice benefit would extend over a longer period than for the typical case. This suggests there is more room for hospice use to affect both total spending and days alive for compliers than it would for the typical hospice entrant.

Estimated Effects of Hospice by Spending Component

Table 8 reports estimates of the effects of hospice entry on Medicare spending separately by type of spending. Here and for the remainder of the individual-level analyses, we use LS and 2SLS model specifications that include three-digit zip code fixed effects. We focus on 2SLS results, shown in table 8 and in table 6 previously, because we find evidence that hospice use is endogenous overall and for many subcomponents of spending.

The overall 2SLS point estimate shows a large and statistically significant reduction in spending ($25,878). The main contributors to this total effect are increased hospice spending of $20,647, reduced SNF spending of $22,153, reduced Part B institutional spending of $18,427, and reduced acute inpatient spending of $14,447. The 2SLS effects are significantly different from zero for most components per the AR tests.

The 2SLS effect of hospice entry on hospice spending provides a basic check on the reasonableness of the IV approach. Whereas the sign of the net spending effect of hospice entry on Medicare spending is uncertain, we know for certain that the effect of hospice entry on hospice spending should be positive, and the result shows it is, as well as being statistically significant. From the point estimate of $20,647, we can also infer that the average length of hospice stay for compliers is around 122 days based on average spending per hospice day of $169 in 2018.[12] In contrast, the mean hospice spending of hospice entrants of $13,443 in table 4 implies an overall hospice length of stay of 80 days. A 122-day implied average length of stay for compliers is broadly consistent with the complier sample being disproportionately composed of ADRD patients, as shown in table 7, given the average length of stay of hospice patients with neurological conditions.

Results for some of the spending subcomponents seem unlikely given the structure of the benefit and the literature on hospice (e.g., higher IRF spending of $5,290 and higher evaluation and management spending of $6,826). The IRF result is the hardest to square with the benefit, because most IRF admissions follow a hospital stay, yet we estimate lower acute inpatient spending among hospice entrants, and intensive rehabilitative therapy is at odds with the purpose of hospice. The evaluation and management result implies that hospice entrants saw primary care physicians or nonprocedural specialists more often than nonusers, which is also difficult to reconcile with receiving most basic care from the hospice provider. Several factors may account for these patterns. We note that our model estimates spending from the quarter of hospice entry through an 18-month window (or through death, whichever is sooner). If a beneficiary entered hospice at the end of a quarter, up to three months of prehospice spending may be incorporated into our estimates, and these services may have been used before hospice entry. In addition, unlike for total spending, the overidentification test statistic is statistically significant for several of these individual spending components (including IRF, evaluation and management, SNF, Part B institutional, and Part B noninstitutional spending), raising concerns about instrument validity for those individual component contributions or indicating that the instruments may be picking up heterogeneous effects. Many individual observations have no spending in certain categories, like IRF (which is rarely used among hospice populations), meaning that there may be little effect on the margin for some groups of patients, contributing to heterogeneous effects.

TABLE 8

Effects of Hospice Entry on Medicare Spending by Spending Category, Least Squares and Two-Stage Least Squares Estimates

	LS coefficient		2SLS coefficient		AR test (p-value)	Endog. test (p-value)	Overid. test (p-value)
Total Parts A, B, D Spending	-$7,435	***	-$25,878	***	0.009	0.022	0.417
	($228)		($8,186)				
Part A Spending							
Total Part A	$5,262	***	-$7,248		0.510	0.044	0.618
	($164)		($6,195)				
Acute inpatient	-$1,330	***	-$14,447	***	0.000	0.000	0.802
	($61)		($3,321)				
Other inpatient	-$119	***	-$686	*	0.384	0.157	0.948
	($7)		($406)				
Post-acute care (PAC)	-$5,103	***	-$12,762	***	0.000	0.010	0.003
	($106)		($3,308)				
PAC: Home health	-$1,900	***	$2,711		0.061	0.001	0.390
	($64)		($1,919)				
PAC: Long-term care facility	-$228	***	$1,391		0.077	0.069	0.135
	($20)		($890)				
PAC: Skilled nursing facility	-$2,586	***	-$22,153	***	0.000	0.000	0.000
	($44)		($3,438)				
PAC: Inpatient rehabilitation facility	-$389	***	$5,290	***	0.000	0.000	0.000
	($14)		($1,200)				
Hospice	$11,814	***	$20,647	***	0.000	0.029	0.266
	($96)		($3,380)				
Part B Spending							
Total Part B	-$9,350	***	-$11,866	***	0.002	0.355	0.915
	($78)		($2,606)				
Part B institutional	-$3,732	***	-$18,427	***	0.000	0.000	0.000
	($39)		($2,444)				
Outpatient	-$2,996	***	-$13,035	***	0.000	0.000	0.004
	($38)		($2,002)				
Part B noninstitutional	-$5,617	***	$6,561	**	0.000	0.000	0.000
	($82)		($2,877)				
Evaluation and management	-$1,390	***	$6,826	***	0.000	0.000	0.000
	($25)		($1,295)				
Part D Spending							

	LS coefficient		2SLS coefficient		AR test (p-value)	Endog. test (p-value)	Overid. test (p-value)
	-$3,347	***	-$6,764	***	0.000	0.051	0.006
	($43)		($1,579)				
N	18,924,700		18,924,700				

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: AR = Anderson-Rubin; LS = least squares; 2SLS = two-stage least squares. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending is measured over an 18-month observation window including the index quarter and is adjusted to 2018 dollars. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by 3-digit zip code, are in parentheses. The Kleibergen-Paap rk Wald F-statistic for the excluded instruments is 17.2 in all models. Endog. is a Durbin-Wu-Hausman test of the null that hospice entry is exogenous; Overid. is the overidentification test.

*** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

Effects of Hospice on Spending per Day and Survival Time

Mechanically, hospice could result in lower Medicare spending on net if it (1) reduced average costs per day and had no effect on length of life, (2) reduced length of life and did not affect costs per day, or (3) affected both, depending on their degrees and how the two interact. Hospice could lower average spending per day by reducing use of high-cost conventional care or could increase it given the per diem hospice payment. If more intensive use of costly health care, on the margin, extends length of life to any degree, we would expect hospice to have at least some effect on timing of death. As we have noted, prior literature has produced a wide range of estimates of the effects of hospice on mortality for different populations, or in many cases, has assumed there is no effect.

Table 9 reports LS and 2SLS estimates of the effects of hospice entry on total Medicare spending using the model specification with three-digit fixed effects shown in tables 6 and 8, but it examines three additional outcomes: spending per day, days alive in the observation window, and the share who died within the observation window. Recall that the observation window includes the index quarter and the five subsequent quarters for a total window of 18 months. With total spending equaling the product of spending per day times days alive in the observation window at the individual level, this analysis examines how hospice entry affects both factors.

We focus on the 2SLS results, as all four endogeneity tests in table 9 reject exogeneity of hospice entry. In the 2SLS estimates, we find a negative but not statistically significant effect of hospice entry on spending per day (-$139). However, we find a large and significant effect of hospice entry on days alive in the observation window, with an effect of -213 days (about seven months) over the 18-month window. We see a similarly large and statistically significant effect of hospice entry on death within the observation window (32.9 percentage points).

For spending per day, days alive, and whether the beneficiary died in the observation window, the 2SLS estimates are closer to zero than their corresponding LS point estimates. If the differences in these variables estimated by LS are biased away from zero because of uncontrolled differences in health status, then we would expect consistent 2SLS estimates to be closer to zero, as the findings show. For spending per day, the estimate changes sign rather than simply moving toward zero, from $268 under LS to -$139 under 2SLS. Because total spending for each person is days alive multiplied by spending per day, the effect on total spending depends not only on the per-day and days-alive effects shown, but also on how daily spending and survival interact and covary across individuals. Thus the reported effect on total spending cannot be built up solely from other quantities in table 9. Also, for the hospice and nonhospice groups, the products of mean spending per day and mean days alive exceed mean total

spending because the two factors are negatively correlated at the individual level, consistent with high medical spending concentrated near the end of life.

TABLE 9

Effect of Hospice Entry on Total Medicare Spending, Spending per Day, and Survival Time, Least Squares and Two-Stage Least Squares Models

	Total spending	Spending per day	Days alive in the observation window	Death within the observation window
Models with three-digit zip code fixed effects				
LS estimates	-$7,435***	$268***	-330***	0.638***
SE	($228)	($3.3)	(0.8)	(0.001)
R-squared	0.088	0.071	0.561	0.451
2SLS estimates	-$25,878***	-$139	-213***	0.329***
SE	($8,186)	($96)	(30.3)	(0.064)
AR test (p-value)	0.009	0.483	0.000	0.004
Endogeneity test (p-value)	0.022	0.000	0.001	0.000
Overidentification test (p-value)	0.417	0.788	0.076	0.223
Dependent variable means				
Hospice	$32,995	$413	142	0.912
Nonhospice, weighted	$40,034	$136	473	0.272
N	18,924,700	18,924,700	18,924,700	18,924,700

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: AR = Anderson-Rubin; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending, days alive, and death are measured over an 18-month observation window including the index quarter, and spending is adjusted to 2018 dollars. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by 3-digit zip code, are in parentheses.
*** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

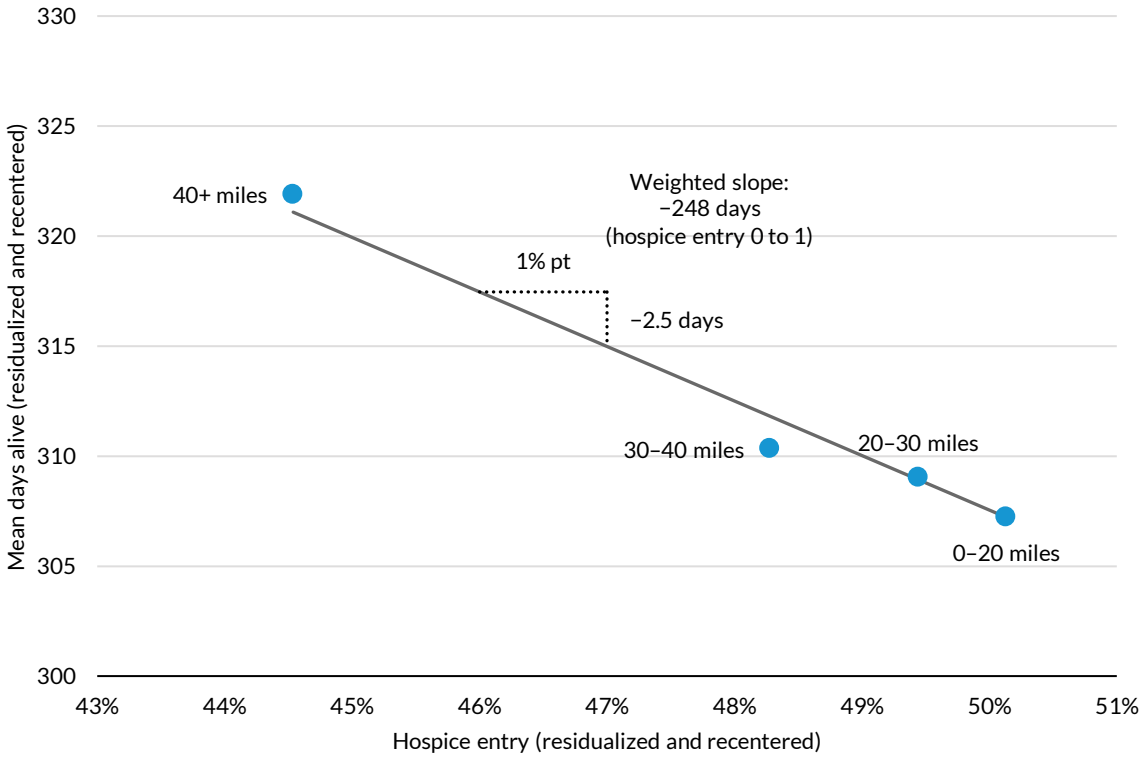
In the context of the few existing findings in the literature reporting mortality differences, the magnitude of the 2SLS-estimated effects of hospice on death-related measures is large, but not an outlier. Prior findings span a wide range. Observational studies that examine survival directly have found hospice use to be associated with similar or longer survival than in comparison groups (Connor et al. 2007; Pyenson et al. 2004). Using an IV approach, Gruber et al. (2025) found that hospice use reduced Medicare spending and decreased survival by an average of seven months among beneficiaries with ADRD who were compliers with their distance-to-for-profit hospice instrument. Thus, the 2SLS estimate for days alive is similar to that in the earlier study using IV, despite other differences between the studies, which we describe in the section on findings in the context of the literature. Compliers in

the current study are more likely than in the overall weighted sample to be diagnosed with ADRD, a condition associated with longer hospice stays and likely with larger-than-average effects of hospice.

Figure 7 shows the visual IV plot for days alive in the observation window. The plot shows a close linear pattern, with hospice use being lower and the number of days alive higher at longer distances to hospice.

FIGURE 7

Regression-Adjusted Days Alive in Observation Window and Hospice Use by Distance to Hospice Group with Weighted Linear Regression Fit Line



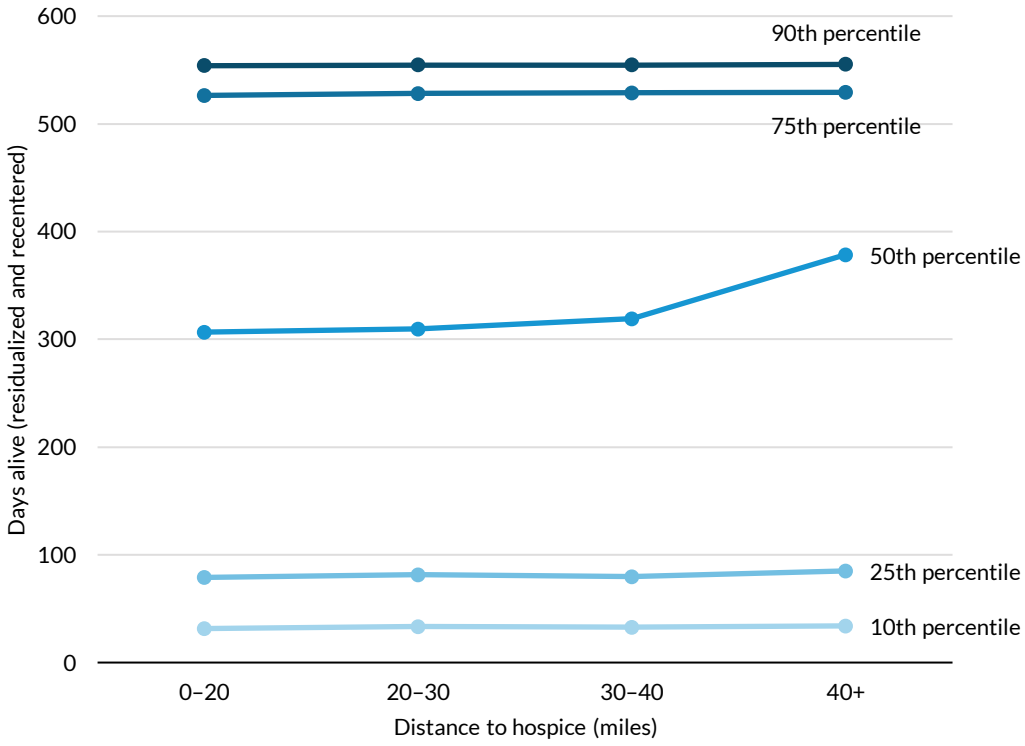
URBAN INSTITUTE

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Note: Data are from an analytic sample limited to fee-for-service beneficiaries ages 66 and older (N = 18,924,700) with propensity score-based weights applied.

Figure 8 shows percentiles of days alive in the observation window by distance to hospice. The graph shows that differences are concentrated around the median of the distribution, with a higher number of days alive rather than at the lower percentiles. The effect for upper percentiles is limited by the fixed length of the observation window.

FIGURE 8

Quantiles of Regression-Adjusted Days Alive in Observation Window and Hospice Use by Distance to Hospice Group



URBAN INSTITUTE

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Note: Data are from an analytic sample limited to fee-for-service beneficiaries ages 66 and older (N = 18,924,700) with propensity score-based weights applied.

Bivariate Probit Model of Death and Hospice

We can gain further insight into how uncontrolled factors may determine both hospice entry and death within the observation window by modeling the two outcomes as jointly determined in a bivariate probit model. Bivariate probit models two binary outcomes as jointly determined by observed factors, as well as unobserved factors for each outcome that are potentially correlated (i.e., share a common unobserved component) (Ashford and Sowden 1970). Correlation is estimated by the model parameter rho. A rho value greater than zero would indicate the presence of unobserved factors that affect the likelihood of hospice entry and death in the same direction, while a negative rho value would indicate the presence of unobserved factors that affect the likelihood of hospice entry and death in opposite directions.

In table 10, we present results for bivariate probit models with the probability of hospice entry specified as a function of distance to hospice and covariates and the probability of death in the observation window specified as a function of hospice entry and covariates. Maddala and Lee (1976) showed that with this recursive specification of the model, with the outcome of the first equation entered as an explanatory variable in the second equation, joint maximum likelihood estimation isolates the causal effect of the first outcome (hospice) on the second (death), while controlling for selection bias and endogeneity. The bivariate probit models are weighted using our base model weights. We present findings for two specifications, first without geographic controls and then with three-digit zip code fixed effects.

TABLE 10

Bivariate Probit Model of Entry to Hospice and Death

	Explanatory Variable	Model Without Geographic Controls		Model with Controls for Three-Digit Zip Code	
		Coefficient	SE	Coefficient	SE
Equation 1. Hospice entry	Distance to hospice				
	20–30 miles	-0.032***	0.007	-0.024***	0.006
	30–40 miles	-0.062***	0.015	-0.050***	0.012
	40+ miles	-0.159***	0.035	-0.129***	0.020
Equation 2. Death	Hospice	1.603***	0.087	1.794***	0.036
Implied marginal effect of hospice on the probability of death		0.526***	0.023	0.574***	0.009
Rho		0.250***	0.043	0.154***	0.020
N		18,924,700		18,924,700	

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Death is measured over an 18-month observation window including the index quarter. Models are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. The bivariate probit consists of two equations: an equation for hospice as a function of distance to hospice, covariates described in methods section and index quarter; and an equation for death as a function of hospice and same covariates as the hospice equation. Rho is the correlation between unmeasured factors affecting both equations. The second model also includes three-digit zip code fixed effects.

*** Estimate is statistically significant at the $p < 0.01$ level, ** at the $p < 0.05$ level, * at the $p < 0.10$ level, using two-sided tests.

Both specifications in table 10 show that longer distance to hospice significantly predicts lower probability of hospice entry. Both show positive and significant effects of hospice entry on the likelihood of death in the observation window. We compute the marginal effect of hospice on the probability of death from the coefficient on hospice to make the estimate comparable to the linear model effects reported in table 9. The marginal effect of hospice entry on death is 52.6 percentage

points in the first model and 57.4 percentage points in the second model. These are somewhat smaller in magnitude than the LS result in table 9 (63.8 percentage points) and larger than the 2SLS result (32.9 percentage points).

The bivariate probit specifications provide direct evidence of unmeasured factors that increase the likelihood of both hospice entry and death. In the first model, rho is 0.250 and is significantly different from zero. In the second specification that controls for three-digit zip code fixed effects, rho is smaller (0.154) and remains statistically significant. Adding the geographic controls in the second model reduced the positive correlation in unobservables, as we would expect. Thus, both specifications indicate the presence of an omitted factor or factors that affect hospice entry and death in the same direction. The inclusion of more detailed controls results in a larger estimated hospice effect, however. These findings are consistent with findings from the 2SLS models that rejected the hypothesis that hospice is exogenous in the model of death. Accounting for that endogeneity with 2SLS led to smaller estimated impacts of hospice entry on death and survival time than the LS model indicated.

With the estimation approach used here, with weights that balance hospice and nonhospice cases on observable covariates and modeling hospice entry and death as jointly determined outcomes with hospice having a potential direct effect on death within the observation window, we find there is a positive, statistically significant correlation in uncontrolled factors affecting both hospice entry and death. The bivariate probit model acts to reduce the magnitude below the LS estimate but reduces it by less than does the 2SLS estimate in table 9. Because rho is positive, estimates that do not account for this correlation overstate the mortality effect, and the estimates are ordered as follows: 63.8 percentage points under LS, 52.6 and 57.4 under the bivariate probit, and 32.9 under 2SLS. The estimated bivariate probit effects are identified by the distance exclusion from the second equation together with the model's functional form, whereas the 2SLS estimates are identified purely from the exclusion restriction. Whereas 2SLS estimates the LATE for compliers, the bivariate probit estimates use its joint distributional structure to estimate the average effect of treatment on the treated for hospice entrants (Chiburis, Jishnu Das, and Lokshin 2012).

Alternative Model Specification That Controls for Death

The large differences between hospice users and nonusers in the probability of death in the observation window suggest that health status is not sufficiently controlled in the hospice and weighted comparison group samples, despite their balance on many important covariates. Hospice cases are, by definition, expected to die within six months, and the comparison group is unlikely to face a similar risk of mortality

given unobserved factors. Analyses above indicated the presence of uncontrolled factors affecting hospice entry and death in the same direction. Although this clearly biases the LS estimates, it should not impair the 2SLS estimates with a valid instrument. Endogeneity tests reject the exogeneity of hospice entry, as would be expected in this setting if the IV approach is working as intended.

We conducted validation checks for the distance instrument described in the methods section, including a placebo test showing that distance to hospice is unrelated to accidental death rates, and a lead-lag test confirming that baseline mortality rates do not predict subsequent distance. Nonetheless, the core exclusion restriction that distance affects Medicare spending only through its effect on hospice use remains an untestable, maintained assumption. Readers who hold restrictive priors over how large mortality effects of hospice can be, and find the magnitude of the 2SLS effects in table 9 for days alive and likelihood of death in the observation window to be implausible, may question the validity of the instrument and estimated effects on spending on that basis. A large estimated effect on days alive may also reflect who the compliers are rather than a failure of the instrument. As shown in table 7, compliers are disproportionately beneficiaries with ADRD, those in institutional settings, and those age 85 or older, groups whose hospice stays are typically far longer than the median stay, so the beneficiaries whose hospice use responds to distance are not those who enter hospice within days of death. In analyses that follow, we present results from alternative models that (1) control for death in the observation window in this section and (2) fully control for days alive in the observation window in the next section.

For the first alternative model, we re-estimated the logit model of the propensity of hospice entry using death in the observation window as an additional covariate. We can view the death indicator as a powerful proxy variable, one that borrows information from the future, for uncontrolled health status differences that capture unobserved factors known to doctors, other medical professionals, the beneficiary, or family/caregivers, but are not already captured by the included covariates. Note that we do not match on days alive at this stage, so the analysis still allows for hospice to affect days alive within the 18-month observation window. When we recompute the weights using propensity scores from the augmented logit model of hospice entry, hospice and weighted comparison groups match on the share of beneficiaries who die during the observation window as well as on the covariates used in the original model (both about 91 percent).

Table 11 presents the alternative estimates with death included in the propensity score model and controlled in the regressions alongside the base model estimates. With death in the observation window held fixed by construction, it does not appear as an outcome in this analysis as it did in table 9. The LS and 2SLS estimates of the effect of hospice on Medicare spending are more similar (-$15,333 and -

$23,682, respectively) in the alternative model than they were in the base model. Both estimates are statistically significant. The endogeneity test does not reject the exogeneity of hospice for total spending in the alternative model. Viewing the death indicator as capturing many factors unobserved in the base model, we should expect less endogeneity in the alternative model, as the findings show. Importantly, with the better control the alternative model provides, the LS result becomes more negative and closer to the 2SLS result, while the 2SLS result for total spending remains about the same as in the base model.

For spending per day, 2SLS estimates show a statistically significant effect of hospice of -$253 per day in the alternative model as compared with a point estimate of -$139 per day in the base model (not significant). With the AR test, however, the p-value of the 2SLS estimate is 0.107.[13] This could arise if hospice entrants both die sooner and incur spending concentrated on hospice care, while the comparison group receives more costly conventional care for at least a portion of the time.

The 2SLS effect of hospice on days alive in the observation window is -84 days (statistically significant), which is smaller in magnitude than the LS estimate of -170 days and smaller than the 2SLS estimate in the base model of -213 days. Unlike for total spending, endogeneity tests for the alternative model reject exogeneity of hospice for spending per day and days alive in the observation window ($p \leq 0.01$ for each outcome). The overidentification test rejects for days alive in the alternative model ($p = 0.003$). A valid exclusion restriction for the base model does not carry over to the alternative model if there is a causal effect of hospice on survival, which could cause the overidentification test to strongly reject here when survival is used as a control.

Controlling for death changes the least squares estimates substantially but leaves the 2SLS estimate of total spending little changed (−$23,682, compared with −$25,878 in the base model). That the 2SLS estimate barely moves, and that the endogeneity test no longer rejects exogeneity for total spending, both point toward the interpretation that death in the observation window proxies for much of the uncontrolled health-status difference that affects the LS estimates, but that 2SLS was correcting for in the base model. These findings indicate the 2SLS total spending result is robust to the additional control. These results also indicate that the estimated spending reduction is not primarily a mechanical artifact of shorter survival, as controlling for mortality changes the baseline 2SLS estimate by only about $2,200.

TABLE 11

Effect of Hospice Entry on Total Medicare Spending, Spending per Day, and Survival Time, Least Squares and Two-Stage Least Squares Models, Base Models and Models with Death in Propensity Model

	Total Spending		Spending per Day		Days Alive in the Observation Window	
	Base model	With death in propensity model	Base model	With death in propensity model	Base model	With death in propensity model
LS estimates	-$7,435***	-$15,333***	$268***	$138***	-330***	-170***
SE	($228)	($290)	($3.3)	($3.2)	(0.8)	(0.6)
R-squared	0.088	0.115	0.071	0.045	0.561	0.522
2SLS estimates	-$25,878***	-$23,682***	-$139	-$253**	-213***	-84***
SE	($8,186)	($7,627)	($96)	($111)	(30.3)	(20.6)
AR test (p-value)	0.009	0.027	0.483	0.107	0.000	0.000
F-statistic for instruments	17.2	24.6	17.2	24.6	17.2	24.6
Endogeneity test (p-value)	0.022	0.277	0.000	0.000	0.001	0.000
Overidentification test (p-value)	0.417	0.961	0.788	0.890	0.076	0.003
Dependent variable means						
Hospice	$32,995	$32,995	$413	$413	142	142
Nonhospice, weighted	$40,034	$48,002	$136	$267	473	313
N	18,924,700	18,924,700	18,924,700	18,924,700	18,924,700	18,924,700

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: AR = Anderson-Rubin; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending and days alive are measured over an 18-month observation window including the index quarter, and spending is adjusted to 2018 dollars. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. In the base model, the predictors in the propensity score model are restricted to measures observed prior to the index quarter. In the alternative model, the propensity score model also includes whether the beneficiary died within the observation window. Standard errors, clustered by 3-digit zip code, are in parentheses. Regression models estimate the effect of hospice entry on each outcome using LS and 2SLS. The 2SLS models use distance to hospice as an instrument for hospice entry.

*** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

Alternative Model Specification That Controls for Timing of Death

In this section, we consider models that fully condition on the number of days alive in the observation window. In other words, we impose as a maintained assumption that hospice has no effect on how long people live through the observation window. This assumption, made explicitly in some of the empirical literature on hospice spending (Zuckerman, Stearns, and Sheingold 2016) and often implicitly, gives us conditional license to analyze the data in productive ways, but which we would avoid if we were not examining how imposing the assumption affects the estimates.

We condition on the number of days alive first by estimating propensity score models within bands of days alive. The bands we used are shown in table 12. In the propensity score models, we further control for individual numbers of days or fine groupings of days. We generate calibrated weights from these models using the same procedures we describe for the main analysis. Covariates are well balanced within each group, and weighted Ns are the same for hospice and nonhospice cases within each group. The total N falls very slightly relative to the base sample (by three cases) because within the days subgroups, singleton cases are created in which a three-digit zip code has only one hospice case or one nonhospice case, and these cases get dropped in the estimation of 2SLS estimates. We dropped these cases to maintain consistent samples within a class of related models.

In the first panel of table 12 we find that total spending increases steeply with number of days alive as we would expect, ranging for hospice entrants from $14,445 for those alive 1 to 29 days to $79,555 for those who survived through the observation window (more than about 549 days). We emphasize that individuals in the hospice group may be in hospice for all or only part of the observation period, so the days measure is not equivalent to the number of days in hospice. Days are measured at the start of the index quarter and not the start of hospice.

Overall, with the days-conditioned weights, average total spending is higher for hospice entrants ($32,995) than for those who did not use hospice ($29,228). Under the base weights reported in table 4, the comparison group average is $40,034. In the first three bands, average spending is less for hospice entrants than for the matched nonhospice comparison cases. But in the two upper bands of 180 days or more, spending is higher for hospice entrants.

TABLE 12

Mean Total Spending and Spending per Day, by Hospice Entry Status and Days Alive in Observation Window, With Weights That Condition on Days

Days alive in observation window or survived	Weighted N (hospice, nonhospice)	Hospice entry (mean)	Unweighted N (nonhospice)	Nonhospice (mean)	Difference in mean (hospice—nonhospice)
Total spending ($)					
1 to 29	122,995	14,445	52,434	15,274	-830
30 to 89	398,154	23,585	92,801	24,778	-1,193
90 to 179	185,075	29,759	126,986	33,084	-3,325
180 to 549	116,648	58,292	879,166	49,950	8,342
Survived through observation window	78,963	79,555	16,871,475	33,748	45,806
All	901,835	32,995	18,022,862	29,228	3,767
Spending per day ($)					
1 to 29	122,995	1,042	52,434	1,061	-20
30 to 89	398,154	410	92,801	433	-23
90 to 179	185,075	256	126,986	283	-26
180 to 549	116,648	191	879,166	167	24
Survived through observation window	78,963	145	16,871,475	62	84
All	901,835	413	18,022,862	421	-8
Days alive					
1 to 29	122,995	18	52,434	18	0
30 to 89	398,154	60	92,801	60	0
90 to 179	185,075	119	126,986	119	0
180 to 549	116,648	312	879,166	314	-2
Survived through observation window	78,963	548	16,871,475	548	0
All	901,835	142	18,022,862	142	0

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Spending and days alive are measured over an 18-month observation window including the index quarter, and spending is adjusted to 2018 dollars. Nonhospice comparison group cases are weighted with propensity score-based weights so that they match hospice cases on covariates and on days alive.

The second panel of table 12 focuses on spending per day. Spending per day falls substantially and monotonically from $1,042 for hospice entrants alive from 1 to 29 days to $145 per day for those who survive through the observation window. Spending per day for nonhospice cases falls similarly by group, from $1,061 to $62. Differences between hospice users and nonusers show modestly lower spending for the first three groups, ranging from $20 to $26 per day. Hospice entrants are more expensive for those alive 180 to 549 days by $24 per day, and spending is higher by $84 per day for those who survive through the observation window. Overall, spending is $8 lower per day for hospice entrants relative to the matched comparison cases. Notably, this is in contrast with the result for total spending, which shows overall total spending is higher for hospice entrants.

Having conditioned on days alive in the propensity scores and weighting, the average number of days alive should be nearly equal, by design, within and across the day groups. The third panel of table 12 reports these differences as a diagnostic. Rounded to the day, average days alive are the same for hospice users and nonusers overall (142 days) and in four of the five subgroups. Days alive differ by hospice status by two days in the 180 to 549 days group.

With the average number of days held nearly fixed, how can there be a positive effect of hospice on total spending and a negative effect on per-day spending? The likely explanation stems from the combination of two factors: (1) heterogeneity in effects of hospice by days alive (evident in the table) and (2) different weighting of days alive for total spending (= per-day spending × days alive) and per-day spending, both within and across groups. The two bands of 180 days or more, where hospice spending is much higher, contribute relatively more to the total spending average, whereas the three shorter-survival bands, where hospice costs less per day, contribute relatively more to the per-day spending average. For regression models evaluating overall effects of hospice, total spending is the more appropriate measure.

Table 13 reports LS and 2SLS estimates of the effects of hospice on total spending using the weights that condition on days alive, estimated separately by group. Because the conditioning was imperfect, however, as shown by the small differences in days in table 12 for some groups, we include days alive as an additional regression control in the models. Models that pool over individual groups at the bottom of the table also control for days-group indicators.

The LS estimates for total spending in table 13 show differences that closely match the differences in means by group shown in table 12. They are highly statistically significant for the five primary groups and show that hospice is associated with lower spending for the three groups with days below 180 and higher spending for the two groups alive 180 days or more, including those who survive through the observation window. The effects are also highly heterogeneous. With all groups combined, hospice spending is estimated to be $4,059 higher than for matched enrollees not using hospice. Weighting up the individual group estimates into one combined estimate gives an overall effect of $3,989, indicating that the pooled estimate and the weighted average of the separate group effects are similar. The overall effect on total spending for the four nonsurvivor groups pooled is small ($45) and insignificant. The straight weighted average of the individual group effects for nonsurvivors is $7.

TABLE 13

Effects of Hospice Entry on Total Medicare Spending Conditional on Days Alive in Observation Window, by Days Subgroups

Days alive in observation window or survived	N	Total spending (LS)		Total spending (2SLS)	AR test (p-value)	F-statistic for instruments	Endog. test (p-value)	Overid. test (p-value)
1 to 29	175,429	-587	***	-9,456	0.221	8.2	0.241	0.261
		(157)		(7,528)				
30 to 89	490,955	-945	***	-797	0.466	12.7	0.931	0.284
		(158)		(8,411)				
90 to 179	312,061	-3,046	***	-15,505	0.010	8.8	0.132	0.024
		(228)		(9,562)				
180 to 549	995,814	8,726	***	-8,524	0.541	6.4	0.134	0.509
		(230)		(13,008)				
Survived through observation window	16,950,438	45,484	***	15,272	0.401	4.2	0.059	0.254
		(243)		(15,453)				
All	18,924,697	4,059	***	-6,435	0.505	15.2	0.118	0.459
		(136)		(6,433)				
Weighted average of group effects		3,989		-4,589				
Overall effect for nonsurvivors	1,974,259	45		-6,542	0.547	15.4	0.323	0.545
		(132)		(6,342)				
Weighted average of group effects for nonsurvivors		7		-6,494				

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: AR = Anderson-Rubin; LS = least squares; 2SLS = two-stage least squares. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending and days alive are measured over an 18-month observation window including the index quarter, and spending is adjusted to 2018 dollars. Model includes covariates as described in the methods section, days alive, index quarter fixed effects, and three-digit zip code fixed effects. Models that pool groups also include days-group indicators. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates and on days alive. Standard errors, clustered by 3-digit zip code, are in parentheses. Overid. is the overidentification test; Endog. is the endogeneity test. Reported F-statistic is the Kleibergen-Paap rk Wald F-statistic.

*** Estimate is statistically significant at the $p < 0.01$ level, ** at the $p < 0.05$ level, * at the $p < 0.10$ level, using two-sided tests.

Table 13 also presents 2SLS analogs to the LS estimates, along with diagnostic tests for the instruments. The aggregate 2SLS point estimates are negative: -$6,435 for all groups combined, -$6,542 for nonsurvivors pooled, -$4,589 for the weighted average of all group effects, and -$6,494 for the weighted average of nonsurvivor group effects. None of the conventional 2SLS estimates are statistically different from zero. Instruments are stronger in aggregate models, with F-statistics of 15.2 for all groups combined and 15.4 for nonsurvivors pooled, but range from 4.2 to 12.7 across the five primary groups. Instrument strength is therefore weak or borderline for several subgroup estimates, especially the survived-through-observation-window group. The weak-instrument-robust AR tests generally do not reject the null of no effect. The 90-to-179-day group is the exception in both cases ($p = 0.010$). The overidentification test also rejects the consistency of the exclusion restrictions in that group. With the weaker instruments within the subgroup models, we take the subgroup-specific estimates as individually less informative than the aggregate-level findings in table 13.

The interpretation of the effects shown in table 13 is dependent on whether the maintained assumption of no effect of hospice on timing of death is true. If it holds, conditioning on days alive can be viewed as removing much of the endogeneity in hospice entry. And in that case, if exogeneity of hospice entry is not rejected—as is the case in the aggregate models, where the instruments are strongest—then what remains is a negative but statistically insignificant 2SLS estimate alongside LS estimates that are consistent and considerably more precise, making the LS estimates the ones to rely on. If hospice entry affects timing of death in any amount, however, then the LS and 2SLS results are both invalid as causal estimates because they control for an endogenous outcome variable.

Comparison of Alternative Estimates of Overall Effect of Hospice on Total Spending

In table 14, we compare the three sets of estimates of the overall effect of hospice on total Medicare spending presented above. Each set makes a different assumption in measuring a potential effect of hospice on timing of death. By design, the model we started with is prospective. It conditions only on information known at the start of the index quarter and estimates outcomes that follow from that point. Recognizing that patient health conditions will evolve and likely affect subsequent hospice entry and total spending, we use the instrumental variables approach with the distance to hospice instrument. This approach is designed to estimate the full effect of hospice entry on total spending, including differences in utilization and potential effects of hospice on timing of death.

We also present findings from two alternative model specifications that impose assumptions about the effect of hospice on the timing of death. In the first alternative, the estimation approach assumes there is no effect of hospice on whether an individual dies in the observation window. The second

alternative makes the still stronger assumption that there is no effect of hospice on the number of days an individual is alive during the observation window.

To recap, table 14 compares results presented previously. 2SLS point estimates are negative in the prospective model and in the first alternative model when death in the observation window is controlled. Exogeneity of hospice is rejected in the prospective model, but not in the alternative model. In the second alternative model that fully controls for days alive in the observation period, which amounts to imposing the assumption of no effect of hospice on the timing of death, the LS estimate is positive and significant, and the 2SLS estimate is negative and not significant, and exogeneity is not rejected (p = 0.118).

What should we make of the results in the third column of table 14 in the context of the other two? If we take $4,059 as the correct point estimate and insist there is no effect of hospice on mortality, it would suggest hospice increases spending overall. If we call the evidence in the third column a wash and declare the overall point estimate zero when days alive are held fixed and we also allow a negative effect of hospice on timing of death of any size, hospice reduces overall Medicare spending. If we consider alternative 1 to be the best compromise between controlling for important omitted factors that death proxies for while still allowing an effect of hospice on timing of death, we see substantively and statistically significant reductions in Medicare spending from hospice.

TABLE 14

Summary of Estimates of Effects of Hospice Entry on Total Medicare Spending under Alternative Assumptions for the Effect of Hospice on Timing of Death

	Prospective model (base): Conditions on covariates known at start of index quarter	Alternative model 1: Conditions additionally on death in the observation window	Alternative model 2: Conditions additionally on days alive in the observation window
LS estimates	-$7,435***	-$15,333***	$4,059***
SE	($228)	($290)	($136)
2SLS estimates	-$25,878***	-$23,682***	-$6,435
SE	($8,186)	($7,627)	($6,433)
AR test (p-value)	0.009	0.027	0.505
F-statistic for instruments	17.2	24.6	15.2
Endogeneity test (p-value)	0.022	0.277	0.118
Overidentification test (p-value)	0.417	0.961	0.459
Dependent variable means			
Hospice	$32,995	$32,995	$32,995
Nonhospice, weighted	$40,034	$48,002	$29,228
N	18,924,700	18,924,700	18,924,697

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: AR = Anderson-Rubin; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending is measured over an 18-month observation window including the index quarter and is adjusted to 2018 dollars. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by 3-digit zip code, are in parentheses.
*** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

Exploratory Analyses with an Alternative IV

As an alternative source of identifying variation, we conducted preliminary analyses that explore the potential for a leave-one-out IV. The alternative IV uses variation in zip-code hospice use rates to identify effects of hospice based on local use patterns. To sever the link between individual-level hospice use and endogenous individual-level factors, the zip code mean is computed by quarter for each case, excluding its own data from the calculation. Preliminary estimates using this approach are reported in appendix table D.1.

The preliminary estimates using this approach are qualitatively similar to those we obtain with the distance instrument but are much more precisely estimated with an F-statistic of 251 for main specifications. The overall effect on Medicare spending is smaller in magnitude (-$7,611), and the estimated effect on days alive is higher (240 days). Several of the questionable findings in the main analyses for Medicare spending by component, however, are reduced in magnitude (for SNF) or not present (for IRF) with the alternative IV. Variation in local practice style, such as referral patterns and hospice availability, that raises hospice use among people with similar needs helps identify causal effects. A limitation of this approach is that it could also pick up endogeneity due to community-level factors that the distance IV may avoid, such as shared preferences for comfort-focused care among patients in a local area. If the leave-one-out rate reflects these characteristics, high-hospice zip codes would differ in spending for reasons other than hospice, and the estimates would partly reflect those differences. Further analyses could consider the value of this or related IV approaches to obtain estimates for subgroups little affected by the distance IV and could produce estimates for a potentially different group of compliers.

Zip-Code-Level Analysis

Table 15 reports summary statistics for the analysis file used to estimate zip-code-level models of the effect of hospice use on total Medicare spending per beneficiary. The dataset contains observations for 386,748 zip-code-years, and the mean total spending is $10,433 (unweighted).

TABLE 15

Summary Statistics for Zip-Code-Group-Level Medicare Spending, Hospice Use, Distance to Hospice, and Covariate Measures

	Mean	Standard deviation
Total Medicare spending	$10,433	2,732
Hospice spending	$330	306
Inpatient spending	$3,482	1,318
Outpatient hospital spending	$1,573	658
Part B excluding outpatient hospital	$2,570	832
Skilled nursing facility spending	$851	555
Home health spending	$487	451
Part D spending	$1,139	620
Share who died	0.048	0.020
Share with any hospice use	0.028	0.016
Distance to hospice (miles, capped at 50)	12.9	11.3
Distance to hospice: 20 to 30 miles	0.128	0.334
Distance to hospice: 30 to 40 miles	0.049	0.215
Distance to hospice: 40 to 50 miles	0.018	0.134
Distance to hospice: 50 miles or more	0.019	0.135
Share female	0.534	0.052
Share age 70 to 74	0.237	0.042
Share age 75 to 79	0.174	0.036
Share age 80 to 84	0.126	0.036
Share age 85 to 89	0.079	0.030
Share 90 or above	0.049	0.026
Share any month long-term institutionalized	0.035	0.036
Share Black or African American	0.069	0.149
Share Asian/Pacific Islander	0.014	0.046
Share Hispanic	0.041	0.107
Share American Indian/Alaska Native	0.011	0.073
Share other race	0.006	0.011
Share unknown race/ethnicity	0.007	0.010
Share with any month Part B coverage	0.928	0.051
Share with any month Part C coverage	0.247	0.166
Share with any month Part D coverage	0.551	0.129
Share full dual, partial dual, or with low-income subsidy at least one month	0.165	0.117
Share with cancer diagnosis, not heart, not Alzheimer's/dementia	0.042	0.018
Share heart diagnosis, not cancer, not Alzheimer's/dementia	0.240	0.060
Share Alzheimer's/dementia diagnosis, not cancer, not heart	0.042	0.021
Share with cancer and heart diagnoses, not Alzheimer's/dementia	0.031	0.015
Share with cancer and Alzheimer's/dementia diagnoses, not heart	0.004	0.005
Share with heart and Alzheimer's/dementia diagnoses, not cancer	0.056	0.029
Share with cancer, heart, and Alzheimer's/dementia diagnoses	0.007	0.007
Mean N of zip code	882	1,250
N	386,748	

Source: Urban Institute analysis of zip-code-year level data derived from MBSF Base, MBSF Cost and Use, Geographic Variation Database, and census data, compiled by MedPAC, 2007–2018.

Notes: Unweighted summary statistics for zip-code-group-year observations, 2007 to 2018. The sample is limited to zip code groups with more than 25 cases in each year. Spending measures are Medicare program spending per enrollee and do not include beneficiary cost sharing.

For the average zip-code-year, the share of individuals who died is 0.048, and the share with any hospice use is 0.028. The average distance to hospice is 12.9 miles using the distance measure capped at 50 miles. For use in the instrumental variable models, we include in first-stage regression models categorical indicators of ranges in distance: 20 to 30 miles, 30 to 40 miles, 40 to 50 miles, and 50 miles or more, with 0 to 20 miles serving as the base category. Distance to hospice is 0 to 20 miles for 78.6 percent of zip-code-years, 20 to 30 miles for 12.8 percent of zip-code-years, and 4.9, 1.8, and 1.9 percent for the other categories, increasing in distance. Table 15 also shows summary statistics for the demographic, coverage, and health condition measures used as control variables in the regression analyses. The average number of Medicare enrollees in each zip code over the sample years is 882, with a standard deviation of 1,250.

Table 16 shows main findings for three first-stage model specifications that predict the share of Medicare enrollees using any hospice in the zip-code-year. Coefficients for the included distance-to-hospice measures are shown. The regression models also control for the covariates listed in table 15, year fixed effects, and zip-code fixed effects (five- or three-digit depending on specification). Standard errors are clustered by zip code at the level of the included fixed effects.

The first model is the first stage of an IV model estimated using 2SLS and includes five-digit zip-code fixed effects. Coefficients on the distance categories are all negative and increase in magnitude with distance. The distance measures are individually statistically significant at the 5 percent level for the farthest two categories and at the 10 percent level for the 30-to-40-mile category. The set of four distance measures is jointly statistically significant with an F-statistic of 11.0 (robust to heteroskedasticity, clustering, and autocorrelation). The second model is the same as the first except it uses three-digit zip-code fixed effects. The coefficients on each of the distance variables are individually statistically significant, and they are jointly significant with a robust F-statistic of 26.3.

The third model is like the second except it is specified as a fractional logit model, a nonlinear variation of the linear second model. Predicted hospice use from the linear models can, and do, have predicted hospice rates outside of the [0, 1] range (negative predicted rates in the present case). The fractional logit model constrains the predicted hospice use rates to fall within the [0, 1] interval. The coefficients in the fractional logit model are on a different scale than in the linear models, so they are not directly comparable. But as in the second model, each of the distance variables is individually statistically significant. They are jointly statistically significant with a chi-squared statistic of 70.6. We have only estimated the fractional logit model with three-digit fixed effects, as it was computationally infeasible to estimate the model with the many thousands of five-digit fixed effects.

TABLE 16

First-Stage IV Models of Hospice Use Using Linear and Fractional Logit Regressions

	Linear regression with five-digit zip FE		Linear regression with three-digit zip FE		Fractional logit regression with three-digit zip FE	
Distance to hospice: 20 to 30 miles	-0.000150		-0.000805	***	-0.031259	***
SE	(0.000212)		(0.000223)		(0.007940)	
t-stat	-0.71		-3.61		-3.94	
Distance to hospice: 30 to 40 miles	-0.000563	*	-0.001849	***	-0.068368	***
SE	(0.000334)		(0.000375)		(0.014609)	
t-stat	-1.68		-4.93		-4.68	
Distance to hospice: 40 to 50 miles	-0.001430	***	-0.003643	***	-0.144241	***
SE	(0.000525)		(0.000726)		(0.032759)	
t-stat	-2.72		-5.02		-4.40	
Distance to hospice: 50 miles or more	-0.004899	***	-0.006966	***	-0.360230	***
SE	(0.000754)		(0.000723)		(0.049885)	
t-stat	-6.49		-9.63		-7.22	
F-statistic for instruments	11.0		26.3		—	
p-value	< 0.0001		< 0.0001		—	
Chi-squared statistic	—		—		70.6	
p-value	—		—		< 0.0001	
N	386,748		386,748		386,748	

Source: Urban Institute analysis of zip-code-year level data derived from MBSF Base, MBSF Cost and Use, Geographic Variation Database, and census data, compiled by MedPAC, 2007–18.
Notes: Models also control for covariates, year fixed effects, and zip code group fixed effects. Weighted by mean five-digit zip code group N. Robust standard errors clustered by zip code group. *** p < 0.01, ** p < 0.05, * p < 0.10. F-statistics are Kleibergen-Paap Wald rk robust F-statistics for first-stage instruments.

Changes in Medicare spending driven by changes in hospice use may be experienced in the same year as the hospice utilization and in the subsequent year (or years). Unlike in the individual-level analysis where we focused on spending associated with individual enrollees over an 18-month period, here we focus on area-level changes in spending that are not tied to specific individuals but are the aggregate experience of Medicare enrollees within each zip-code-year. For a given year, we expect early-year hospice use may affect subsequent spending (hospice and nonhospice) through the remainder of the year. But conceptually, we do not think of late-year-initiated hospice use as causally affecting early-year spending.

On the other hand, both early- and late-year hospice use in the current year may have a direct causal effect (i.e., that may derive from underlying hospice use and spending of specific individuals) on spending in the subsequent year. Hospice use in the current year should affect spending in the following

year only for beneficiaries who remain alive into that year. We may also expect that an area-wide change in hospice use patterns in a year may have a continuing effect on spending into the subsequent and future years.

Table 17 reports estimates of the effect of hospice use on total spending per capita, including LS and 2SLS regressions. We estimate the effect of hospice use in the current year on spending in the contemporaneous year (top panel) as well as spending in the following year (bottom panel), conditional on current-year covariates. The models in table 17 control for covariates in year t, year fixed effects, and zip-code fixed effects at the specified level. With lead-year spending as the dependent variable, the size of our estimation sample drops by one year of observations from 386,748 to 354,519.

Focusing on the effect of hospice use on spending in the current year, using 2SLS with five-digit zip-code fixed effects, the estimated effect of hospice use on total Medicare spending per capita is -$32,589. The effect is statistically significant at the 10 percent level but not at the 5 percent level (p-value = 0.063). The endogeneity test indicates the 2SLS effect is significantly less than the estimated LS effect of $17,536, indicating endogeneity. Using 2SLS with three-digit zip-code fixed effects shows a smaller effect of -$1,392 (not significant), and the endogeneity test does not reject exogeneity. The IV estimate using the fractional logit first stage with three-digit fixed effects shows an effect of hospice use of -$15,207 (statistically significant at the 10 percent level, p-value = 0.090). The model rejects the exogeneity of hospice use. Therefore, the significant and positive LS effect ($14,639) of hospice on spending with three-digit fixed effects should be viewed as biased because of endogeneity, given the IV result (though not based on the corresponding 2SLS result).

TABLE 17

Estimates of the Effect of Zip-Code-Level Hospice Use on Total Medicare Spending per Enrollee

	LS five-digit zip code FE	2SLS five-digit zip code FE	LS three-digit zip FE	2SLS three-digit zip FE	IV with fractional logit first-stage three-digit zip FE
Contemporaneous					
Hospice use (share with any)	17,536 ***	-32,589 *	14,639 ***	-1,392	-15,207 *
Standard error	(368)	(17,551)	(1,178)	(14,594)	(8,965)
p-value	0.000	0.063	0.000	0.924	0.090
R-squared	0.904	—	0.830	—	—
AR test (p-value)		0.196		0.382	
CLR test (p-value)		0.047		0.844	
F-statistic of instruments		11.0		26.3	—
Endogeneity test (p-value)		0.003		0.289	0.000
Overidentification test (p-value)		0.601		0.246	—

Number of observations	386,748	386,748	386,748	386,748	386,748
Following year					
Hospice use (share with any)	-2,314 ***	-48,521 ***	944	3,845	-15,400
Standard error	(383)	(15,877)	(1,207)	(14,596)	(9,420)
p-value	0.000	0.002	0.434	0.792	0.102
R-squared	0.892	—	0.806	—	—
AR test (p-value)		0.003		0.279	
CLR test (p-value)		0.001		0.847	
F-statistic of instruments		10.1		25.0	—
Endogeneity test (p-value)		0.004		0.819	0.068
Overidentification test (p-value)		0.106		0.169	—
Number of observations	354,519	354,519	354,519	354,519	354,519

Source: Urban Institute analysis of zip-code-year level data derived from MBSF Base, MBSF Cost and Use, Geographic Variation Database, and census data, compiled by MedPAC, 2007–18.
Notes: AR test = Anderson-Rubin test and CLR = conditional likelihood ratio test. Both are weak-instrument-robust tests for the significance of the hospice use coefficient. FE = fixed effects; LS = least squares; 2SLS = two-stage least squares. Models also control for zip-year covariates and zip code group and year fixed effects. Weighted by mean zip N. Robust standard errors are clustered by zip code group. *** p < 0.01, ** p < 0.05, * p < 0.10.

Focusing on the effect of hospice use on spending in the following year, the estimated effect of hospice use on total Medicare spending per capita is -$48,521 using 2SLS in the model with five-digit zip-code fixed effects. The effect is statistically significant (p < 0.01). The endogeneity test rejects exogeneity. In the 2SLS model with three-digit zip-code fixed effects, the 2SLS estimate is $3,845, and endogeneity is not rejected. Treating hospice as exogenous implies an insignificant hospice effect. In the IV model using the fractional logit first stage, the estimated effect of hospice on total spending in the following year is -$15,400, which is not statistically significant at the 5 percent level (p-value = 0.102). The p-value of the endogeneity test is 0.068, so exogeneity is rejected at the 10 percent level.

Because the first-stage F-statistics indicate that instrument strength is moderate, table 17 also reports AR and conditional likelihood ratio (CLR) tests, which remain valid for inference on the hospice use coefficient when instruments are weak. The AR test is known to be overly conservative with multiple instruments, with the CLR test preferred in that case. We report them both for comparison but rely on the CLR test for inference. For contemporaneous spending in the model with five-digit zip-code fixed effects, the AR test does not reject the null of no effect (p-value = 0.196) while the CLR test rejects it at the 5 percent level (p-value = 0.047). For spending in the following year, both tests reject the null in the five-digit model, consistent with the statistically significant 2SLS estimate. These weak-instrument-robust tests are broadly consistent with the conventional p-values, with the exception of the contemporaneous five-digit model.

Between the two linear specifications, more detailed control for area-level differences should reduce bias in the model with five-digit fixed effects relative to the model with three-digit fixed effects,

while the first stage retains enough variation in distance to identify effects, albeit with an F-statistic of only moderate strength. Between the linear and nonlinear specifications with three-digit fixed effects, the model with the nonlinear first stage is preferred because respecting the [0, 1] range better approximates expected hospice use given the instruments and covariates.[14] Because exogeneity is rejected in the IV model with the nonlinear first stage, the specification we prefer among those with three-digit fixed effects, we do not give weight to the three-digit LS estimate, notwithstanding that the endogeneity test in the three-digit 2SLS model does not reject exogeneity. We therefore rely on the 2SLS estimates with five-digit fixed effects and the IV estimates with a nonlinear first stage and three-digit fixed effects. Because these two specifications differ both in the level of geographic control and in the functional form of the first stage, we treat the consistency of their negative estimated effects as more informative than either estimate alone.[15]

Table 18 reports 2SLS estimates with five-digit zip-code FEs and IV estimates with the fractional logit first stage using three-digit FEs for total spending in the contemporaneous year (repeating estimates in table 17) and adds corresponding estimates for components of total spending. Effects of hospice on total spending in both models were substantial but statistically significant only at the 10 percent level: -$32,589 (2SLS, column 1) and -$15,207 (IV, column 2). We find positive and significant effects of hospice use on hospice spending of $17,092 (2SLS) and $15,824 (IV). We find negative and significant effects of hospice use on nonhospice spending of -$49,681 (2SLS) and -$31,031 (IV). Thus, the increase in hospice spending is more than offset by a statistically significant reduction in nonhospice spending.

We also find statistically significant negative effects of hospice use on outpatient hospital spending, SNF spending, and Part D spending. The effect on outpatient hospital spending is only statistically significant in the second IV model (-$8,199). The 2SLS point estimate is similar (-$9,727) but not statistically significant. The effect of hospice use on SNF spending is significant in both models (-$11,966 for 2SLS and -$10,553 for IV with fractional logit). The effect on Part D is significant in the second IV model (-$11,500), and the 2SLS estimate of -$5,857 is not significant. The 2SLS estimate for inpatient spending is negative and significant at the 10 percent level (-$15,271), though the corresponding IV estimate is not. No other estimates are statistically significant.

Because the F-statistic of the instruments in the 2SLS models is moderate (11.0), we also report CLR test p-values for the 2SLS estimates in table 18, which provide robust inference for the 2SLS hospice coefficients when instruments are weak. In most cases, the CLR test confirms the conventional significance levels. In three cases, the CLR test provides somewhat stronger evidence of effects than the

conventional standard errors do: total spending (p-value = 0.047, compared with a conventional p-value of 0.063), outpatient hospital spending (p-value = 0.036), and home health spending (p-value = 0.075). Because conventional t-tests rely on a normal approximation that may be unreliable when identification is weak, and the size of the test can therefore be distorted, we base inference on the weak-instrument-robust CLR test. [16]

In analogous estimates (not shown in the table) that examined effects of hospice on spending in the following year, effects were similar in direction, magnitude, and significance. Effects of hospice use on hospice spending the following year are $16,907*** (2SLS) and $13,476*** (IV). The effects of hospice use on nonhospice spending in the following year are -$65,428*** (2SLS) and -$28,876*** (IV).

TABLE 18

Estimates of the Effect of Zip-Code-Level Hospice Use on Contemporaneous Year Medicare Spending per Enrollee, by Type of Spending

Effect of hospice use (share with any) on spending per enrollee, by type of spending	2SLS five-digit FE	IV with fractional logit first-stage three-digit zip FE
Total spending	-$32,589*	-$15,207*
Standard error	(17,551)	(8,965)
CLR test p-value	0.047	—
Hospice spending	$17,092***	$15,824***
Standard error	(1,620)	(409)
CLR test p-value	0.000	—
Nonhospice spending	-$49,681***	-$31,031***
Standard error	(17,609)	(8,792)
CLR test p-value	0.003	—
Inpatient	-$15,271*	-$1,739
Standard error	(8,126)	(4,370)
CLR test p-value	0.070	—
Outpatient hospital	-$9,727	-$8,199***
Standard error	(6,711)	(1,649)
CLR test p-value	0.036	—
Part B, excluding outpatient hospital	-$2,126	$205
Standard error	(4,933)	(3,688)
CLR test p-value	0.606	—
Skilled nursing facility	-$11,966**	-$10,553***
Standard error	(5,379)	(2,365)
CLR test p-value	0.014	—
Home health	-$4,735	$754
Standard error	(3,664)	(2,086)
CLR test p-value	0.075	—
Part D	-$5,857	-$11,500***
Standard error	(4,336)	(4,314)
CLR test p-value	0.210	—
Number of observations	386,748	386,748

Source: Urban Institute analysis of zip-code-year level data derived from MBSF Base, MBSF Cost and Use, Geographic Variation Database, and census data, compiled by MedPAC, 2007–18.

Notes: Models also control for zip-year covariates and zip-code-group and year fixed effects. Weighted by mean zip N. Robust standard errors are clustered by zip code group. *** p < 0.01, ** p < 0.05, * p < 0.10. CLR = conditional likelihood ratio weak-instrument-robust significance test for the hospice coefficient. FE = fixed effects; LS = least squares; 2SLS = two-stage least squares.

Table 19 reports estimates of the effect of hospice use on the share of Medicare enrollees in the zip code who died, using the same model specifications as in table 17. We estimate the effect of hospice use in the current year on death rates in the contemporaneous year (top panel) and in the following year (bottom panel), conditional on current-year covariates.

In the current year, the LS estimate with five-digit zip-code fixed effects is 56.3 percentage points (p-value = 0.000). However, the corresponding 2SLS estimate is -1.2 percentage points and is not statistically significant (p-value = 0.887). The endogeneity test rejects exogeneity (p-value = 0.000). With three-digit zip-code fixed effects, the LS estimate is 47.8 percentage points (p-value = 0.000), while the corresponding 2SLS estimate is -4.6 percentage points and is again not statistically significant, with the endogeneity test again rejecting exogeneity. The IV estimate using the fractional logit first stage with three-digit fixed effects is 29.5 percentage points and is statistically significant (p-value = 0.000), and this model also rejects the exogeneity of hospice use (p-value = 0.000). In the following year, the LS estimate with five-digit zip-code fixed effects is 1.0 percentage point (p-value = 0.000), while the corresponding 2SLS estimate is larger at 6.6 percentage points but is not statistically significant (p-value = 0.480), and the endogeneity test does not reject exogeneity (p-value = 0.583). With three-digit zip-code fixed effects, the LS estimate is 12 percentage points (p-value = 0.000), while the corresponding 2SLS estimate (−1.5 percentage points) is not statistically significant, and exogeneity is not rejected at the 5 percent level (p-value = 0.058). The IV estimate using the fractional logit first stage is 15 percentage points, and exogeneity is not rejected.

The 2SLS point estimates are not statistically distinguishable from zero whether looking at the contemporaneous effects of hospice use or the effects in the following year, while the IV estimates using the fractional logit first stage are positive and statistically significant in both panels. As in the spending results, we prefer the fractional logit first stage over the linear specification with three-digit fixed effects. The preferred three-digit estimates therefore indicate higher area mortality with greater hospice use, a pattern the 2SLS estimates with five-digit fixed effects do not corroborate.

TABLE 19

Estimates of the Effect of Zip-Code-Level Share of Medicare Beneficiaries Who Died

	LS five-digit zip code FE	2SLS five-digit zip code FE	LS three-digit zip code FE	2SLS three-digit zip code FE	IV with fractional logit first-stage three-digit zip FE
Contemporaneous					
Died (share of beneficiaries)	0.5630 ***	-0.0123	0.4780 ***	-0.0460	0.2946 ***
Standard error	(0.0027)	(0.0865)	(0.0058)	(0.0659)	(0.0362)
p-value	0.000	0.887	0.000	0.485	0.000
R-squared	0.735		0.658		
F-statistic for instruments		11.0		26.3	—
Endogeneity test (p-value)		0.000		0.000	0.000
Overidentification test (p-value)		0.872		0.557	—
Number of observations	386,748	386,748	386,748	386,748	386,748
Following year					
Died (share of beneficiaries)	0.0096 ***	0.0656	0.1188 ***	-0.0147	0.1513 ***
Standard error	(0.0024)	(0.0928)	(0.0053)	(0.0636)	(0.0401)
p-value	0.000	0.480	0.000	0.817	0.000
R-squared	0.641	0.150	0.562	0.444	0.562
F-statistic for instruments	—	10.1	—	25.0	—
Endogeneity test (p-value)	—	0.583	—	0.058	0.397
Overidentification test (p-value)	—	0.152	—	0.282	—
Number of observations	354,519	354,519	354,519	354,519	354,519

Source: Urban Institute analysis of zip-code-year level data derived from MBSF Base, MBSF Cost and Use, Geographic Variation Database, and census data, compiled by MedPAC, 2007–18.

Notes: Models also control for zip-year covariates and zip code group and year fixed effects. Weighted by mean zip N. Robust standard errors are clustered by zip code group. *** p < 0.01, ** p < 0.05, * p < 0.10. F-statistics are Kleibergen-Paap Wald rk robust F-statistics for first-stage instruments. FE = fixed effects; LS = least squares; 2SLS = two-stage least squares.

Findings in the Context of the Literature on Effects of Hospice on Spending

The literature on the effect of hospice on Medicare spending reaches mixed conclusions on whether the Medicare hospice benefit increases or decreases program spending. Existing studies often focus on particular groups of hospice users and are not designed to address whether hospice, overall, has a net positive or negative effect on program spending. Some studies show lower Medicare spending for certain subgroups, particularly those with cancer (Fiala et al. 2020; Hung, Hsu, and Wang 2020; Kalidindi et al. 2020; Zuckerman, Stearns, and Sheingold 2016), while more aggregate studies across geographic areas have tended to find increased spending associated with increasing rates of hospice use

(Hogan 2015; Iyer et al. 2020). The Medicare beneficiaries who are eligible for and choose to enter hospice may differ in many ways from otherwise similar Medicare beneficiaries who do not elect hospice, and many of these differences are not observable with claims data (e.g., frailty, risk of death, care-seeking behavior, and family support), making it difficult to hold all such factors fixed in a comparison group that only accounts for observable differences. It is therefore essential to attempt to use a research design that can estimate causal effects in the presence of such unobservable differences. In addition, much of the literature implicitly or explicitly assumes hospice use does not affect mortality. Gruber et al. (2025), however, used a quasi-experimental instrumental variables research design to estimate the causal effects of hospice use on Medicare spending for ADRD patients and found significantly lower Medicare spending and large increases in mortality because of hospice use.

This study drew substantially from Hogan (2015) and Gruber et al. (2025) for the analyses it presents and its methodological design. Hogan, based on analysis of aggregate trends, a literature review seeking to reconcile prior hospice cost studies, and a market-level (CBSA) analysis, concluded that the preponderance of the evidence suggested that Medicare's hospice benefit had modestly increased spending for noncancer patients. Comparing our hospice use and decedent spending results to Hogan (2015) suggests that the national aggregate-level relationship between hospice use and spending in the last year of life changed between our respective study periods (2002–12 for Hogan and 2007–18 in this study). In more recent years, we no longer see evidence at the aggregate level that increasing hospice use is associated with higher relative spending in the last year of life. Instead, we see a slight downward trend in the share of Medicare spending on decedents, coinciding with a substantial increase in the share of Medicare decedents electing hospice. We also see this pattern in CBSA-level regression models that update and extend those of Hogan (appendix A), where hospice use is not significantly associated with decedent spending.

Like Hogan (2015), we find that methodological differences appear to substantially affect whether studies find lower net Medicare spending associated with hospice use. Hogan found that the timing of hospice patient inclusion in the study sample substantially affected results, in part because hospice users tend to have high Medicare spending immediately before hospice entry. Whether that spending proximate to the time of hospice entry is included or excluded from the studies' spending measures dramatically affects findings in a systematic way, with studies that omit such costs being more likely to find lower Medicare spending than those that include them. Our individual-level analysis sought to largely include these prehospice spending spikes by defining hospice use as hospice entry during an index quarter following a fixed starting date (e.g., January 1 of the year). The zip-code-level analysis examines effects on use in both the current and subsequent year after hospice use.

Gruber et al. (2025) estimated the effect of for-profit hospice enrollment on Medicare spending and survival among beneficiaries diagnosed with ADRD. They find that for-profit hospice use reduces Medicare spending by $29,028 over a five-year post-diagnosis window, and separately, that it reduces survival by roughly seven months among the group of patients who would not otherwise have used any hospice. To identify these effects, they use distance to the nearest for-profit hospice as an instrument for for-profit hospice use. Distance-based instruments are common in such studies because proximity to a provider plausibly affects a patient's likelihood of using that provider without directly affecting their health or spending through any other channel, addressing the selection problems introduced by comparing hospice users to nonusers. We adopt that approach as well.

Gruber et al. found a much stronger association between distance to hospice and hospice use than we did here. Part of this difference likely owes to what the instrument predicts. Their instrument is distance to a for-profit hospice, used to predict for-profit hospice use specifically. Our instrument is distance to any hospice, which we use to predict hospice entry generally, regardless of ownership type. We find substantially greater variation within zip code over time in distance to for-profit hospice than in distance to any hospice (and, secondarily, greater variation in the earlier period [2002–2013] used in their study than in the period we examine). The greater variation in the distance-to-for-profit measure versus distance-to-any-hospice likely results from changes in the for-profit status of hospices, which shows up as variation in the distance-to-for-profit-hospice measure, but not in the distance-to-any-hospice measure. Second, distance to for-profit hospice is not a strong predictor of any hospice use in our data. Instead, it appears to only be a strong predictor of for-profit hospice use, which is not the focus of our analysis.

Their design also differs from ours in the window over which spending is measured. They estimate the effect on total spending for five years following diagnosis for ADRD, finding significantly lower spending associated with for-profit hospice use ($29,000 over five years). Their design pools all beneficiaries who enter hospice at any point in the five years following diagnosis into a single estimate. The effect includes all beneficiaries who enter hospice during the five-year window following diagnosis, including those who enter hospice quickly after diagnosis and those who do not enter until year 5. Their approach has the advantage that it includes any increased spending preceding hospice election in the estimates of total spending for the hospice user population. However, the amount of anticipatory spending included depends on the timing of hospice relative to the five-year window. They report a robustness check using a shorter two-year window that gives similar results ($22,100 in lower spending). Our model, in contrast, covers the 18 months following the start of the quarter when

beneficiaries enter hospice for the first time. We control for prequarter spending with our propensity-score-based weighting approach.

Gruber et al. found that hospice use decreased survival by an average of seven months among compliers who would otherwise not have used hospice, which exceeds the six-month expected survival window that doctors are required to certify for hospice entry. When we do not match our treatment and control groups on days until death, we similarly find that hospice users survive 213 fewer days, or 7.1 fewer months, in our 2SLS estimates. Thus, even with differences in instruments, their strength, and whether they are predicting for-profit hospice use or all hospice use, level of data (cohort-level versus individual-level), population (ADRD versus all), and differences in the time periods over which effects are estimated, both prospective IV studies produce estimates that show lower net Medicare spending and effects on mortality of similar direction and magnitudes.

Prior studies estimating the effect of hospice on mortality include Pyenson et al. (2004), who used a matched prospective/retrospective case-control design comparing hospice and nonhospice Medicare beneficiaries. Pyenson et al. is primarily a cost study, finding lower mean and median Medicare costs for hospice patients in most cohorts, but it also reported that those lower costs were not associated with shorter duration until death. Connor et al. (2007), using a similar matched-cohort design across six Medicare populations (congestive heart failure and five cancer types), found hospice patients survived a mean of 29 days longer than nonhospice patients overall, with the difference statistically significant for congestive heart failure, lung cancer, and pancreatic cancer. Temel et al. (2010), in a randomized controlled trial of 151 early palliative care patients with metastatic nonsmall-cell lung cancer, found a 2.7-month median survival benefit for the group receiving early palliative care. Thus, the small literature on the effects of hospice on mortality largely found no effect or a positive effect on survival time.

Discussion

Medicare instituted the hospice benefit in 1982 to improve end-of-life care. The hospice benefit provides Medicare beneficiaries with important end-of-life care options, including pain relief, symptom management, and psychosocial supports (MedPAC 2026). Eligibility for hospice requires that a physician certify that the beneficiary has a life expectancy of six months or less, and beneficiaries receive care for their terminal illness and related conditions under the hospice benefit while forgoing conventional care for that illness provided outside of the hospice benefit. Exchanging conventional terminal-illness care for supportive care at the end of life could lead beneficiaries to receive less intensive care, including less hospital care, and may therefore be expected to lower Medicare end-of-life

spending and improve end-of-life experiences for patients and their families. On the other hand, Medicare pays hospices a daily fee for the patients they enroll, which could increase Medicare costs on net, particularly for those with long hospice stays.

When the benefit was added to Medicare, the Congressional Budget Office estimated that hospice would result in net lower spending in the Medicare program.[17] However, studies of the effects of hospice over subsequent decades have not produced full or consistent answers to the question of whether hospice use, on the whole, reduces or increases Medicare program costs.

In this study, we sought to estimate the effects of hospice use on Medicare spending overall. To this end, we used a variety of approaches, including national aggregate-level analyses, individual-level analyses using inverse probability weighting and instrumental variables methods, and zip-code-level analyses using instrumental variables. We designed our individual-level analyses to be fully prospective to avoid problems we identify with the existing literature. We have further conducted individual-level analyses that estimate effects of hospice under maintained assumptions that hospice entry does not affect whether an individual dies within an 18-month observation window, and alternately, that hospice has no effect on the timing of death within the observation window. Each of these approaches is designed to address selection effects inherent in studying hospice use while including nearly all hospice users and Medicare spending for beneficiaries ages 66 and older. We also estimated the effects of hospice entry on factors that contribute to total spending, spending per day, and days alive in the observation window, as well as whether an individual died in the observation window.

In prospective individual-level 2SLS models we find significantly lower Medicare spending associated with hospice entry. These and other 2SLS effects pertain to the population of instrument compliers, that is, individuals whose use of hospice is affected by whether the location of the nearest hospice is near or far. We find that the compliers are more likely to be beneficiaries with longer-term illness, including those in institutional settings, those with ADRD, and those age 85 or older. Compliers have a similar likelihood of being at high mortality risk (among those with inpatient hospital stays), having any ICU/CCU day, and having a very high HCC risk score. They are less likely than average to be non-Hispanic Black, to have conditions involving five or more body systems, and to have cancer. With the prospective base models, 2SLS estimates also find a statistically insignificant negative effect of hospice on spending per day, a large statistically significant negative effect of hospice on days alive in the observation window, and a large statistically significant positive effect on the likelihood of death in the observation window. Thus, it appears that the effect of hospice entry on total Medicare spending we estimate is driven substantially by its effect on days alive. The magnitudes of the total spending and

days alive effects are similar to those reported by Gruber et al. (2025), who used a similar instrumental variables approach, though differing from this study in several other respects we describe above.

In interpreting these life-expectancy effects, we note the 2SLS estimates rely on the assumption that the distance-to-hospice instrument is valid. Even if the instrument is valid, these estimates are LATEs specific to the complier population and are not necessarily generalizable to the overall population of hospice users. Finally, hospice compliers have chosen a model of care focused on quality rather than length of life, so an estimated reduction in survival time does not necessarily indicate a poor outcome for those beneficiaries if it reflects their own preference for hospice care over conventional care.

In individual-level models that control for death, which amounts to imposing an assumption that hospice use has no effect on likelihood of death during the 15- to 18-month observation window, the 2SLS effect of hospice use on total spending is negative, statistically significant, and of the same magnitude as in the base model. When we control for death at any time during the observation window, which is a potentially endogenous outcome measure, but also a measure that proxies for a bundle of otherwise uncontrolled factors with very high explanatory power, our overall estimate of the effect of hospice on total Medicare spending remains about the same. Again, this effect pertains to the population of instrument compliers specific to this estimation model. With this approach, we find a substantially smaller (than in the base model) but still large negative effect of hospice use on days alive in the observation window.

When we fully control for the number of days alive in the observation window, the estimated effects change substantially. Fully controlling for days alive amounts to imposing the assumption that there is no effect of hospice on survival time. The 2SLS estimate with this alternative model is statistically insignificant, with a smaller negative point estimate than in the other models. Taken as a group, these findings show that estimated effects of hospice entry on total spending depend greatly on what is assumed and imposed in the study design about hospice's effects on mortality and survival time. The base model makes no such assumption about mortality effects, but it relies on the assumption that the instrument is valid, which is untestable in the data. If there is a causal effect of hospice entry on mortality or survival time, the two alternative estimation approaches are invalid as causal estimates because they condition on an endogenous outcome measure (mortality and days alive) and therefore should be viewed as descriptive. The 2SLS estimates in the two alternative approaches are only valid if their respective assumptions hold, and if the distance-to-hospice instrument is valid within those models.

The zip-code-level analyses provide a complementary test with aggregated data of whether hospice use affects Medicare spending in overall area-level spending patterns. Models estimated with 2SLS with five-digit zip code fixed effects and IV with three-digit zip code fixed effects and a fractional logit first stage find statistically significant positive effects of hospice use on hospice spending and significant negative effects on nonhospice spending, with the implied reductions in nonhospice spending exceeding the increase in hospice payments. The resulting estimates generally point toward lower net Medicare spending in the current year, with magnitudes broadly similar to individual-level results that do not control for days alive. The net spending estimates are statistically significant at the 10 percent level for current-year spending and at the 1 percent level for following-year spending in models that include zip code fixed effects. Parallel zip-code-level analyses examining mortality find smaller and statistically insignificant effects of hospice use on mortality, whereas the nonlinear first-stage control-function specification suggests a positive relationship between hospice use and death rates, though smaller than the mortality effects estimated in individual-level models that do not control for days until death. Taken together, the overall spending effects of the zip-code results are broadly consistent with the results of the individual-level models. They do not provide consistent evidence on the effect of hospice use on mortality. All 2SLS effects are insignificant; the IV models with nonlinear first stages are positive and significant in both the current and following years.

Further research using methods for causal inference is needed to disentangle what drives lower or higher Medicare spending associated with hospice entry for specific subgroups. We find that estimates using distance to hospice as an instrumental variable estimate effects for a particular subgroup of compliers with characteristics associated with longer hospice stays. Estimates for such cases are valuable because they focus on the types of patients for which the literature has not converged on the direction of spending effects, unlike for cancer. However, the approach does not appear to be an applicable identification method to estimate effects for other types of hospice users. More analysis is needed to better understand the mechanics behind any effect of hospice on mortality and/or the reasons why distance-to-provider IV or other approaches might overestimate such effects. Qualitative work could also shed light on what may be plausible effects of hospice on days alive by condition groups, allowing for better modeling of the effects of hospice on Medicare spending. The value of the hospice benefit for patients and their families, which this work did not examine, is of prime importance regardless of whether Medicare's hospice benefit on net reduces or increases program spending.

Appendix A: Aggregate Analyses by Core Based Statistical Areas

Associations Between Medicare Spending and Hospice Use at the CBSA Level

We produced updated regression analyses analogous to those reported by Hogan (2015) that examine whether changes in hospice use rates were associated with changes in the ratio of decedent-to-survivor spending at the CBSA level. The CBSA-level analysis uses the same data for each year with the same sample exclusions used for the national patterns of decedent and survivor spending. To conduct this analysis, we created 430 CBSA groups—individual CBSAs and indicators for areas in each state outside a designated metropolitan CBSA. To maintain consistent geographic areas each year, we assigned CBSA groups based on the county assignments to CBSAs in 2018 and applied these to the beneficiary counties for all years from 2007 to 2018.[18]

Spending Measures

For each CBSA group, we calculate the spending for decedents and survivors per person-year using the same calculations described above for the national analysis and applying them by CBSA group. Here, we do not reweight to control for changes over time in the distribution of beneficiary age and sex within CBSA group. Instead, we control for characteristics of beneficiaries who died during the calendar year in the CBSA group in regression analyses.

Regression Analyses

The key outcome in our primary analysis is the CBSA-group average spending per person-year in the last year of life. We use three model specifications to examine how this outcome is associated with hospice use: (1) in a model controlling for survivor spending; (2) logged in a model controlling for logged survivor spending; and (3) with the outcome deflated by average spending per person-year for survivors. In each case, we use the overall CPI to adjust the estimates to 2018 dollars. We control for survivor spending in models 2 and 3 to control for underlying spending trends in CBSAs.

We follow Hogan (2015) in defining independent variables for those who died in the CBSA during the calendar year. The key independent variable is the share of decedents using hospice. Along with any controls for survivor spending, we include controls for the average age, sex, and shares of duals, cancer patients (as measured in the MBSF), non-Hispanic white, and non-Hispanic Black, as well as the share of

beneficiaries ages 66 and older in the CBSA enrolled in Medicare Advantage. All variables are measured by CBSA group and year.

We estimate each model using linear regression, with the controls described above and CBSA group and time fixed effects. The controls (other than Medicare Advantage penetration) match those used in Hogan's work and align in time with the measure of hospice use.

We estimated the models both unweighted and weighted by the number of calendar-year decedents in the CBSA in the year. The unweighted results give equal weight to each CBSA group—and thus essentially give an estimate of the average effect expected for a CBSA group. The CBSA group is not necessarily a unit of interest on its own, particularly for the "rest of state" CBSAs we constructed from non-CBSA counties within a state. The weighted models rely more on the relationships observed in the larger CBSAs, yielding overall estimates more in line with relationships where the decedents tend to live. The difference between the unweighted and weighted estimates offers insight into potential heterogeneity in the effects of hospice.

As noted above, a major limitation of the prior hospice literature is that matching hospice users to nonusers using observed characteristics available in claims data does not account for potentially important unmeasured differences. Estimates of the effects of hospice use on spending that rely on matched comparison groups may fail to account for differences in care-seeking and other preferences between hospice users and nonusers. In the CBSA analysis, which has no matched comparison group, a similar concern could arise if CBSAs that had larger increases in hospice use over the study period also had populations whose spending was changing for reasons other than hospice use, such as provider, beneficiary, or health system changes. To address this issue, we also attempted an IV approach. Our instrument is the annual CBSA-group average of the zip-code-level distance to hospice for those who died in the year. However, in the weighted model, the instrument does not meet the usual standards for the first-stage relationship, lacking a statistically significant relationship with hospice use. In the unweighted model, the relationship was statistically significant, but the F-statistic is 7, which is below the standard of 10 used to identify weak instruments (Staiger and Stock 1997). We therefore do not report analogous IV analyses in this report.

Association Between Hospice Use and Medicare Spending by Core Based Statistical Area

Table A.1 provides summary statistics for the data used in the CBSA-level analyses. Average spending in our sample is $51,127 for decedents and $8,454 for survivors. For the average CBSA, 48.1 percent of decedents use hospice, but this varies across CBSAs with a standard deviation of 10.6 percentage

points. In the average CBSA, the distance to hospice (with a 50-mile cap applied) is 7.6 miles, with a standard deviation of 5.5 miles.

APPENDIX TABLE A.1

Descriptive Statistics for CBSA-Level Models

	Mean	Standard deviation
Average spending per person-year for decedents*	$51,127	$8,261
Average spending per person-year for survivors*	$8,454	$1,085
Ratio decedent-to-survivor average spending per person-year, no Part D	6.570	0.644
Ratio decedent-to-survivor average spending per person-year, with Part D	6.046	0.573
Share of decedents using hospice	0.481	0.106
Distance to hospice (miles, capped at 50)	7.629	5.534
Share of residents ages 66 and older in Medicare Advantage	0.252	0.141
Share of decedents who are female	0.540	0.030
Average age of decedents	82.376	1.025
Share of decedents who are white	0.876	0.123
Share of decedents who are Black	0.064	0.076
Share of decedents with cancer	0.175	0.022
Share of decedents dually eligible for Medicare and Medicaid	0.273	0.075
Share of decedents who are hospice users with stays up to 14 days	0.220	0.057
Share of decedents who are hospice users with stays of 15 days to 6 months	0.198	0.043
Share of decedents who are hospice users with stays over 6 months	0.062	0.028
Share of decedents who are hospice users with cancer	0.100	0.021
Share of decedents who are hospice users without cancer	0.381	0.092

Source: Urban Institute analysis of 2007–18 Medicare claims for Parts A and B, Geographic Variation Database, and Part D event file.
Notes: CBSA = Core Based Statistical Area. *Average spending includes Part D. CBSA averages for Medicare fee-for-service beneficiaries ages 66 and older, demographics for decedents in calendar year. Decedent spending is measured within the last year of life, adjusted to person-year, 2018 dollars. Models include CBSA fixed effects. Summary statistics are unweighted.

Table A.2 reports effects of hospice use among those who died during the calendar year on decedents' Medicare spending per person-year, estimated by linear regression models in three model specifications, weighted by the number of decedents in each CBSA-year. Each model controls for share with Medicare Advantage, average age, share white, share Black, share with cancer, share dual eligible, year fixed effects, and CBSA fixed effects.

The first model uses decedent spending per person-year within the last year of life as the dependent variable. The model controls for spending for survivors and the other covariates listed in table A.1. Survivor spending is used to control for overall price and utilization changes that affect Medicare spending broadly. Increasing the share of decedents using hospice from 0 to 100 percent is associated

with a $3,042 increase in Medicare spending for decedents, but the association is not statistically significant.

Model 2 is analogous to model 1 except that the dependent variable is logged to make the results more easily interpretable. As shown in table A.2, a 10 percentage-point increase in hospice use among decedents is associated with a 0.6 percent increase in decedent spending per person-year, which translates to about $306, but this effect is not statistically significant.

Finally, model 3, which is most analogous to the Hogan approach, shows that increasing the share using hospice from 0 to 100 percent is associated with a 35.4 percentage-point increase in the decedent-to-survivor spending ratio, though the association is not statistically significant. This is directionally the same as Hogan's results, though the association has weakened.[19]

APPENDIX TABLE A.2

Effect of CBSA Hospice Use Among Decedents on Decedent Medicare Spending Measures per Person-Year

Model/outcome measure	Coefficient on hospice use share among CBSA decedents	Standard error
1. Decedent spending per person-year, controlling for survivor spending	$3,042	$2,190
2. Log decedent spending per person-year, controlling for survivor spending	0.063	0.040
3. Decedent spending per person-year, deflated by survivor spending (ratio)	0.354	0.283

Source: Urban Institute analysis of 2007–18 Medicare claims for Parts A and B, Geographic Variation Database, and Part D event file.

Notes: CBSA = Core Based Statistical Area. ** $p < 0.01$, * $p < 0.05$. CBSA averages for Medicare fee-for-service beneficiaries ages 66 and older, demographics for decedents in calendar year. Decedent spending is measured within the last year of life, adjusted to person-year, 2018 dollars. Models include CBSA fixed effects and year dummies. CBSAs are weighted by number of decedents. Control variables are the measures of Medicare Advantage, female, age, Black, white, cancer, and duals as described in table A.1. Program spending includes Part D.

Table A.3 provides full model results for model 1 in table A.2. We find that the Medicare Advantage share is more strongly associated with spending of decedents than hospice use. Demographic shares have large effects but are generally not statistically significant, except for average age.

APPENDIX TABLE A.3

Effect of CBSA Hospice Use on Decedent Total Medicare Spending per Person-Year, Including Part D

	Coefficient		Standard error
Hospice use, share, CBSA decedents	$3,042		$2,190
Survivors, average payments per person-year, CBSA, with Part D	$3.39	**	$0.505
Medicare Advantage, share, ages 66 and older	-$7,702	**	$1,976
Female, share, CBSA decedents	-$624		$2,607
Age, average, CBSA decedents	-$869	**	$232
White, share, CBSA decedents	-$8,394		$6,385
Black, share, CBSA decedents	-$12,535		$12,114
Cancer, share, CBSA decedents	$1,615		$3,484
Dual eligible, share, CBSA decedents	$1,217		$3,184
2008	$2,171	**	$167
2009	$4,124	**	$438
2010	$3,476	**	$464
2011	$3,822	**	$463
2012	$2,423	**	$393
2013	$2,778	**	$417
2014	$1,665	**	$438
2015	$2,642	**	$591
2016	$2,367	**	$621
2017	$2,526	**	$669
2018	$3,178	**	$784
Number of observations	5,148		
Number of CBSAs	429		
R-squared	0.986		

Source: Urban Institute analysis of 2007–18 Medicare claims for Parts A and B, Geographic Variation Database, and Part D event file.

Notes: CBSA = Core Based Statistical Area. ** p < 0.01, * p < 0.05. CBSA averages for Medicare fee-for-service beneficiaries ages 66 and older; demographics for decedents in the calendar year. Models include CBSA fixed effects and control for survivor spending per person-year. CBSAs are weighted by number of decedents. Hospice use by decedents includes prior years.

Table A.4 estimates the effects of hospice use on Medicare spending per person-year for decedents at the CBSA level by subgroup. The model specification is the same as that used for all hospice users in table A.3 (and model 1 of table A.2). Subgroups are defined by presence of cancer and length of hospice stay.

Among decedents, hospice use was associated with about $16,000 less spending for patients with cancer, a statistically significant effect. In contrast, hospice use was associated with about $6,800 more spending for decedents who did not have cancer. We found no statistically significant association between hospice use and Medicare spending among length-of-stay subgroups. However, the coefficients suggest that hospice stays longer than six months are associated with much higher spending, though the difference was not statistically significant.

APPENDIX TABLE A.4

Effects of CBSA Hospice Use Among Decedents on Medicare Decedent Spending per Person-Year by Presence of Cancer and by Hospice Length of Stay

	Decedent spending per person-year, including Part D		Standard error
Impact of hospice use across subgroups			
By Presence of Cancer			
Hospice and cancer	-$16,338	**	$5,362
Hospice and no cancer	$6,797	*	$2,656
By length of hospice stay			
Hospice up to 14 days	$1,527		$2,347
Hospice 14 days–6 months	$850		$3,220
Hospice over 6 months	$11,848		$6,772

Source: Urban Institute analysis of 2007–2018 Medicare claims for Parts A and B, Geographic Variation Database, and Part D event file.

Notes: CBSA = Core Based Statistical Area. ** $p < 0.01$, * $p < 0.05$. CBSA averages for Medicare fee-for-service beneficiaries ages 66 and older, demographics for decedents in calendar year. Models include CBSA fixed effects and control for survivor spending per person-year. CBSAs are weighted by number of decedents. Hospice use by decedents includes prior years. Cancer is defined based on the Master Beneficiary Summary File for the year of death. See appendix table A.2 for a full list of control variables.

Overall, these findings, in contrast to Hogan's earlier work, do not show an overall increase in spending associated with hospice. However, we do find a statistically significant, positive association between hospice use and Medicare spending for beneficiaries without cancer ($p < 0.05$) and a statistically significant, negative association between hospice use and Medicare spending for beneficiaries with cancer ($p < 0.01$). The point estimates are at least suggestive that longer hospice stays are more expensive and that only for the cancer patient subgroup do we find lower Medicare spending associated with hospice use.

Appendix B: Descriptive Statistics for Predictors Used in Individual-Level Propensity Models

Appendix table B.1 reports descriptive statistics for the full set of predictors used in the individual-level propensity models of hospice entry, along with the distance-to-hospice measures used as instruments in the 2SLS models. Separate columns show all beneficiaries, hospice entrants, and nonhospice comparison cases with and without propensity-score-based weights applied.

APPENDIX TABLE B.1

Descriptive Statistics for Predictors Used in Individual-Level Propensity Models, With and Without Weighting Comparison Cases

		All, unweighted	Hospice	Nonhospice, weighted	Nonhospice, unweighted
Hospice	Entered in index quarter	0.048	1.000	0.000	0.000
Instrument for two-stage least squares					
Distance to hospice	0 to 5 miles	0.516	0.532	0.537	0.516
	5 to 10 miles	0.210	0.211	0.205	0.210
	10 to 20 miles	0.178	0.172	0.169	0.179
	20 to 30 miles	0.062	0.058	0.060	0.062
	30 to 40 miles	0.020	0.018	0.019	0.020
	>40 miles	0.013	0.009	0.011	0.013
Predictors for propensity models					
Age at end of index year	66 to 69	0.223	0.067	0.051	0.231
	70 to 74	0.250	0.110	0.093	0.257
	75 to 79	0.202	0.150	0.143	0.205
	80 to 84	0.158	0.202	0.210	0.155
	85 to 89	0.105	0.229	0.248	0.099
	90 to 94	0.048	0.169	0.180	0.042
	95 to 99	0.013	0.063	0.064	0.010
	100+	0.002	0.010	0.010	0.001
Sex	Female	0.587	0.598	0.608	0.587
Race/ethnicity using Research Triangle Institute race categorization	Unknown	0.007	0.002	0.002	0.007
	White, non-Hispanic	0.855	0.883	0.881	0.854
	Black, non-Hispanic	0.064	0.061	0.062	0.065
	Other	0.007	0.004	0.003	0.007
	Asian/Pacific Islander	0.023	0.013	0.013	0.023
	Hispanic	0.041	0.034	0.035	0.041
	American Indian/Alaska Native	0.004	0.003	0.003	0.004
Risk score based on inpatient, outpatient, and physician claims in the previous year	≤25th percentile	0.729	0.250	0.223	0.753
	25th–50th percentile	0.157	0.250	0.290	0.153
	50th–75th percentile	0.077	0.250	0.275	0.068

		All, unweighted	Hospice	Nonhospice, weighted	Nonhospice, unweighted
	75th–90th percentile	0.025	0.150	0.138	0.019
	90th–95th percentile	0.006	0.050	0.039	0.004
	95th–99th percentile	0.004	0.040	0.028	0.003
	>99th percentile	0.001	0.010	0.007	0.001
Risk score type	Community, full dual	0.072	0.096	0.102	0.071
	Community, nondual	0.831	0.707	0.693	0.837
	Community, partial dual	0.027	0.027	0.028	0.027
	Institutional	0.032	0.163	0.172	0.026
	New enrollee	0.037	0.006	0.004	0.039
Original reason for entitlement category	OASDI	0.917	0.905	0.900	0.918
	Disability	0.083	0.095	0.100	0.083
Highest severity of illness from acute stay APR-DRG in the previous three quarters	No acute stays	0.840	0.510	0.465	0.857
	1 (least severe)	0.030	0.028	0.030	0.030
	2	0.068	0.155	0.164	0.064
	3	0.050	0.237	0.221	0.041
	4 (most severe)	0.012	0.070	0.062	0.009
Highest risk of mortality from APR-DRG in the previous three quarters	No acute stays	0.840	0.510	0.465	0.857
	1 (least risk)	0.045	0.026	0.028	0.046
	2	0.060	0.164	0.178	0.054
	3	0.044	0.229	0.209	0.035
	4 (most risk)	0.012	0.070	0.062	0.009
Number of ER visits in the previous year	No ER visits	0.717	0.366	0.362	0.734
	1 to 2	0.225	0.384	0.403	0.217
	3 to 7	0.054	0.223	0.212	0.045
	8+	0.004	0.027	0.023	0.003
Number of ICU/CCU days, previous two quarters	No ICU/CCU days	0.953	0.806	0.819	0.960
	1 to 2	0.018	0.053	0.051	0.016
	3 to 5	0.015	0.063	0.060	0.013
	6 to 10	0.009	0.045	0.041	0.007
	11 to 20	0.004	0.024	0.021	0.003
	21+	0.001	0.010	0.008	0.001
Part A spending, previous quarter	None	0.894	0.554	0.572	0.911
	<25th percentile	0.037	0.112	0.121	0.033
	25th–50th percentile	0.028	0.111	0.109	0.024
	50th–75th percentile	0.024	0.111	0.103	0.019
	75th–90th percentile	0.011	0.067	0.057	0.008
	90th–95th percentile	0.003	0.022	0.019	0.002
	95th–99th percentile	0.003	0.018	0.015	0.002
	≥99th percentile	0.001	0.005	0.004	0.001
Part A spending, second previous quarter	None	0.904	0.682	0.681	0.916
	<25th percentile	0.027	0.079	0.083	0.025
	25th–50th percentile	0.025	0.080	0.080	0.022
	50th–75th percentile	0.024	0.080	0.078	0.021
	75th–90th percentile	0.012	0.048	0.046	0.010
	90th–95th percentile	0.004	0.016	0.016	0.003
	95th–99th percentile	0.003	0.013	0.013	0.003
	≥99th percentile	0.001	0.003	0.003	0.001
Part B institutional spending, previous 2 quarters	None	0.361	0.203	0.203	0.369
	<25th percentile	0.457	0.392	0.403	0.460
	25th–50th percentile	0.107	0.188	0.191	0.103
	50th–75th percentile	0.047	0.116	0.113	0.043
	75th–90th percentile	0.017	0.053	0.049	0.015
	90th–95th percentile	0.006	0.019	0.017	0.005

		All, unweighted	Hospice	Nonhospice, weighted	Nonhospice, unweighted
	95th–99th percentile	0.006	0.023	0.020	0.005
	≥99th percentile	0.001	0.007	0.005	0.001
Part B noninstitutional spending, previous two quarters	None	0.027	0.011	0.011	0.028
	<25th percentile	0.524	0.247	0.250	0.538
	25th–50th percentile	0.251	0.247	0.265	0.251
	50th–75th percentile	0.136	0.247	0.253	0.130
	75th–90th percentile	0.044	0.148	0.139	0.039
	90th–95th percentile	0.011	0.049	0.042	0.009
	95th–99th percentile	0.007	0.040	0.031	0.005
	≥99th percentile	0.001	0.010	0.008	0.001
Part D spending, previous two quarters	None	0.526	0.418	0.406	0.531
	<25th percentile	0.182	0.146	0.144	0.184
	25th–50th percentile	0.129	0.146	0.147	0.128
	50th–75th percentile	0.098	0.146	0.150	0.095
	75th–90th percentile	0.042	0.087	0.092	0.040
	90th–95th percentile	0.012	0.029	0.031	0.011
	95th–99th percentile	0.010	0.023	0.024	0.009
	≥99th percentile	0.002	0.006	0.006	0.002
Alzheimer's Alzheimer's/dementia indicator from MBSF	Alzheimer's	0.056	0.260	0.271	0.046
	Alzheimer's or dementia	0.125	0.497	0.523	0.107
Cancer from inpatient, outpatient, and physician claims over the previous four quarters	Secondary malignancy	0.016	0.125	0.093	0.011
	Respiratory	0.012	0.072	0.057	0.009
	Pancreatic	0.002	0.015	0.009	0.001
	Nervous system or brain	0.001	0.010	0.007	0.001
	Malignant, unknown site	0.004	0.033	0.022	0.003
	Liver or bile duct	0.001	0.012	0.008	0.001
	Esophagus	0.001	0.007	0.005	0.001
	Stomach	0.002	0.008	0.006	0.001
	Ovary	0.002	0.009	0.008	0.002
	Other	0.268	0.371	0.353	0.262
Nonhierarchical HCC groupings based on inpatient, outpatient, and physician claims from the previous four quarters (PACE/ESRD groupings)	Alcohol or drug disease	0.011	0.026	0.026	0.010
	Cancer	0.138	0.297	0.272	0.130
	Cardiac and vascular	0.347	0.672	0.691	0.330
	Complications of device or graft	0.014	0.041	0.038	0.013
	Eye disorders	0.028	0.040	0.042	0.028
	GI and liver	0.038	0.111	0.105	0.034
	Head and spine	0.039	0.137	0.133	0.034
	Hematologic and immunologic diseases	0.051	0.147	0.137	0.046
	HIV/AIDS	0.001	0.001	0.001	0.001
	Mental illness	0.046	0.107	0.109	0.043
	Metabolic endocrine	0.290	0.413	0.407	0.283
	Neuro excluding stroke	0.047	0.130	0.133	0.043
	Obesity	0.026	0.028	0.028	0.026
	Orthopedic	0.066	0.090	0.091	0.065
	Renal	0.054	0.213	0.204	0.046
	Respirator dependence	0.002	0.008	0.008	0.002

		All, unweighted	Hospice	Nonhospice, weighted	Nonhospice, unweighted
	Respiratory	0.151	0.353	0.353	0.141
	Septicemia and other systemic infections	0.024	0.116	0.106	0.020
	Skin disorders	0.030	0.102	0.100	0.027
	Stroke	0.044	0.132	0.133	0.040
	Transplant	0.002	0.003	0.003	0.002
Number of body systems, 5+	5 or more	0.049	0.238	0.219	0.039
Index quarter	2008 Q1	0.170	0.161	0.158	0.170
	2008 Q3	0.160	0.143	0.139	0.161
	2013 Q1	0.169	0.186	0.188	0.169
	2013 Q3	0.160	0.157	0.156	0.160
	2018 Q1	0.175	0.191	0.195	0.174
	2018 Q3	0.166	0.163	0.164	0.166
N		18,924,700	901,837	18,022,863	18,022,863
Weighted N		—	901,837	901,837	—

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: APR-DRG = all patient refined diagnosis-related group; ER = emergency room; ESRD = end-stage renal disease; GI = gastrointestinal; HCC = hierarchical condition category; ICU/CCU = intensive care unit/coronary care unit; MBSF = Master Beneficiary Summary File; OASDI = Old-Age, Survivors, and Disability Insurance; PACE = Program of All-Inclusive Care for the Elderly; Q1 = first quarter; Q3 = third quarter. Fee-for-service beneficiaries ages 66 and older. Em dashes indicate values that are not applicable.

Appendix C: Modeling Potential Heterogeneous Effects of Hospice Across Three Condition Subgroups

With evidence that the complier population differs from the overall weighted population, heterogeneous treatment effects of hospice could lead to LATE estimates that are not indicative of the ATE for the full sample, which is our primary interest. Therefore, we turn to estimating effects of hospice separately for three mutually exclusive subgroups: those diagnosed with cancer, those diagnosed with ADRD but not cancer, and those with all other diagnoses. The weighted population shares for the three groups are 40.7 percent, 36.1 percent, and 23.2 percent, respectively. We conducted this analysis less to produce subgroup-specific estimates of the effects of hospice entry than to evaluate what potential heterogeneity in hospice effects means for our overall estimated hospice effects.

Table C.1 shows first-stage model results that estimate the effects of distance to hospice on hospice entry separately by subgroup. Longer distance is associated with lower hospice use in each model, significantly so at all distance bands for the ADRD and all other conditions subgroups and at distances beyond 40 miles for cancer. F-statistics of the instruments range from 10.0 for cancer to 14.5 for ADRD. The higher F-statistic for the ADRD group is consistent with the complier analysis showing cases with Alzheimer's or dementia are more likely to be distance compliers.

APPENDIX TABLE C.1

First-Stage Model Coefficients of Distance to Hospice on Hospice Entry, by Subgroup (cancer, Alzheimer's and related dementias, and all others)

	Cancer				ADRD (no cancer)				All other conditions (no cancer or ADRD)			
	Estimate	SE	T-stat	P-value	Estimate	SE	T-stat	P-value	Estimate	SE	T-stat	P-value
20–30 miles	-0.0030	0.0030	-0.99	0.321	-0.0142	0.0031	-4.52	0.000	-0.0091	0.0031	-2.94	0.003
30–40 miles	-0.0102	0.0054	-1.90	0.058	-0.0288	0.0065	-4.40	0.000	-0.0156	0.0056	-2.77	0.006
>40 miles	-0.0452	0.0083	-5.46	0.000	-0.0471	0.0100	-4.73	0.000	-0.0530	0.0090	-5.92	0.000
N				5,216,924				1,812,535				11,895,237
Weighted N				366,756				325,781				209,300
Weighted share				0.407				0.361				0.232
F-statistic of instruments				10.0				14.5				13.5
P-value				0.000				0.000				0.000

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: ADRD = Alzheimer's disease and related dementias; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. The 18-month observation window includes the index quarter. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights so that nonhospice comparison group cases match hospice cases on covariates. Standard errors are clustered by three-digit zip code. Weighted share is the subgroup share of the weighted hospice-entrant total.

Table C.2 reports results for our base model applied separately to each subgroup. LS estimates of the effect of hospice entry indicate statistically significant negative effects on total spending for cancer (-$19,244) and ADRD without cancer (-$1,750). For the third group of all others, the LS estimate is significant and positive ($2,485). The 2SLS estimates are negative for all three subgroups: marginally significant for cancer (-$23,277, $p < 0.10$), statistically insignificant for ADRD (-$12,690), and significant for the all-other conditions group (-$43,911).

The point estimates on total spending are highly variable across subgroups, overall and by estimation method. The 2SLS estimates remain imprecise, with standard errors of roughly $9,000 to $12,000, and reach conventional significance only for the all-other conditions group.

For cancer, we cannot reject that hospice entry is exogenous in the LS model ($p = 0.858$), and the LS estimate suggests substantially lower spending. The 2SLS estimate is larger but only marginally significant, and the weak-instrument-robust AR test does not reject the null of no effect ($p = 0.225$). For ADRD, the LS result points to more modestly lower spending, and the endogeneity test does not reject exogeneity ($p = 0.221$) for this subgroup. Although the 2SLS point estimate implies a larger reduction in spending than the LS estimate (-$12,690 versus -$1,750), it is not statistically distinguishable from zero. For the group including all other diagnoses, the endogeneity test is statistically significant. Deferring then to the 2SLS estimate, the finding suggests significantly lower spending from hospice. The point estimate is very large (-$43,911) relative to the mean cost (about $32,000) with a standard error of $12,428, so the 2SLS estimate is imprecise. However, because we reject exogeneity of hospice entry, we cannot rely on the more precise LS estimate. The overidentification test is not rejected in any subgroup ($p = 0.387$, 0.926, and 0.138), which is consistent with the distance instruments being validly excluded from the spending equation.

By recombining the separate estimates into an overall weighted-average estimate for each outcome, we can see whether modeling the heterogeneity as we do here leads to a meaningfully different overall estimate. Recombining the subgroup estimates using the weighted shares reported at the bottom of table C.2 implies an overall LS effect on total spending of -$7,888 and an overall 2SLS effect of -$24,242. Each is close to the pooled estimate from the otherwise similar model estimated on the full sample in table 6, so this recombination recovers essentially the same overall effect.

APPENDIX TABLE C.2

Effect of Hospice Entry on Total Medicare Spending by Subgroup (cancer, Alzheimer's and related dementias, and all others) with Weights from Base Propensity Score Model Estimated Separately by Condition

	Cancer		ADRD (no cancer)		All other conditions (no cancer or ADRD)	
LS estimates	-$19,244	***	-$1,750	***	$2,485	***
SE	($268)		($227)		($194)	
R-squared	0.107		0.098		0.090	
2SLS estimates	-$23,277	*	-$12,690		-$43,911	***
SE	($12,347)		($9,104)		($12,428)	
AR test (p-value)	0.225		0.542		0.000	
Endogeneity test (p-value)	0.858		0.221		0.000	
Overidentification test (p-value)	0.387		0.926		0.138	
Dependent variable means						
Hospice	$31,715		$33,111		$35,057	
Nonhospice, weighted	$50,712		$34,844		$32,153	
N	5,216,924		1,812,535		11,895,237	
Weighted N	366,756		325,781		209,300	
Weighted share	0.407		0.361		0.232	

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: AR = Anderson-Rubin; ADRD = Alzheimer's disease and related dementias; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending is measured over an 18-month observation window including the index quarter and is adjusted to 2018 dollars. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights, estimated separately by condition, so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by 3-digit zip code, are in parentheses. Weighted share is the subgroup share of the weighted hospice-entrant total. *** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

Table C.3 extends the analyses that condition on days alive to the three condition subgroups presented in the panels. The first column repeats the total spending estimates from table C.2 for reference. The second column estimates the effect of hospice entry on days alive in the observation window separately by subgroup. The third column again examines total spending as the outcome but uses weights that condition on days alive and includes days alive as a regression control, so that estimated effects of hospice on spending are net of differences in survival time. As in the earlier analysis, the estimates in the third column impose the maintained assumption that hospice does not affect how long people live within the observation window. Under that assumption, these are estimates of the effect of hospice on spending. If hospice instead has a causal effect on survival time, the estimates in the third column may be interpreted as an accounting decomposition rather than as a causal effect, because conditioning on days alive conditions on a variable that hospice itself affects. As a decomposition, it is still informative. Comparing beneficiaries with similar survival time in the observation window, the spending difference for cancer, for example, is a small fraction of the difference estimated without survival time held fixed. Nearly all the spending difference in the first column is therefore associated with a shorter observation period rather than with lower spending at a given survival time.

The second column shows estimated effects of hospice entry on days alive that are large, negative, and statistically significant in every subgroup and under both estimation methods. LS estimates are -338 days for cancer, -295 days for ADRD, and -365 days for all other conditions. The 2SLS estimates are smaller in magnitude for cancer (-198 days) and ADRD (-141 days) and nearly unchanged for all other conditions (-359 days), showing the same pattern of moderation seen in the full-sample results in table 9 for two cases but not the third. Exogeneity of hospice is rejected for cancer ($p = 0.020$) and for ADRD ($p < 0.001$), so the 2SLS estimates are preferred there, while for all other conditions the test does not reject ($p = 0.916$) and the LS and 2SLS estimates are close in magnitude (-365 and -359 days). In no condition subgroup does the overidentification test reject the validity of the distance instruments in the days alive equations ($p = 0.144$, 0.302, and 0.448, respectively), consistent with the instruments being validly excluded from the survival equation.

Holding survival time fixed changes the LS spending coefficients substantially. The 2SLS estimates, however, are uninformative with days alive controlled, because the F-statistic for the first stage is 4.1. The first stage is too weak in this specification to support the 2SLS estimates, so we rely more on the LS estimates, which show a significant negative effect (-$1,274).

For ADRD cases, the pattern is the opposite. The endogeneity test suggests the LS results are endogenous (p = 0.029), and the first-stage F-test is 14.4. Together, these indicate we should prefer the 2SLS estimate, which is not statistically significant (point estimate = -$5,803).

For the subgroup with all other conditions, the 2SLS estimate with days controlled is preferred, because the LS estimate shows evidence of endogeneity. The 2SLS model shows that hospice entry reduces spending by $18,448, but this estimate is only marginally significant using conventional standard errors, and it is not significant using the weak-instrument-robust AR test.

Thus, we find that with days controlled, none of the condition subgroups' 2SLS estimates are statistically significant. In the two cases of positive LS estimates (ADRD and all other conditions), 2SLS point estimates are negative, and exogeneity of hospice was rejected. Overall, we can refer to the pooled 2SLS estimate with days controlled in table 13 of -$6,435, which was not statistically significant.

APPENDIX TABLE C.3

Effect of Hospice Entry on Total Medicare Spending and Days Alive in Observation Window by Subgroup (cancer, Alzheimer's and related dementias, and all others), with Base Model Weights and Weights That Condition on Days Alive in Observation Window

	Total spending		Days alive in observation window		Total spending, with days alive in observation window controlled	
Cancer subgroup						
LS estimates	-$19,244	***	-338	***	-$1,274	***
SE	($268)		(1)		($201)	
R-squared	0.107		0.594		0.238	
2SLS estimates	-$23,277	*	-198	***	$4,538	
SE	($12,347)		(56)		($12,516)	
AR test (p-value)	0.225		0.018		0.566	
F-statistic of instruments	10.0		10.0		4.1	
Endogeneity test (p-value)	0.858		0.020		0.675	
Overidentification test (p-value)	0.387		0.144		0.408	
Dependent variable means						
Hospice	$31,715		118		$31,715	
Nonhospice, weighted	$50,712		457		$33,206	
N	5,216,924		5,216,924		5,216,915	
ADRD (no cancer) subgroup						

	Total spending		Days alive in observation window		Total spending, with days alive in observation window controlled	
LS estimates	-$1,750	***	-295	***	$8,810	***
SE	($227)		(1)		($149)	
R-squared	0.098		0.459		0.353	
2SLS estimates	-$12,690		-141	***	-$5,803	
SE	($9,104)		(44)		($6,756)	
AR test (p-value)	0.542		0.032		0.637	
F-statistic of instruments	14.5		14.5		14.4	
Endogeneity test (p-value)	0.221		0.000		0.029	
Overidentification test (p-value)	0.926		0.302		0.630	
Dependent variable means						
Hospice	$33,111		170		$33,111	
Nonhospice, weighted	$34,844		465		$24,673	
N	1,812,535		1,812,535		1,812,532	
All other conditions (no cancer or ADRD) subgroup						
LS estimates	$2,485	***	-365	***	$5,819	***
SE	($194)		(1)		($171)	
R-squared	0.090		0.651		0.221	
2SLS estimates	-$43,911	***	-359	***	-$18,448	*
SE	($12,428)		(38)		($10,317)	
AR test (p-value)	0.000		0.000		0.290	
F-statistic of instruments	13.5		13.5		10.8	
Endogeneity test (p-value)	0.000		0.916		0.014	
Overidentification test (p-value)	0.138		0.448		0.952	
Dependent variable means						
Hospice	$35,057		139		$35,057	
Nonhospice, weighted	$32,153		506		$29,449	
N	11,895,237		11,895,237		11,895,226	

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.
Notes: AR = Anderson-Rubin; ADRD = Alzheimer's disease and related dementias; LS = least squares; 2SLS = two-stage least squares; SE = standard error. Analytic sample includes fee-for-service beneficiaries ages 66 and older; hospice entry indicates first hospice use in the index quarter—the first or third quarter of 2008, 2013, or 2018. Medicare spending is measured over an 18-month observation window including the index quarter and is adjusted to 2018 dollars. Model includes covariates as described in the methods section, index quarter fixed effects, and three-digit zip code fixed effects. Regressions are weighted with propensity score-based weights, estimated separately by condition, so that nonhospice comparison group cases match hospice cases on covariates. Standard errors, clustered by 3-digit zip code, are in parentheses. In the third column, the weights also condition on days alive, which is measured over the same 18-month window, and the model controls for days alive.
*** Estimate is statistically significant at the $p < 0.01$ level, ** at the $p < 0.05$ level, * at the $p < 0.10$ level, using two-sided tests.

Appendix D. Exploratory Analyses Using a Leave-One-Out IV Approach

As an alternative source of identifying variation, we conducted preliminary analyses that explore the potential for using a leave-one-out IV in the individual-level analyses. The alternative IV uses variation in zip-code hospice use rates to identify effects of hospice based on local use patterns. To sever the link between individual-level hospice use and endogenous individual-level factors, the zip code mean is computed by quarter for each case, excluding its own data from the calculation. Preliminary estimates using this approach are reported in appendix table D.1.

APPENDIX TABLE D.1

Preliminary 2SLS Model Results with a Leave-One-Out Hospice Use Instrument

Reference table in report	N	Dependent variable	Hospice coefficient		Standard error	Coefficient p-value	Endogeneity test p-value	F-test for instruments
Table 6	18,924,700	Total Medicare spending	-7,611	**	3,150	0.0159	0.9112	250.88
Table 8	18,924,700	Part A Medicare spending	3,162		2,520	0.2099	0.3973	250.88
Table 8	18,924,700	Acute care	-10,794	***	1,507	0.0000	0.0000	250.88
Table 8	18,924,700	Other inpatient	419	*	250	0.0949	0.0332	250.88
Table 8	18,924,700	Post-acute care	-10,875	***	1,236	0.0000	0.0000	250.88
Table 8	18,924,700	Home health	-2,322	***	614	0.0002	0.4827	250.88
Table 8	18,924,700	Long-term care facility	-2,011	***	558	0.0003	0.0011	250.88
Table 8	18,924,700	Skilled nursing facility	-6,924	***	740	0.0000	0.0000	250.88
Table 8	18,924,700	Inpatient rehabilitation facility	383		266	0.1505	0.0038	250.88
Table 8	18,924,700	Hospice	24,412	***	1,338	0.0000	0.0000	250.88
Table 8	18,924,700	Part B Medicare spending	-8,054	***	933	0.0000	0.1821	250.88
Table 8	18,924,700	Part B institutional	-3,819	***	583	0.0000	0.8494	250.88
Table 8	18,924,700	Outpatient	-2,724	***	556	0.0000	0.6508	250.88
Table 8	18,924,700	Part B noninstitutional	-4,235	***	706	0.0000	0.0523	250.88
Table 8	18,924,700	Evaluation and management	-1,377	***	247	0.0000	0.9917	250.88
Table 8	18,924,700	Part D Medicare spending	-2,719	***	555	0.0000	0.3465	250.88
Table 9	18,924,700	Spending per day alive	90	*	54	0.0929	0.0010	250.88
Table 9	18,924,700	Days alive	-240	***	12	0.0000	0.0000	250.88
Table 9	18,924,700	Died	0.52	***	0.03	0.0000	0.0000	250.88
Table C.2	5,216,924	Total Medicare spending, cancer subgroup	-16,334	**	7,863	0.0381	0.7293	73.54
Table C.3	5,216,924	Days alive, cancer subgroup	-206	***	26	0.0000	0.0000	73.54

Table C.2	1,812,535	Total Medicare spending, ADRD subgroup	-4,035		3,702	0.2761	0.5471	184.54
Table C.3	1,812,535	Days alive, ADRD subgroup	-225	***	18	0.0000	0.0001	184.54
Table C.2	11,895,237	Total Medicare spending, other conditions subgroup	724		4,042	0.8578	0.6511	189.84
Table C.3	11,895,237	Days alive, other conditions subgroup	-292	***	18	0.0000	0.0001	189.84

Source: Urban Institute analysis of Medicare claims and Part D event file for 2007–09, 2012–14, and 2017–19.

Notes: *** Estimate is statistically significant at the p < 0.01 level, ** at the p < 0.05 level, * at the p < 0.10 level, using two-sided tests.

Notes

[1] "Hospice," CMS, accessed September 30, 2026, https://www.cms.gov/medicare/payment/fee-for-service-providers/hospice.

[2] "Tax Measure Offers New Benefits for Hospice Care of Terminally Ill," *The New York Times*, September 1, 1982, https://www.nytimes.com/1982/09/01/us/tax-measure-offers-new-benefits-for-hospice-care-of-terminally-ill.html.

[3] "Tax Measure Offers New Benefits for Hospice Care of Terminally Ill."

[4] This measure is inflated because it doesn't net out the manufacturer rebates/pharmacy direct and indirect remuneration fees.

[5] Beneficiaries without Part B often have employer-sponsored coverage with Medicare Part A as a secondary payer, meaning substantial health care utilization would be missing from our dataset. Similarly, the Medicare Advantage Encounter Data does not include payment/spending data for beneficiaries with Part C; see "Encounter Carrier," ResDAC, accessed September 18, 2026, https://resdac.org/cms-data/files/carrier-encounter.

[6] To address changes in zip codes over time, we apply the Bailey and Helmuth crosswalk to obtain zip code clusters that hold over the period 2007 to 2018. We lose a small share of beneficiaries living in zip codes that can be cleanly assigned to a cluster. See Bailey, Amy Kate and Allison Suppan Helmuth(2023) and "National ZIP Code Crosswalk, 1990-2020," DSDR, November 10, 2025, https://doi.org/10.3886/ICPSR39431.v1.

[7] We estimate these models using the ivreghdfe command in Stata, which performs IV estimation with high-dimensional fixed effects. ivreghdfe combines the ivreg2 IV/GMM estimation engine (Baum, Schaffer, and Stillman 2007) with the high-dimensional fixed-effects algorithm of Correia (2016).

[8] With only one distance indicator, or a single continuous (i.e., linear) distance measure, the hospice coefficient would be just-identified in the 2SLS models, and instrument validity (that the distance measure only affects the outcome through its effect on hospice) is a maintained assumption that is untestable in the data.

[9] The same calculation applied within a subgroup gives the share of compliers among beneficiaries who have a given binary characteristic X: Pr(Complier | X) = Pr(Hospice | X and Near) − Pr(Hospice | X and Far). The mean of that characteristic among compliers then follows from the definition of conditional probability and Bayes' rule: Pr(X | Complier) = Pr(X and Complier) / Pr(Complier) = Pr(Complier | X) · Pr(X) / Pr(Complier) (Angrist and Pischke 2009, equation 4.4.8), where Pr(X) is the mean of the characteristic in the full weighted sample, calculated directly.

[10] Weighting zip codes by the number of enrollees does not assure that estimates reflect a national average effect. The IV estimates remain LATEs for complier geographies, which can differ from an average treatment effect if treatment effects are heterogeneous. Weighting does, however, make the estimate an enrollee-weighted average across complier geographies rather than one in which small zip codes count as much as large ones.

[11] We note that the comparison group includes some hospice spending because the comparison group sample includes some beneficiaries who entered hospice two or more quarters after the index quarter.

[12] Total Medicare hospice spending was $19.2 billion over 113.5 million hospice days in 2018, which implies an average payment of about $169 per hospice day (MedPAC 2020)

[13] The effect on per-day spending has a p-value of 0.107 with the AR test, which is known to be overly conservative in overidentified models relative to the preferred CLR test (Keane and Neal 2024). It was not feasible to estimate CLR tests or potentially asymmetric confidence regions from AR or CLR tests given the size of our samples.

[14] With a correctly specified nonlinear first stage, the control-function estimator is asymptotically more efficient than 2SLS, and the generalized residual absorbs correlation between the unobserved determinants of hospice use and those of the outcome. Both gains require the first-stage conditional mean and the link between first- and second-stage unobservables to be correctly specified, which 2SLS does not require.

[15] Neither specification dominates the other. The two fixed-effects levels draw on different identifying variation and therefore imply different complier geographies, and the control-function estimator identifies its effect under first-stage assumptions that 2SLS does not require. The two estimate related but distinct parameters, so we report both. While standard linear 2SLS estimates a local average treatment effect for instrument compliers, the parametric control-function model identifies a structural average treatment effect under its specified assumptions about the generalized residual.

[16] Conventional t-tests on the 2SLS coefficient rely on the estimator being approximately normally distributed. When identification is weak, that approximation can fail. The 2SLS estimator may be skewed and biased toward least squares, and lead to conventional standard errors misstating uncertainty and the test size distorted in either direction. The CLR test (Moreira 2003) conditions on a statistic that captures first-stage strength, which gives correct size regardless of instrument strength. Confidence sets may be asymmetric and, under very weak identification, unbounded. Among weak-instrument-robust tests, CLR also has near-optimal power (Andrews, Moreira, and Stock 2006). Where the CLR p-value differs from the conventional one, the difference can reflect the failure of the normal approximation for the latter.

[17] "Tax Measure Offers New Benefits for Hospice Care of Terminally Ill," *The New York Times*.

[18] We exclude the CBSA for Fairbanks, Alaska, because in the 2SLS estimation, it had an outsized effect on the 2SLS estimates, particularly unweighted. Fairbanks had few decedents and a very large change over time in the distance measure due to the entry of a single hospice into the Fairbanks market (where payments had previously been trending down).

[19] Using data from 2012 and 2013, Hogan (2015) found a statistically significant 51.0 percent increase in the decedent-to-survivor spending ratio ($p = 0.023$).

References

Aldridge, Melissa D., Lauren J. Hunt, Krista L. Harrison, Karen McKendrick, Lihua Li, and R. Sean Morrison. 2023. "Health Care Costs Associated with Hospice Use for People with Dementia in the US." *Health Affairs* 42 (9): 1250–59. https://doi.org/10.1377/hlthaff.2023.00036.

Andrews, Donald W. K., Marcelo J. Moreira, and James H. Stock. 2006. "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression." *Econometrica* 74 (3): 715–52.

Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11: 727–53. https://doi.org/10.1146/annurev-economics-080218-025643.

Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55. https://doi.org/10.2307/2291629.

Ashford, J. R., and R. R. Sowden. 1970. "Multi-Variate Probit Analysis." *Biometrics* 26 (3): 535–46. https://doi.org/10.2307/2529107.

Austin, Peter C. 2009. "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-score Matched Samples." *Statistics in Medicine* 28 (25): 3083–107. https://doi.org/10.1002/sim.3697.

Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. 2007. "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing." *The Stata Journal* 7 (4): 465–506.

Chiburis, Richard C., Jishnu Das, and Michael Lokshin. 2012. "A Practical Comparison of the Bivariate Probit and Linear IV Estimators." *Economics Letters* 117 (3): 762–66. https://doi.org/10.1016/j.econlet.2012.08.037.

CMS (Centers for Medicare & Medicaid Services). 2021. "Geographic Variation in Medicare Cost & Quality." Baltimore, MD: CMS.

———. 2025. *Hospice Monitoring Report*. Baltimore, MD: CMS.

Connor, Stephen R., Bruce Pyenson, Kathryn Fitch, Carol Spence, and Kosuke Iwasaki. 2007. "Comparing Hospice and Nonhospice Patient Survival among Patients Who Die within a Three-Year Window." Journal of Pain and Symptom Management 33 (3): 238–46. https://doi.org/10.1016/j.jpainsymman.2006.10.010.

Correia, Sergio. 2016. "Linear Models with High-Dimensional Fixed Effects: An Efficient Algorithm." Working paper.

Davidson, R., and MacKinnon, J. G. 1993. *Estimation and Inference in Econometrics*. Oxford University Press. (Chapter 7).

Fiala, Mark A., Torie Gettinger, Cara L. Wallace, Ravi Vij, and Tanya M. Wildes. 2020. "Cost Differential Associated with Hospice Use among Older Patients with Multiple Myeloma." Journal of Geriatric Oncology 11 (1): 88–92. https://doi.org/10.1016/j.jgo.2019.06.010.

Gruber, Jonathan, David Howard, Jetson Leder-Luis, and Theodore L. Caputi. 2025. "Dying or Lying? For-Profit Hospices and End of Life Care." American Economic Review 115 (1): 264–94. https://doi.org/10.1257/aer.20230328.

Gutman, Rom, Ehud Karavani, and Yishai Shimoni. 2024. "Improving Inverse Probability Weighting by Post-Calibrating Its Propensity Scores." *Epidemiology 35 (4)*: 473–80. https://doi.org/10.1097/EDE.0000000000001733.

Hirano, Keisuke, and Guido W. Imbens. 2001. "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services and Outcomes Research Methodology* 2 (3): 259–78. https://doi.org/10.1023/A:1020371312283.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): 1161–89.

Hogan, Christopher. 2015. *Spending in the Last Year of Life and the Impact of Hospice on Medicare Outlays*. Washington, DC: MedPAC.

Hung, Peiyin, Sylvia H. Hsu, and Shi-Yi Wang. 2020. "Associations Between End-of-Life Expenditures and Hospice Stay Length Vary by Clinical Condition and Expenditure Duration." *Value in Health* 23 (6): 697–704. https://doi.org/10.1016/j.jval.2020.01.021.

Imbens, Guido W. 2009. "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua 2009." NBER Working Paper 14896. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w14896.

Iyer, Anand S., Christine A. Goodrich, Mark T. Dransfield, Shama S. Alam, Cynthia J. Brown, C. Seth Landefeld, Marie A. Bakitas, and Jeremiah R. Brown. 2020. "End-of-Life Spending and Healthcare Utilization Among Older Adults with Chronic Obstructive Pulmonary Disease." *American Journal of Medicine* 133 (7): 817–824.e1. https://doi.org/10.1016/j.amjmed.2019.11.024.

Kalidindi, Yamini, Jeah Jung, Joel Segel, and Douglas Leslie. 2020. "Impact of Length of Hospice on Spending and Utilization Among Medicare Beneficiaries With Lung Cancer." *American Journal of Hospice and Palliative Care* 37 (11): 918–24. https://doi.org/10.1177/1049909120909304.

Keane, Michael P., and Timothy Neal. 2024. "A Practical Guide to Weak Instruments." *Annual Review of Economics* 16: 185–212. https://doi.org/10.1146/annurev-economics-092123-111021.

Leibowitz, Arleen A., Diane Tan, and Jennifer L. Gildner. 2020. "The Effect of Hospice on End-of-Life Costs for Terminal Medicare Patients With HIV." *Inquiry: A Journal of Medical Care Organization, Provision and Financing* 57: 46958020969381. https://doi.org/10.1177/0046958020969381.

Maddala, G. S., and Lung-Fei Lee. 1976. "Recursive Models with Qualitative Endogenous Variables." *Annals of Economic and Social Measurement* 5 (4): 525–45.

MedPAC (Medicare Payment Advisory Commission). 2020. "Hospice Services." In *March 2020 Report to the Congress: Medicare Payment Policy*. Washington, DC: MedPAC.

MedPAC (Medicare Payment Advisory Commission). 2026. *March 2026 Report to the Congress: Medicare Payment Policy*. Washington, DC: MedPAC.

Moreira, Marcelo J. 2003. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica* 71 (4): 1027–48. https://doi.org/10.1111/1468-0262.00438.

NORC (NORC at the University of Chicago). 2023. "Final Report: Value of Hospice in Medicare." Chicago: NORC at the University of Chicago.

Obermeyer, Ziad, Maggie Makar, Samer Abujaber, Francesca Dominici, Susan Block, and David M. Cutler. 2014. "Association Between the Medicare Hospice Benefit and Health Care Utilization and Costs for Patients With Poor-Prognosis Cancer." JAMA 312 (18): 1888–96. https://doi.org/10.1001/jama.2014.14950.

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11: 619–32.

Pyenson, Bruce, Stephen Connor, Kathryn Fitch, and Barry Kinzbrunner. 2004. "Medicare Cost in Matched Hospice and Non-Hospice Cohorts." *Journal of Pain and Symptom Management* 28 (3): 200–210. https://doi.org/10.1016/j.jpainsymman.2004.05.003.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39 (1): 33–38. https://doi.org/10.2307/2683903.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–86.

Storick, Virginia, Aoife O'Herlihy, Sarah Abdelhafeez, Rakesh Ahmed, and Peter May. 2019. "Improving Palliative Care with Machine Learning and Routine Data: A Rapid Review." *HRB Open Research* 2 (August): 13. https://doi.org/10.12688/hrbopenres.12923.2.

Temel, Jennifer S., Joseph A. Greer, Alona Muzikansky, Emily R. Gallagher, Sonal Admane, Vicki A. Jackson, Constance M. Dahlin, Craig D. Blinderman, Juliet Jacobsen, William F. Pirl, J. Andrew Billings, and Thomas J. Lynch. 2010. "Early Palliative Care for Patients with Metastatic Non-Small-Cell Lung Cancer." The New England Journal of Medicine 363 (8): 733–42. https://doi.org/10.1056/NEJMoa1000678.

Wooldridge, Jeffrey M. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources* 50 (2): 420–45.

Zuckerman, Rachael B., Sally C. Stearns, and Steven H. Sheingold. 2016. "Hospice Use, Hospitalization, and Medicare Spending at the End of Life." *The Journals of Gerontology*: Series B 71 (3): 569–80. https://doi.org/10.1093/geronb/gbv109.

About the Authors

Bowen Garrett is an economist and senior fellow at the Urban Institute. His research focuses on health policy and health care reform, including health insurance and labor markets, Medicare's prospective payment systems, and Medicare financing. He leads the development of the Urban Institute's Medicare policy simulation model (MCARE-SIM). Garrett received his PhD in economics from Columbia University in 1996 and was a postdoctoral research fellow in the Robert Wood Johnson Foundation's Scholars in Health Policy Research Program at the University of California, Berkeley, from 1996 to 1998.

Doug Wissoker is an economist and nonresident fellow in the Health, Income, and Benefits Policy Division at the Urban Institute. In Urban's Statistical Methods Group, he collaborated with researchers on a wide range of topics, including Medicare payment policy, welfare reform, audit studies of discrimination in housing and employment, and survey nonresponse. He has worked extensively with MedPAC staff on potential reforms to traditional Medicare payment systems for post-acute care. Wissoker has a PhD in economics from the University of Wisconsin.

Laura Skopec is a senior research associate in the Health Policy Division at the Urban Institute, where her research focuses on health insurance coverage, health care access, and health equity. Before joining Urban, she worked on Affordable Care Act implementation at the Office of the Assistant Secretary for Planning and Evaluation in the US Department of Health and Human Services, and on transparency in health insurance and health care at the American Cancer Society Cancer Action Network. Skopec holds a BS in biopsychology and cognitive science from the University of Michigan and an MS in public policy and management from Carnegie Mellon University.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org