Summer 8-10-2017

Direct Effects of Bundled Payments for Care Improvement Initiative: A County-Level Approach

Konstantinos Panitsas Mr

Direct Effects of Bundled Payments for Care Improvement Initiative: A County-Level Approach

by

Konstantinos Panitsas

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Economics
Hunter College, City University of New York

2017

Thesis sponsor:

August 4, 2017	Partha Deb
Date	First Reader
August 4, 2017	Matthew Baker
Date	Second Reader

Dedication

This thesis is dedicated to my family, primarily to my loving parents, Panagoula and Ioannis Panitsas. Their constant support, coupled with their words of encouragement, from the very beginning of my thesis until its completion, were the greatest and most essential sources of inspiration and drove me to achieve my goal. I would like to thank them wholeheartedly for everything they have done for me and let them know that I am grateful for having them in my life. I also dedicate this thesis to my uncle, Antonios Spiliotopoulos and aunt, Dr. Stavroula Kousteni, who were always next to me, every time I needed them, ready to provide me with useful pieces of advise, encourage me and support the effort I put in order to complete my work. Lastly, I want to thank my friend and companion in life, Elisavet Vlachou. Even though she was not physically present during the time I worked on my thesis, her warm, supportive words were invaluable.

Acknowledgements

It is with great pleasure and honor that I acknowledge my deepest and most sincere thanks and gratitude to professor Partha Deb for his suggestion of this topic and his insightful supervision. He has been supportive from the very beginning of this journey. The inexhaustible source of his scientific knowledge, the clarity of his comments and his eagerness to continually interact with me, were key contributors towards the completion of my diploma thesis. Lastly, I would like to thank professor Matthew Baker for his insightful comments and suggestions.

Contents

1 Introduction

Plethora of scientific papers and articles from the official literature have examined the vulnerabilities of health care system in the United States, e.g. [Grubaugh and Santerre, 1994], [Spithoven, 2009] and [Lorenzoni et al., 2014]. Based on a Bloomberg[1]report, in 2015, annual health care per capita expenditures in the US were approximately $9,146, with only two countries surpassing this number; Norway and Switzerland with $9,715 and $9,276 per capita respectively. The author, by assigning a health-care efficiency score to 55 high-spending OECD countries, used three weighted metrics; life expectancy, health-care costs per-capita and medical costs as a percentage of GDP. The United States was placed 50$^{\text{th}}$ out of 55 countries. But the Bloomberg report is not the only one addressing the imbalance between the quality of health care services provided in the United States and the corresponding cost at which they are offered. [Garber and Skinner, 2008], applied cross-country comparisons of administrative expenses between the US and a group of peer countries and tried to explore these disparities by addressing the issue of *productive* and *allocative* efficiency. The authors argue that this inefficiency is primarily due to "a predominantly fee-for-service system of reimbursement, coupled with few supply-side constraints fueling the rapid adoption and diffusion of technological advances".

It is widely accepted that a huge variety of wasteful practices should be considered as the major part of the aforementioned problem. [Berwick and Hackbarth, 2012], present six factors for the total cost of waste, which concern not only privately insured US citizens but also those insured by Medicare and Medicaid; (*i*) failure of care delivery and (*ii*) coordination, (*iii*) overtreatment, (*iv*) administrative complexity, (*v*) pricing failure and lastly (*vi*) fraud and abuse. The authors

[1]Source: Moffat, A., R., *"Health-Care Check-Up: Whose System Is Least Efficient?"*, Bloomberg, November 30, 2015

found that more than 20% of the total healthcare expenditures is due to the combined waste from these estimates. It is a clear conclusion that efficient treatment methods are paramount in order to substantially lower costs.

Future trends regarding high-value care are at least ominous. Contemporary findings [CMS, 2016a] suggest that, by 2025, expenditures related to health care are projected to grow at an average of 5.8 percent each year and the health care cost as a percent of GDP is estimated to rise from 17.5 percent in 2014 to 20.1 percent in 2025. At the same time, both state and local governments are estimated to increase their total medical spending by 2 percentage points; from 45 percent in 2014 to 47 percent in 2025. Additionally, from 2017 to 2019, health care spending growth will accelerate and average 5.7 percent, with this number reaching the alarming 6 percent in the next five years. Thorough analysis has been conducted regarding payers' demographic and social characteristics, as well as the type of their health coverage. Both the privately insured and Medicaid (and Medicare) patients, are projected to experience increased medical expenditures; especially in terms of Out-Of-Pocket expenses, which is defined as the amount of money insurance does not cover and must be paid by an individual on its own. Prescription drugs, hospital spending and expenditures on physician and clinical services are all anticipated to increase in the long-term as well. The above arguments justify why policymakers, scientists and researchers have raised concerns about the continuous increase of medical expenditures in the United States and its consequent impact on the domestic economy.

1.1 Research Hypotheses

This paper studies the effects of a cost reducing initiative undertaken by CMS and has two objectives. The first one is to examine whether in areas where BPCI policies were implemented, individuals reported on average lower out of pocket and total medical expenditures. Secondly, it investigates whether the impact of this policy was even larger among the most vulnerable population groups, such as the economically disadvantaged individuals (people reporting income less than $25,000), the elderly (people older than 64 years old) and those who meet both of those criteria; the elderly poor. Individuals who belong to these population groups are of particular interest. Low-income individuals or elderly people are more likely to experience health problems that require medical services while at the same time are less able to afford needed care due to their low-income. In terms of cost we expect people living in areas where and when BPCI policies were implemented to report lower health care costs.

In order to examine whether bundled payment policies had a positive impact on vulnerable population groups, data from the Annual Social and Economic Supplement of the Current Population Survey and the Center for Medicare and Medicaid Services were used. To evaluate the impact health care policy reforms had, a Difference-in-Differences (DiD) estimation method is used, popular in the scientific literature when we want to evaluate the effect of policy changes [Athey and Imbens, 2006]. Difference-in-differences estimation consists of identifying a specific treatment, by comparing the difference in outcomes before and after the intervention for groups affected by the intervention to the same difference for unaffected groups [Bertrand et al., 2004]. Here, the "treatment" refers to policies being run and tested by CMS. The treatment group consists of people living in counties where such policies were implemented. People living in counties where BPCI policies were not tested are the control group.

The rest of this paper is structured in the following way. In subsection 1.2 the BPCI initiative is described in detail. Section 2 is a summary of previous findings regarding the impact that bundled payment policies have on cost reduction, the difficulties which exist when trying to estimate these effects and an analytic description of reasons that justify the need to reassess the current payment model. Moving on, in section 3 I present the datasets I rely on. Section 4 and 5 detail the tools that are used and the obtained results respectively. In Section 6 and 7, conclusions and limitation of this research are presented. All figures, tables and regression results obtained by my analysis, along with descriptive statistics for the data used, are presented in the Appendix.

1.2 The BPCI Initiative

In 2013, the Center for Medicare and Medicaid Innovation Services (CMS), launched the Bundled Payment for Care Improvement (BPCI) program. The objective of this initiative is to examine whether *bundled payments* can actually reduce Medicare's cost and whether they are able to improve the quality of care that is delivered.

Bundled payment is defined as a per-episode reimbursement model under which different physicians, hospitals and post-acute care centers receive a predetermined payment amount designed to cover the expenses for all services provided for an episode of treatment for a specific medical condition [Antonova et al., 2015]. Payment can be made before, during, or after an episode of care [Friedberg et al., 2015]. Thus, BPCI model aims to extend the principle of prospective payment to a package of services that spans multiple providers and extends for longer periods of time. The amount of services delivered during an episode of care is defined as a bundle and varies by model [CMS, 2016b]. By the time this study was conducted, BPCI Model 1 (Retrospective Acute Care Hospital Stay) is excluded from our analysis since, data regarding this model were not available from CMS. The remaining three models of this initiative are described below:

- The **Model 2** bundle includes the triggering hospital stay and all professional and post-discharge services delivered within a 30, 60 or 90 day period.

Individual providers are paid on a fee-for-service basis with retrospective reconciliation against a predetermined target price.

- The **Model 3** bundle starts once a patient is admitted to an episode-initiating post-acute provider following hospitalization, including all services within the designated period of 30, 60 or 90 days. Unlike the previous model, Model 3 includes only the post-discharge services and any readmission within the designated period of time.
- Lastly, the **Model 4** bundle includes the anchor hospital stay and professional services along with any readmissions that may occur within 30 days of discharge. Contrary to Models 2 and 3, awardees are paid a prospectively determined amount and they, in turn, pay the providers involved in an episode.

We also extend the BPCI terminology list by introducing two more key terms. An Episode Initiator (EI), is defined as the participating hospital where a BPCI Model is implemented. An Episode of Care (EoC), is triggered by an inpatient hospitalization for one of the 48 BPCI clinical episodes, defined by patient's Medicare Severity Diagnosis Related Group (MS-DRG) or it begins upon patient's admission to a post-acute care setting.

Models 2, 3 and 4 have some key differences that are worth noting. The most important one is in episode-based payment methods. For both Models 2 and 3, payment is retrospective. Medicare makes fee-for-service (FFS) payments to providers, practitioners and suppliers who offer medical services to beneficiaries. Total payments for a beneficiary's episode is reconciled against a bundled payment predetermined by CMS. For Models 1, 2 and 3, CMS firstly relies on patients' historical spending in each organization that participates. It then sets a target price for each EoC, paying at Medicare's FFS rates. Thereafter, it checks whether, for each episode, actual spending exceeds this target price; if it does not, CMS provides organizations with additional payments and organizations which exceed this price, return the excess amount back to the organization [Mechanic and Tompkins, 2012]. On the other hand, payment methods under Model 4 scheme are prospective; CMS, instead of proceeding to an Inpatient Prospective Payment System, offers a fixed, predetermined bundled payment during a hospitalization and possible readmissions within thirty days.

BPCI policy is implemented in two phases. During the first phase, also called preparation period, which ended on August 5^{th}, 2014, 2,368 potential Awardees (entities which assume financial liabilities for the episode spending) submitted applications to CMS. Subsequently, phase I participants move on to the second phase of the program, also called "the risk-bearing" phase, during which, they have to complete execution of an agreement with CMS. From phase I, 1,306 participants made it to phase II.

2 Literature Review

The voluntary bundled-payments program was officially launched in 2013. Under this initiative, hospitals, Post Acute providers, physicians and other health-related professional organizations assume risk for total spending relative to a target price for up to 48 episodes of care, which account for 70% of total Medicare spending.

Given that medical expenditures rise at a fast pace, there is no need to question why policymakers focus on discovering methods which aim to reduce costs and simultaneously deliver better quality health care services. The bundled payment approach is simple. Under the assumption that the reduction of volume of services delivered leads to reduced spending, multiple providers are reimbursed a single sum of money related to an episode of care. In mathematical notation, as [Desisle, 2013] demonstrates, a generalized bundled payment formula can be depicted as shown in Eq. (1) below:

$$\text{Bundled Payment} = (A - \mathscr{X}) + \sum_{i=1}^{n} B_i, \tag{1}$$

where, we let A denote hospital payments (Part A), B is the total number of physician payments (Part B) and $\mathscr{X}$ is the "reduced negotiated reimbursement price" driven by participation in the bundled payment model. On the contrary, under the current Medicare *fee-for-service* payment model, health care providers are reimbursed for each service they provide.

2.1 The Downside of FFS System

Several studies have highlighted the falacies in FFS system, since it leads to higher costs and sometimes also results in sub-optimal quality. Yet, it has been the dominant form of reimbursement in the US for decades. This set of vulnerabilities are worthy of thorough examination.

The most difficult barrier that policymakers have to face is handling an extensive list of codes which is linked to the type of services provided by physicians or nursing facilities. This makes it almost unmanageable, due to the depth of its detail. This practice complicates their effort to measure the *unit of service* provided each time, creating substantial administrative costs related to coding [Chernew, 2011]. However, even when this problem is overcome, there is a lack of incentive for taking initiatives and implementing innovative policies on behalf of providers. This argument relies on the fact that less provided services imply lesser earnings.

[Steele et al., 2015], showed how the FFS model makes it difficult to practice innovative, advanced medical procedures, justifying why it is important to reform the existing payment system. The authors initially present two techniques that are applied to patients with liver cancer. The first one (balloon occlusion) is a method proposed by them while the second technique (coil embolization) is widely used by other physicians. Even though their technique was associated with lower cost, similar patient outcomes and faster procedure times relative to the other one, it failed to gain widespread acceptance due to a significant decrease in hospital and physician revenue under the existing FFS system. The authors conclude that switching to the most efficient treatment, the one they proposed, under the existing system is not possible as it leads to loss of profits.

The adoption of Medicare's Sustainable Growth Rate (SGR) model by Congress, in 1997, in an effort to address increasing spending on certain services, has been proven to be unsustainable. This inefficiency is what led Congress to repeal SGR and initiate procedures to implement payment reforms [Steinbrook, 2015]. However, fee reduction is not a panacea. Even if both private and public payers see a reduction in fees, spending may rise, due to increased utilization [Chernew, 2010]. Primarily driven by volume, FFS payment scheme leads to overtreatment, services that would not be needed otherwise, sometimes coupled with poor outcomes. On top of that it limits collaboration and coordination of all participants.

A survey conducted in 2015 by Medicare Current Beneficiary Survey (MCBS) [Noel-Miller, 2015], highlights the variation of OOP medical expenditures for beneficiaries in the FFS program. Relying on a panel survey of 11,000 Medicare beneficiary respondents, in 2011 alone, Medicare insured individuals reported on average an increase in their OOP expenditures of approximately $3,500 or 20 percent of their median spending as a percentage of income. They also noted that under an FFS payment scheme, significant differences occur in terms of OOP health care spending, based on insurers' demographic attributes. Indicatively, women were found to spend more than men, OOP spending increases significantly as people age and Whites pay more than any other race or ethnicity. Even in cases where beneficiaries had supplemental insurance, reported OOP expenditures were higher.

2.2 Review of Bundled Payment Schemes

Following the creation of the inpatient prospective system, in 1983, Medicare began paying a fixed amount per inpatient hospital stay, based on the patient's diagnosis. The first integrated bundled payment project was launched in 1991 by the Centers for Medicare and Medicaid Service and concerned Coronary Artery Bypass Graft (CABG) surgeries [Mechanic, 2016], for all hospitals and services provided to patients within the hospitalization period and any readmissions within 90 days. The project lasted for five years and seven hospitals participated in it. Upon its completion, the program evaluation found a 10% decrease in Medicare spending, along with a significant reduction in death rates. Since then, CMS has initiated plenty of other BP programs with mixed results. For example, in 2009, it began a three-year hospital-physician collaboration program, but the agency did not find significant increase in Medicare cost savings. That same year, CMS launched the Acute Care Episode (ACE) Initiative, which lasted for three years, achieving approximately $600 savings per EoC.

[Froimson et al., 2013], studied whether, aligning the financial incentives of hospitals and surgeons could be achieved by evaluating the Acute Care Episode Demonstration project which was initiated in 2009 by CMS and lasted for three years. They used one and two-year data, based on an extensive financial documentation regarding all the services offered, for two MSAs cities; Albuquerque, New Mexico and Tusla, Oklahoma. After the completion of this pilot program scientists compared their findings, based on data following the start of the ACE project, relative to baseline values. Their findings were at least promising; in both cities their was a 7-10% total cost reduction per episode, along with an increase in hospital revenue. Also, each hospital managed to reduce the overall length of stay (LOS) for patients who underwent a knee/hip replacement surgery. The corresponding cost reduction led to surgeons receiving bonus payments, ranging from

\$275 to \$450 per EoC, accounting for up to 25% of the professional fee [Rana and Bozic, 2014]. These results clearly demonstrate a better financial performance and also certify that better collaboration between hospitals and medical professionals leads hospital centers to improve patient care.

Encouraging results were found by [Carey, 2014] as well. Using data from the State Inpatient Database and supplemental files for revisit analysis, for Medicare beneficiaries older than 65 years old, the author investigated the relationship between the probability of discharge of readmission from an Acute Care Hospital (ACH) and the patient's length of stay. Comparing the expected cost of readmission with an additional day of stay, she finds that the cost of an additional day of stay is approximately 15% lower than the expected cost of readmission. This finding has important policy implications, mainly for hospitals where bundled payment schemes are tested.

Previous bundled payment demonstrations suggest that the application of episode-based payment methods delivered by beneficial effects. Prominent examples like Cardiovascular Care Providers Inc., Medicare Participating Bypass Center and ProvenCare Demonstrations resulted not only in reduction of Out-of-Pocket expenses but also delivery of top-quality medical care and an overall reduction in hospital charges [Shih et al., 2015].

The implementation of bundled payment methods has faced many challenges as well. Despite the fact that various episode-based payment models have been launched and tested for more than 40 years, little evidence exists in the official literature regarding their efficiency. This is mainly due to the complexity of the system implementation and the difficulty to draw firm, robust conclusions. Plenty of research papers have already addressed the complexity when transitioning from one model to the other.

[Hussey et al., 2011], evaluated how well the PROMETHE-US Demonstration

Project performed during the first two years of its implementation by the Health Care Incentives Improvement Institute, in 2009. This pilot program was designed to pay for all of the multiple services provided during a clinical episode. The authors presented six challenges that needed to be addressed before implementing any bundled payment methods: (*i*) a precise description of the services included in the bundle, (*ii*) a clear definition of the payment method, (*iii*) implementation of quality measurement, (*iv*) determining accountability, (*v*) engaging providers to participate in this process and finally (*vi*) care redesign coupled with improvements in care delivery. Taking into account all these details, providers, physicians and other parties will then be able to carefully avoid complications.

[Bozic et al., 2013] demonstrate some of the challenges that were previously presented, by examining bundled payments in Total Joint Arthroplasty procedures. Using data on payments to all Medicare providers for TJA EoC, authors highlight the importance of quality monitoring under bundled-payment scheme. Because the data used by the authors varied widely in terms of patient, procedure and hospital characteristics, safe and generalized conclusions regarding the benefits of bundled-payment policies were difficult to establish. However, as they conclude, thoroughly breaking down all services delivered during an episode is mandatory in order to test the feasibility of bundled-payment reforms. Similar findings were also reported by [Cram et al., 2015] who highlighted many of the complexities of implementing bundled payment reforms for elective primary Total Knee Arthroplasty (TKA) procedures.

Under the precondition that hospitals are financially responsible for post-acute care delivered by providers, [Lau et al., 2014] examined existing post-acute hospital referral networks for Skilled Nursing Facilities (SNF) and Home Health Agencies (HHA). The authors relied on a complete set of non-managed Medicare hospital, nursing home and home health discharges with datasets containing details regarding hospitals' certain characteristics and SNF/HHA referrals, which by 2008, accounted for 2.4 millions in total across the US. The authors point major difficulties that hospitals faced in an effort to coordinate their financial incentives with Post-Acute Care (PAC) providers. This phenomenon occurs mostly because hospitals, being financially responsible for the PAC services delivered to their patients, must smoothly cooperate with providers, supplying them with continuous updates throughout the implementation of pilot programs.

It is evident that the literature in bundled payment reforms gains ground constantly. Irrespectively of the complexity of this health care reform, evidence regarding the quality of care delivered, coupled with cost reduction under bundled-payment schemes is encouraging. Decreased readmission rates at hospitals, reductions in spending levels by reducing the use of costly post-acute care services and better health-care provided to patients are sound examples of the positive effect BP reforms have on a broad range of health-care outcomes.

3 Data & Summary Statistics

In an effort to examine the impact BPCI policies have on health-care expenditures, I use two datasets. The first dataset comes from the Centers for Medicare and Medicaid Services website. It provides us with data concerning the Health Care Organizations (HCO) where various EoC took place and BPCI Models 2-4 were tested. This dataset contains useful information regarding 11,178 episodes of care that took place during the implementation of this pilot program in 1,302 counties, out of 3,144 ones across the United States.

The second dataset consists of individual-level, observational data and were obtained from Current Population Survey (CPS). I use the Annual Social Economic Supplement (ASEC) CPS dataset files, which are widely used in the official literature and are issued jointly by the US Census Bureau and the Census Bureau of Labor Statistics (BLS) annually. The CPS-ASEC data set contains various demographic, socioeconomic and health related variables. Given that BPCI program was launched in 2013, I rely on cross-sectional, ASEC-CPS data files for the last six years; from 2011 up to 2016.

In my analysis, Out of Pocket (OOP) Medical Expenditures and Total Health Expenditures (THE) are the dependent variables with mean values equal to $4,195 and $3,014 respectively. Individuals who are less than 40 years old and live in non-metropolitan counties are dropped from the dataset, in an effort to have a more homogeneous sample, in terms of medical expenses. In addition, I omit observations with county FIPS codes equal to zero as these countries are suppressed in CPS public use data for reasons of confidentiality. Our full and finalized sample consists of 213,058 observations. The dataset is composed of an almost equal number of males and females, with males representing 47 percent of it and the remaining 53 percent corresponding to females. Decomposing race variable, Whites outnumber all other races combined[2] as they account for approximately 55 percent of our sample. As with previous studies, certain demographic variables such as age, educational attainment, marital status etc. are included. A full description of the explanatory variables that are used, along with the dependent variables, can be found in Appendix, Table 1

[Appendix: *Tables*, (Tab. 1)]

Given that the full, finalized sample consists of 368 counties, we end up with 198 treated counties where BPCI policies were implemented. Fig. 1 in the Appendix depicts the places where BPCI Models were run and tested, from a state-level perspective.

[Appendix: *Figures*, (Fig. 1)]

[2] African Americans, Hispanics, Asians and others

4 Empirical Model & Estimation Methods

As mentioned above, one of the objectives of this paper is to examine whether, in counties where BPCI policies were implemented, health care expenses decreased. In this context, the following empirical model is specified:

$$(\mathrm{ME})_{st} = \beta_0 + \alpha_i \boldsymbol{X}_{st} + \beta_1(\mathrm{time})_t + \beta_2(\mathrm{treat})_s + \beta_3((\mathrm{time})_t * (\mathrm{treat})_s) + \lambda_t + \xi_s + \epsilon_{st} \quad (2)$$

In Eq. (2), dependent variable ME is twofold, as it refers to both OOP and Total medical expenditures. Vector $\boldsymbol{X}$ consists of predictors that are commonly used in the official literature as explanatory variables when examining health care costs. Continuous variables age, its square and discrete variables gender, race, marital status, educational attainment, income and region compose parameter vector α. It consists of eight coefficients, each one of them corresponding to the appropriate predictor. λ_t and ξ_s are used to control for time and county effects respectively and ϵ_{st} is the error term. In addition, I let, $treat_s$ and $time_t$, denote two dichotomous variables.

Variable $treat_s$, refers to the counties where BPCI policies were either implemented or not; 0 represents counties which were excluded from the implementation and 1 represents counties that participated in the program. Variable $time_t$ "splits" the six-years period into two separate ones, also called the pre- and post- treatment periods. The "treated" group, is exposed to a "treatment" during the second, three-year period (also called post-treatment period) while counties belonging to the "control" group are not exposed to the treatment during either period. Note that, the coefficient of the interaction term between $time_t$ and $treat_s$, β_3, is of primary interest.

4.1 Two Part Regression

Using nonlinear regression methods for cost driven data is a common practice. Particularly in the area of Health Economics Research, the main reason for applying generalized linear models is, not only to overcome problems of skewed data but also to deal with zero mass issues (excess number of zeroes present in a dataset) [Malehi et al., 2015]. [Nelder and Wedderburn, 1972] were the first researchers to propose the use of Generalized Linear Models (GLMs) as an appropriate method of handling observations with highly skewed distribution. Since then, they are widely used in the field of Health Economics and in studies which involve non-normal dependent variables. Hence, due to the fact that I rely on a dataset which consists of a large number of zero observations, a single index model for such types of observations is not desirable. Instead, our attention turns into exploiting the advantages that Two-Part (2PM) regression models[Belotti et al., 2015] offer in cases when mixed discrete-continuous outcomes are studied [Matsaganis et al., 2008].

Two-Part models have been extensively used in the official literature. [Mihaylova et al., 2010] reviewed the widespread use of these models in healthcare resources and costs. Since the 1970s, scientists have demonstrated their usage in a broad range of topics; from meteorology, e.g. [Cole and Sherriff, 1972], to topics related to health care. While tempting, dropping observations with zero outcome does not fully expose the impact treatments, policies or other covariates have on the entire population, including those who report zero values. Thus, incorporating those zero outcomes in an analysis, help us evaluate the correct treatment effects and/or incremental effects of covariates. Two-part models offer this flexibility and this is why they are preferred over single-equation estimation techniques, e.g. OLS.

The first part of a 2PM regression is a binary choice model for the probability of observing a positive outcome versus a zero one. In this case, a logistic regression analysis is used,

$$Pr(y_i > 0) = \frac{\exp{(\beta \boldsymbol{x}_i)}}{1 + \exp{(\beta \boldsymbol{x}_i)}} = \frac{1}{1 + \exp{(-\beta \boldsymbol{x}_i)}} \tag{3}$$

letting y_i denoting positive or zero medical expenditures. Thus, for individuals who report zero, either OOP or total expenditures, let $y_i = 0$ and for those who report positive ones, $y_i = 1$.

For the second part, conditional on a positive outcome, an appropriate GLM regression model is fit for nonzero, positive medical expenditures. A log natural link function, $g(\mu_i) = log(\mu_i)$, is used along with Gamma distributional family.

4.2 Difference-in-Differences Estimation

When observing outcomes for different groups of two or more time periods, DiD estimation method is used extensively in order to estimate treatment effects. This is accomplished by comparing the pre- and post- treatment differences in the outcome of a treatment and a control group. Leaving aside cases with multiple time periods and control groups, I focus on the conventional case of this technique using two groups (a control and a treatment one) and two periods of time. This setup, which is also used by other authors, e.g. [Meyer, 1995] and [Blundell and Macurdy, 1999], helps me exploit the advantages difference-in-differences method offers, as discovered by [Ashenfelter and Card, 1985].

4.2.1 Linear DiD Model

For each observation, i, let $G_i = \{0, 1\}$ and $T_i = \{0, 1\}$ denote group and time indicators respectively as well as their interaction term, $G_i * T_i$. G_i equals to one when an observation belongs to the treatment group. When $T_i = 1$, observation refers to the post-treatment period. Hence, treatment effect, τ^{DiD}, is defined as,

$$\tau^{\text{DiD}} = \mathbb{E}[Y^1|T = 1, G = 1, \boldsymbol{X}] - \mathbb{E}[Y^0|T = 1, G = 1, \boldsymbol{X}], \tag{4}$$

while, the participation in the treatment is defined by Eq. (5),

$$I = T * G = 1[T = 1, G = 1] \tag{5}$$

Assuming we estimate an equation, linear in parameters, of the form:

$$Y = \beta' \boldsymbol{X} + \gamma_1 T + \gamma_2 G + \gamma_3 (T * G) + \epsilon \tag{6}$$

Based on Eq. (5) and (6), when treatment occurs, we have that $\mathbb{E}[Y^1 | T = 1, G = 1, X] = \beta' \boldsymbol{X} + \gamma_1 + \gamma_2 + \gamma_3$, while for the unobserved (counterfactual) outcome we have that $\mathbb{E}[Y^0 | T = 1, G = 1, \boldsymbol{X}] = \beta \boldsymbol{X} + \gamma_1 + \gamma_2$. Treatment effect is then identified as $\tau^{\text{DiD}} = \gamma_3$.

Due to the structure of the dataset and the estimation method I use, applying a nonlinear DiD model is imperative.

4.2.2 Nonlinear (Logit) DiD Estimation

The difficulty of directly interpreting τ^{DiD} relies on the fact that treatment effect is bounded between 0 and 1, group effects are not constant across time and correspondingly time effects are not constant between groups. A solution Puhani proposes to the above issue is to assume that the difference between groups across time periods is constant to the unobserved latent linear index rather than the limited dependent variable[Puhani, 2011].

In cases like the one I examine, when a Logit regression model is used and based on Eq. (4), the "Logit DiD" estimation is:

$$\mathbb{E}[Y | T, G, X] = F(\beta' \boldsymbol{X} + \gamma_1 T + \gamma_2 G + \gamma_3 T * G) \tag{7}$$

where, $F(\bullet)$, is the transformation function. When Logit regression techniques are applied, the probability of dependent variable, y_i, taking value 1 is modeled as $P(y_i = 1 | x_i) = F(x_i' \beta)$, where $F(\bullet)$ is a single linear index satisfying the following properties: (i) $F(-\infty) = 0$, (ii) $F(\infty) = 1$ and (iii) $\partial F(z)/\partial z > 0, z = x_i'\beta$, mapping the single index into the $[0, 1]$ space.

Interpreting the coefficient of $T * G$, γ_3, in "logit DiD" models requires extra caution. The sign is equal to the one of the treatment effect since transformation function, $F(\bullet)$, is strictly monotonic. Hence, it can be immediately interpreted, certifying whether treatment effect exists or not. Yet, the magnitude of the coefficient of γ_3 can not be interpreted as in the conventional, linear DiD models since treatment effect is estimated as shown in Eq. (8) below:

$$\tau(T = 1, G = 1, X) = F(\beta'\boldsymbol{X} + \gamma_1 + \gamma_2 + \gamma_3) - F(\beta'\boldsymbol{X} + \gamma_1 + \gamma_2) \tag{8}$$

Eq. (8) relies heavily on the fact that, as Puhani states, in any nonlinear model with a strictly monotonic function, $F(\bullet)$, Treatment Effect is "the cross difference of the conditional expectation of the observed outcome, $\mathbb{E}[Y^1|T, G, \boldsymbol{X}]$, minus the cross difference of the conditional expectation of the counterfactual outcome, $\mathbb{E}[Y^0|T, G, \boldsymbol{X}]$". In cases where, coefficient $\gamma_3 \neq 0$, Treatment Effect is the incremental effect of the coefficient of the interaction term, γ_3.

It is essential to fully present the procedure that is followed in order to estimate τ^{DiD} in Eq. (8). Since Stata is the statistical software that is used for econometric analysis, we fully take advantage of the `margins` postestimation command with one major difference in the regression equation. Instead of specifying interaction term, $T * G$, directly in the regression, a pre-specified interaction variable is created. This new variable is then evaluated at $T = 1, G = 1$ providing the appropriate estimates with the corresponding standard errors. In terms of my analysis and the regression model that is used, evaluating the incremental effect on β_3 is accomplished by specifying the values of covariates, time_t and treat_c, to be equal to 1.

5 Results

For the two types of Medical Expenditures (ME), Out of Pocket and Total, both parts of the two-part model regressions are shown in the Appendix,

[Appendix: *Tables*, (Tab. 2, 3, 4, *&* 5)]

Based on the analysis presented in Section (4), Table (7) is extremely important as it helps us quantify treatment effect and, consequently examine whether the hypotheses that are under investigation are true.

All tables presented in Appendix, consist of four columns. The first one shows results based on original population (full sample). The remaining three columns refer to estimates over three different subsamples; (i) elder people (older than 64 years old), (ii) these who are poor (with income equal or less than $25,000) and lastly (iii) a combination of those two criteria, elders with low income. Tables (2) - (5) report estimated coefficients. In particular, Tables (2) and (4) concern estimates for the Logit part of the Two-Part model to predict the probability of positive medical expenditures, and Tables (3) and (5) refer to the GLM regression coefficients to predict the level of expenditures for those who have reported more than zero.

5.1 Medical Expenditures

In most cases estimates obtained for both types of Medical Expenditures (ME) demonstrate the expected associations. Both for the first and second part of regressions, the vast majority of the explanatory variables maintain a plausible sign, attaining statistical significance at all conventional significance levels. Indicatively, for the first part, as shown in Table (2), individuals who are older, females, highly educated, have higher income or live in the region of Midwestern United States, were likely to report positive health care expenditures, holding all other variables constant. On the other hand, individuals who are married, are less likely to report positive ME. Lastly, relative to Whites, all other races are more likely to declare zero OOP and Total ME. Findings are aligned with those of scientific literature.

Most variables keep their sign and significance when running the second part of the two-part model[3]. Given that for the GLM part, a log-link function has been used, conditional mean has an exponential form. Thus, coefficients can be directly interpreted as percent changes[4]. Looking at Column (1) on Table (3), which corresponds to full sample, with each additional year of age, (OOP) ME are expected to increase by 3.8 percents. Similarly, females spend 1.3% more on health care relative to males and Hispanics spend on average 18.5% less than Whites. The rest of the coefficients are interpreted similarly.

The crux of this paper relies on the results shown in Table (7). The estimates presented on this table, reveal the true effect of BPCI initiative, implemented by CMS in the "treated" counties and during the post-treatment period. For all four samples, looking at Table (7), the sign of estimated coefficient of the interaction term, $(\text{time}_t) * (\text{treat}_c)$, henceforth $\hat{\beta}_3$, is negative. The sign itself certifies that in areas where this policy was implemented, residents of those counties reported on average lower costs, for both OOP and Total ME. Interpreting these results, in counties where bundled payment reforms were implemented, people older than 64 years old were found to have reduced OOP ME by \$180 (an 8% reduction) and Total ME by \$123 approximately (a 13% reduction). In a similar manner, reduced spendings were found both for low-income individuals and those who are poor and older. For example, people with income less than \$25,000 and who are older than 64 years old, reported on average less ME than those who live in counties where

[3]Although variations, in terms of sign and statistical significance are still present; for example coefficients of variable (*Region*) or *Gender*.

[4]For example, assuming that $\hat{\beta_1}$ is the estimate of coefficient β_1, the percent change of β_1, *ceteris paribus*, equals to $\exp(\hat{\beta_1}) - 1$

BPCI policy was not implemented. Table (8) reflects the positive impact BPCI policy had on areas where it was implemented.

[Appendix: *Tables*, (Tab. 8)]

The rest of estimated coefficients reported in Table (7) can also be interpreted this way.

6 Sensitivity Analysis and Robustness of Results

In order to check the robustness of my results, numerous other regression models were tested. The most important results obtained from these analyses are reported in Table (7). For ease of interpretation and convenience, estimates for all other independent variables apart from $\hat{\beta}_3$, which concern the DiD treatment effect estimator, are not shown. The first regression model, Model (I)[5], corresponds to Eq. (1) without the parameter vector, α_i. Model (II), includes only age, gender and income as explanatory variables while for Model (III), all covariates are included except for the year effects, in order to not omit year 2016[6]. Regression analyses were conducted both for the full sample and the three predefined subsamples. However, estimates obtained from regressions based on full sample provided us with insignificant results and they are not shown.

Regardless of specification, relying on estimates presented on Table (7), it is evident that estimated coefficient, $\hat{\beta}_3$, is consistent in terms of maintaining its negative sign and for being statistically different from zero. Small differences with respect to its magnitude are apparent, however such variations are small for both types of medical expenditures. Apart from examining the effect BPCI policy had on elders and poor people, a separate analysis was conducted for the disabled persons, without finding significant results.

[5] first row of Table (7)

[6] Inclusion of year effects, both for the baseline regression and for all other specification models leads to omission of 2016, the last year of the BPCI pilot program, due to collinearity

Before applying the estimation methods I presented above, various other models were run and tested as well; for example, log-Linear estimation methods were used with individuals reporting zero ME being dropped from this set of regressions. Even though such an approach has a comparative advantage when it comes to interpreting the DiD treatment effect coefficient, it leads to biased and inconsistent estimators due to the nature of data used. Results for such types of regressions are excluded and not shown in this paper.

Given that both dependent variables had highly-skewed distributions and dealing with zero-mass problems was inevitable due to the nature of my dataset, Two-Part models offer the ability to relax both the assumption of heteroscedasticity and normality when it comes to obtaining consistent estimators [Cameron and Trivedi, 2009]. The set of explanatory variables that were used in the aforementioned analysis do not violate the assumption of exogeneity.

7 Conclusion

The aim of this paper was to investigate whether BP methods have positive impact in counties where such policies were tested. Using a combination of CMS and ASEC-CPS datasets, statistically significant results were drawn for individuals with low socioeconomic status, living in "treated" counties. With Out-of-Pocket and Total Medical Expenditures as the outcomes of interest, the analysis conducted above found reduced health-care expenses in areas where Bundled Payment reforms were launched and tested. To the extent of my knowledge, there were no other programs being tested during the three-year, post-treatment period I study.

Certain limitations needed to be overcome while conducting this analysis. On the data side, the datasets that were used provide limited information with respect to individuals' medical expenses. This phenomenon restricted my analysis as I could solely rely on just two dependent variables. Despite the fact that MEPS data files provide us with detailed description of US residents' ME, this alternative was not possible. For reasons of confidentiality, MEPS datasets do not include information regarding the residence of respondents, thus it would not let me merge those files with the dataset provided by CMS. Repeating this analysis based on more detailed health-care expenditures would have multiple advantages and provide policymakers with even more insightful results. On the estimator side, two important things need to be mentioned. Firstly, since BPCI is an ongoing process, my study examines the effect it has for only two years post program's initiation[7]. Future analyses could exploit data for more years ensuring an even better, more robust evaluation of BP reforms. Secondly, given the survey design and the fact that individuals are interviewed, there may be circumstances when survey questions are misunderstood by respondents. Thus, classical measurement error is a concerning factor when dealing with data derived from such surveys.

In summary, this study contributes solidly in the existing literature. Firstly, it empirically demonstrates Puhani's famous paper for estimating and evaluating DiD treatment effects in nonlinear settings. From a policy perspective and to the extent of my knowledge there are no other studies that have examined Bundled Payment Reforms, both from a county level perspective as well as in terms of certain population characteristics. Evidently, the encouraging results being found, have important implications for policymakers. The aforementioned concrete estimates strengthen the argument to transition from a widely-applied FFS payment scheme to bundled payments, especially since such implications positively affect certain, vulnerable population groups.

[7]Note that the post-treatment period used in this analysis refers to a three-year period, but the inclusion of year effects implies the omission of final year, 2016.

References

National health expenditures (NHE) Projections 2016-2025 - Forecast Summary. Center for Medicare and Medicaid Services, 2016a.

CMS Bundles Payments for Care Improvement Initiative Models 2-4: Year 2 Evaluation & Monitoring Annual Report. Center for Medicare and Medicaid Services, 2016b.

E. Antonova, M. Boye, M. Sen, A. O'Sullivan, and R. Burge. Can Bundled Payment Improve Quality and Efficiency of Care for Patients with Hip Fractures? *Taylor & Francis, Journal of Aging & Social Policy*, 27:1–20, 2015.

O. Ashenfelter and D. Card. Using the Longitudinal Structure Earnings to Estimate the Effect of Training Programs. *Review of Economics and Statistics*, 67, November 1985.

S. Athey and G. W. Imbens. Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, 24:431–497, 2006.

F. Belotti, P. Deb, W. G. Manning, and E. C. Norton. twopm: Two-part models. *Stata Journal*, 15(1):3–20(18), 2015. URL `http://www.stata-journal.com/article.html?article=st0368`.

M. Bertrand, E. Duflo, and S. Mullainathan. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, pages 249–275, 2004.

D. M. Berwick and M. D. Hackbarth. Eliminating waste in us health care. *Journal of American Association*, 2012.

R. Blundell and T. Macurdy. *Labor Supply: A Review of Alternative Approaches, Handbook of Labor Economics*, volume 3. Elsevier Science, 1999.

J. K. Bozic, L. Ward, T. Vail, and M. Maze. Bundled Payments in Total Joint Arthroplasty: Targeting Opportunities for Quality Improvement and Cost Reduction. *Symposium: Knee Society Proceedings, Clinical Orthopeadics and Related Research*, 2013.

A. C. Cameron and P. K. Trivedi. Microeconometrics Using Stata. *A Stata Press Publication, Chapter 16.5: Two-Part model in Logs*, pages 538–541, 2009.

K. Carey. Measuring The Hospital Length of Stay/Readmission Cost Trade-Off Under A Bundled Payment Mechanism. *Wiley Online Library, Health Economics*, May 2014.

M. E. Chernew. Bundled Payment Systems: Can They Be More Successful This Time. *HSR: Health Services Research and Educational Trust*, 45:1141–1147, October 2010.

M. E. Chernew. Why Physicians Should Like Bundled Payment. *HSR: Health Services Research and Educational Trust*, 46:1963–1967, December 2011.

J. A. Cole and J. D. F. Sherriff. Some Single- And Multi-Site Models Of Rainfall Within Discrete Time Increments. *North-Holland Publishing Company, Journal of Hydrology*, 17, 1972.

P. Cram, X. Lu, and Y. Li. Bundled Paymemnts for Elective Primary Total Knee Arthroplasty: An Analysis of Medicare Administrative Data. *Sage, Geriatric Orthopedic Surgery & Rehabilitation*, 61, 2015.

R. D. Desisle. Big Things in Bundled Packages: Implications Of Bundled Payment Systems in Health Care Reimbursement Reform. *American Journal of Medical Quality*, 28, 2013.

M. Friedberg, C. P., O. Jung, L. Raaen, S. Hirshman, E. Hoch, C. Stevens, P. Ginsburg, L. Casalino, M. Tutty, C. Vargo, and L. Lipinski. *Effects of Health Care Payment Models on Physician Practice in the United States*. RAND Corporation, American Medical Association, 2015.

M. Froimson, A. Rana, R. White, A. Marshall, S. Schutzer, W. Heavy, P. Naas, G. Duabert, R. Iorio, and B. Parsley. Bundled Payments for Care Improvement Incentive: The Next Evolution of Payment Formulations. *American Association of Hip and Knee Surgeons (AAHKS) Bundled Payment Task Force, The Journal of Arthroplasty, Elsevier*, July 2013.

L. Garber and J. Skinner. Is american health care uniquely inefficient? *The Journal of Economic Perspectives, JSTOR*, 22, 2008.

S. Grubaugh and R. Santerre. Comparing the Performance of Health Care Systems: An Alternative Approach. *Southern Economic Journal*, 60:1030–1042, 1994.

P. Hussey, S. Ridgely, and M. Rosenthal. The PROMETHEUS Bundled Payment Experiment: Slow Start Shows Problems in Implementing New Payment Methods. *Health Affairs 30*, 10, 2011.

C. Lau, A. Alpert, P. Huckfeldt, P. Hussey, D. Auerbach, H. Liu, N. Sood, and A. Mehrtra. Post-Acute Referral Patterns For Hospitals and Implications for Bundled Payment Initiatives. *Elsevier, Healthcare*, pages 190–195, May 2014.

L. Lorenzoni, A. Belloni, and F. Sasse. Healthcare Expenditure And Health Policy in the USA Versus Other High-Spending OECD Countries. *The Health of Americans (No. 5), Health Division: Organization for Economic Co-operation and Development (OECD)*, July 2014.

A. S. Malehi, F. Pourmotahari, and K. F. Angali. Statistical Models For The

Analysis of Skewed Healthcare Cost Data: A Simulation Study. *SpringerOpen Journal, Health Economics Review*, 2015.

M. Matsaganis, T. Mitrakos, and P. Tsakloglou. Modelling health expenditure at the household level in Greece. *Springer-Verlag, European Journal of Health Economics*, December 2008.

R. Mechanic. Medicare's Bundled Payment Initiatives: Considerations for Providers. *American Hospital Association (AHA)*, pages 1–16, 2016.

R. Mechanic and C. Tompkins. Lessons Learned Preparing For Medicare Bundled Payments. *The New England Journal Of Medicine*, 2012.

B. D. Meyer. Natural and Quasi-Expirements in Economics. *Journal of Business and Economic Statistics, JBES Symposium on Programming Policy Evaluation*, 13:151–161, April 1995.

B. Mihaylova, A. Briggs, A. O'Hagan, and S. G. Thompson. Review of Statistical Methods For Analyzing Healthcare Resources and Costs. *Wiley Online Library, Journal of Health Economics*, 20, August 2010.

J. A. Nelder and W. M. Wedderburn. Generalized Linear Models. *Wiley, Journal of the Royal Statistical Society*, 135, 1972.

C. Noel-Miller. Medicare Beneficiaries Out-Of-Pocket Spending for Health Care: Insight on The Issues. *American Association of Retires Persons (AARP) Public Policy Institute*, 2015.

P. A. Puhani. The treatment effect, the cross-difference and the interaction term in nonlinear "difference-in-difference" models. *Elsevier, Economic Letters*, December 2011.

A. J. Rana and K. J. Bozic. Bundled Payments in Orthopeadics. *Springer, The Association of Bone and Joint Surgeons: Clinical Orthopaedics and Related Research*, February 2014.

T. Shih, L. M. Chen, and B. K. Nallamothu. Will Bundled Payments Change Health Care? Examining the Evidence Thus Far in Cardiovascular Care. *Circulation, Health Policy & Cardiovascular Medicine*, June 2015.

A. H. G. M. Spithoven. Why U.S. health care expenditure and ranking on health care indicators are so different from Canada's. *International Journal of Health Care Finance Economics*, pages 1–24, 2009.

J. R. Steele, K. A. Jones, P. E. Nihan, and et al. Why Bundled Payment Could Drive Innovation: An Example from Interventional Oncology. *Health Policy*, 2015.

R. Steinbrook. The Repeal of Sustainable Growth Rate For Physician Payment. *Journal of American Medical Association*, 2015.

Appendices

A Tables

Table 1. Descriptive Statistics

Variable	Full Sample (n=213,058) Mean [SD]	Age≥65 (n=57,265) Mean [SD]	Inc<25K (n=96,798) Mean [SD]	Age≥65 & Inc<25K (n=34,535) Mean [SD]
Age	57.199 [12.104]	73.67 [6.376]	59.734 [12.948]	74.45 [6.431]
Age^2	3418.3 [1474.1]	5468 [954.7]	3735.8 [1606.8]	5583.8 [966.5]
Female	0.533 [0.499]	0.565 [0.496]	0.658 [0.474]	0.660 [0.473]
Race				
White	Reference	Reference	Reference	Reference
Black	0.114 [0.318]	0.142 [0.349]	0.157 [0.363]	0.155 [0.362]
Hispanic	0.114 [0.318]	0.116 [0.321]	0.121 [0.326]	0.127 [0.333]
Other	0.199 [0.399]	0.154 [0.361]	0.259 [0.438]	0.199 [0.399]
Married	0.363 [0.429]	0.447 [0.497]	0.420 [0.494]	0.479 [0.499]
Education				
No HS diploma	Reference	Reference	Reference	Reference
HS diploma	0.443 [0.497]	0.474 [0.499]	0.501 [0.500]	0.505 [0.499]
Ass. Degree	0.093 [0.289]	0.069 [0.254]	0.079 [0.270]	0.063 [0.244]
Bachelor's	0.197 [0.398]	0.149 [0.356]	0.125 [0.331]	0.101 [0.301]
Master's +	0.129 [0.335]	0.118 [0.322]	0.055 [0.228]	0.057 [0.231]
Income				
$< 25,000$	Reference	Reference	Reference	Reference
$< 50,000$	0.243 [0.429]	0.219 [0.414]		
$< 75,000$	0.134 [0.341]	0.086 [0.203]		
$\leq 100,000$	0.067 [0.250]	0.037 [0.100]		

Continued on next page

Table 1 – continued from previous page

	Full Sample	Age≥65	Inc<25K	Age≥65 & Inc<25K
Variable	Mean [SD]	Mean [SD]	Mean [SD]	Mean [SD]
$> 100,000$	0.101 [0.301]	0.053 [0.224]		
Region				
Northeast	Reference	Reference	Reference	Reference
Midwest	0.135 [0.341]	0.128 [0.333]	0.121 [0.327]	0.119 [0.324]
South	0.318 [0.466]	0.325 [0.468]	0.321 [0.467]	0.333 [0.471]
West	0.357 [0.479]	0.349 [0.477]	0.373 [0.483]	0.353 [0.478]
Post*Treated	0.319 [0.466]	0.333 [0.471]	0.316 [0.465]	0.325 [0.469]
OOP ME	2187 [4195]	2179 [4469]	1516 [3805]	1758 [4667]
Total ME	805 [3016]	960.8 [3727]	698.6 [3110]	813 [4161]

In parentheses, standard deviation is reported for both the explanatory and dependent variables

Table 2. Logistic Regression Results for OOP ME

Variable	Full Sample (n=212,698)	Age≥65 (n=55,457)	Inc<25K (n=96,220)	Age≥65 & Inc<25K (n=32,764)
Age	0.563***	0.152***	0.059***	0.136***
	(0.009)	(0.055)	(0.010)	(0.074)
Age^2	-0.0003***	-0.001***	-0.0003***	-0.0009*
	(0.00008)	(0.0003)	(0.00009)	(0.0005)
Female	0.267***	0.203***	0.272***	0.175***
	(0.016)	(0.035)	(0.017)	(0.038)
Race				
White	Reference	Reference	Reference	Reference
Black	-0.352***	-0.349***	-0.332***	-0.323***
	(0.039)	(0.090)	(0.044)	(0.099)
Hispanic	-0.430***	-0.638***	-0.480***	-0.697***
	(0.072)	(0.109)	(0.069)	(0.124)
Other	-0.360***	-0.469***	-0.345***	-0.480***
	(0.032)	(0.058)	(0.039)	(0.066)
Married	-0.161***	-0.210***	-0.286***	-0.249***
	(0.025)	(0.035)	(0.028)	(0.041)
Education				
No HS diploma	Reference	Reference	Reference	Reference
HS diploma	0.362***	0.368***	0.339***	0.342***
	(0.022)	(0.049)	(0.024)	(0.051)
Ass. Degree	0.551***	0.530***	0.545***	0.512***
	(0.036)	(0.080)	(0.039)	(0.092)
Bachelor's	0.597***	0.554***	0.629***	0.588***
	(0.043)	(0.068)	(0.052)	(0.077)
Master's +	0.640***	0.671***	0.682***	0.601***
	(0.056)	(0.102)	(0.069)	(0.128)
Income				
$< 25,000$	Reference	Reference	Reference	Reference
$< 50,000$	0.728***	0.685***		
	(0.025)	(0.048)		
$< 75,000$	0.950***	0.630***		
	(0.035)	(0.074)		

Continued on next page

Table 2 – continued from previous page

Variable	Full Sample (n=212,698)	Age≥65 (n=55,457)	Inc<25K (n=96,220)	Age≥65 & Inc<25K (n=32,764)
$\leq 100,000$	1.168***	1.093***		
	(0.055)	(0.142)		
$> 100,000$	1.184***	0.734***		
	(0.050)	(0.105)		
Region				
Northeast	Reference	Reference	Reference	Reference
Midwest	0.292***	0.166***	0.615***	-0.124***
	(0.055)	(0.079)	(0.064)	(0.112)
South	-0.491***	-0.850***	0.060***	-0.742***
	(0.010)	(0.019)	(0.014)	(0.032)
West	0.748***	-0.062***	0.863***	0.021***
	(0.071)	(0.106)	(0.838)	(0.128)
Post*Treated	-0.039	-0.180	-0.120	-0.172
	(0.094)	(0.128)	(0.110)	(0.152)
No of Clusters (Logit)	360	307	350	283

In parentheses, robust standard errors reported, adjusted for clustering at the county level. In addition, (*) declares statistical significance at the 10% Significance Level, (**) at the 5% SL and lastly, (***) at the 1% SL. In all four regressions county and year effects are included

Table 3. Generalized Linear Regression Results for OOP ME

Variable	Full Sample (n=191,366)	Age≥65 (n=52,144)	Inc<25K (n=82,104)	Age≥65 & Inc<25K (n=30,429)
Age	0.038***	-0.029	0.054***	-0.008
	(0.005)	(0.025)	(0.006)	(0.032)
Age^2	-0.0002***	0.0002	-0.0003***	0.00007
	(0.00004)	(0.0001)	(0.00004)	(0.0002)
Female	-0.013***	0.038***	-0.011	0.027*
	(0.002)	(0.012)	(0.014)	(0.015)
Race				
White	Reference	Reference	Reference	Reference
Black	-0.231***	-0.234***	-0.282***	-0.238***
	(0.023)	(0.0248)	(0.024)	(0.030)
Hispanic	-0.170***	-0.223***	-0.196***	-0.271***
	(0.019)	(0.034)	(0.035)	(0.046)
Other	-0.254***	-0.308***	-0.322***	-0.367***
	(0.020)	(0.028)	(0.027)	(0.036)
Married	-0.062***	-0.083***	-0.131***	-0.126***
	(0.011)	(0.016)	(0.019)	(0.024)
Education				
No HS diploma	Reference	Reference	Reference	Reference
HS diploma	0.197***	0.152***	0.181***	0.153***
	(0.024)	(0.033)	(0.024)	(0.035)
Ass. Degree	0.319***	0.267***	0.324***	0.248***
	(0.031)	(0.050)	(0.033)	(0.054)
Bachelor's	0.332***	0.322***	0.376***	0.345***
	(0.029)	(0.038)	(0.035)	(0.047)
Master's +	0.402***	0.392***	0.525***	0.428***
	(0.034)	(0.044)	(0.047)	(0.059)
Income				
$< 25,000$	Reference	Reference	Reference	Reference
$< 50,000$	0.310***	0.225***		
	(0.014)	(0.018)		

Continued on next page

Table 3 – continued from previous page

Variable	Full Sample (n=191,366)	Age≥65 (n=52,144)	Inc<25K (n=82,104)	Age≥65 & Inc<25K (n=30,429)
$< 75,000$	0.415***	0.291***		
	(0.014)	(0.025)		
$\leq 100,000$	0.477***	0.318***		
	(0.019)	(0.036)		
$> 100,000$	0.587***	0.399***		
	(0.020)	(0.035)		
Region				
Northeast	Reference	Reference	Reference	Reference
Midwest	0.133***	-0.150***	0.135***	0.098***
	(0.021)	(0.023)	(0.0266)	(0.036)
South	0.072***	-0.103***	-0.042***	-0.020
	(0.005)	(0.007)	(0.009)	(0.013)
West	0.187***	-0.200***	-0.078**	-0.255***
	(0.029)	(0.035)	(0.034)	(0.042)
Post	0.88***	0.002	0.121***	0.037
	(0.031)	(0.036)	(0.027)	(0.044)
Treated	0.042***	-0.364***	-0.148***	-0.398***
	(0.032)	(0.037)	(0.009)	(0.046)
Post*Treated	-0.025	-0.081*	-0.100**	-0.119***
	(0.032)	(0.041)	(0.040)	(0.049)
No of Clusters(GLM)	368	308	368	367

In parentheses, robust standard errors reported, adjusted for clustering at the county level. In addition, (*) declares statistical significance at the 10% Significance Level, (**) at the 5% SL and lastly, (***) at the 1% SL. In all four regressions county and year effects are included

Table 4. Logistic Regression Results for TME

Variable	Full Sample (n=213,058)	Age≥65 (n=57,135)	Inc<25K (n=96,777)	Age≥65 & Inc<25K (n=34,428)
Age	0.073***	0.092**	0.080***	0.087*
	(0.006)	(0.040)	(0.006)	(0.047)
Age^2	-0.0005***	-0.0006**	-0.0005***	-0.0006*
	(0.00006)	(0.0002)	(0.00005)	(0.0003)
Female	0.322***	0.209***	0.307***	0.186***
	(0.001)	(0.021)	(0.015)	(0.028)
Race				
White	Reference	Reference	Reference	Reference
Black	-0.253***	-0.206***	-0.264***	-0.222***
	(0.029)	(0.042)	(0.030)	(0.051)
Hispanic	-0.263***	-0.392***	-0.338***	-0.473***
	(0.052)	(0.069)	(0.049)	(0.076)
Other	-0.318***	-0.371***	-0.338***	-0.374***
	(0.025)	(0.041)	(0.028)	(0.048)
Married	-0.283***	-0.199***	-0.431***	-0.269***
	(0.017)	(0.028)	(0.021)	(0.033)
Education				
No HS diploma	Reference	Reference	Reference	Reference
HS diploma	0.331***	0.282***	0.314***	0.294***
	(0.020)	(0.033)	(0.021)	(0.036)
Ass. Degree	0.475***	0.345***	0.456***	0.313***
	(0.029)	(0.055)	(0.038)	(0.063)
Bachelor's	0.563***	0.435***	0.595***	0.453***
	(0.031)	(0.053)	(0.038)	(0.059)
Master's +	0.615***	0.552***	0.670***	0.476***
	(0.041)	(0.058)	(0.051)	(0.077)
Income				
$< 25,000$	Reference	Reference	Reference	Reference
$< 50,000$	0.548***	0.437***		
	(0.021)	(0.031)		

Continued on next page

Table 4 – continued from previous page

Variable	Full Sample (n=213,058)	Age≥65 (n=57,135)	Inc<25K (n=96,777)	Age≥65 & Inc<25K (n=34,428)
$< 75,000$	0.696***	0.422***		
	(0.032)	(0.050)		
$\leq 100,000$	0.854***	0.649***		
	(0.038)	(0.071)		
$> 100,000$	0.882***	0.652***		
	(0.044)	(0.074)		
Region				
Northeast	Reference	Reference	Reference	Reference
Midwest	0.476***	0.418***	0.411***	0.225***
	(0.037)	(0.047)	(0.043)	(0.059)
South	-0.024***	0.404***	-0.001	-0.104***
	(0.007)	(0.010)	(0.009)	(0.017)
West	0.834***	1.29*8*	0.478***	0.658***
	(0.051)	(0.061)	(0.056)	(0.063)
Post*Treated	-0.032	-0.061	-0.098	-0.051
	(0.062)	(0.074)	(0.069)	(0.076)
No of Clusters (Logit)	368	368	365	351

In parentheses, robust standard errors reported, adjusted for clustering at the county level. In addition, (*) declares statistical significance at the 10% Significance Level, (**) at the 5% SL and lastly, (***) at the 1% SL. In all four regressions county and year effects are included

Table 5. Generalized Linear Regression Results for TME

Variable	Full Sample (n=153,043)	Age≥65 (n=42,357)	Inc<25K (n=61,290)	Age≥65 & Inc<25K (n=23,744)
Age	0.073***	0.007	0.060***	0.017
	(0.005)	(0.040)	(0.008)	(0.052)
Age^2	-0.0005***	-0.00003	-0.0004***	-0.0001
	(0.00004)	(0.0003)	(0.00006)	(0.0003)
Female	0.048***	0.055***	-0.011	0.039
	(0.014)	(0.019)	(0.021)	(0.026)
Race				
White	Reference	Reference	Reference	Reference
Black	-0.213***	-0.185***	-0.237***	-0.126***
	(0.037)	(0.035)	(0.031)	(0.046)
Hispanic	-0.225***	-0.192***	-0.213***	-0.204***
	(0.021)	(0.043)	(0.041)	(0.061)
Other	-0.237***	-0.227***	-0.275***	-0.261***
	(0.028)	(0.039)	(0.037)	(0.051)
Married	-0.037**	-0.063***	-0.066***	-0.140***
	(0.015)	(0.023)	(0.023)	(0.030)
Education				
No HS diploma	Reference	Reference	Reference	Reference
HS diploma	0.066***	0.055	0.067**	0.077*
	(0.030)	(0.040)	(0.032)	(0.041)
Ass. Degree	0.200***	0.173***	0.228***	0.189**
	(0.042)	(0.060)	(0.053)	(0.075)
Bachelor's	0.172***	0.238***	0.182***	0.278***
	(0.036)	(0.047)	(0.041)	(0.054)
Master's +	0.255***	0.284***	0.313***	0.313***
	(0.042)	(0.055)	(0.063)	(0.079)
Income				
$< 25,000$	Reference	Reference	Reference	Reference
$< 50,000$	0.009	0.124***		
	(0.019)	(0.022)		

Continued on next page

Table 5 – continued from previous page

Variable	Full Sample (n=153,043)	Age≥65 (n=42,357)	Inc<25K (n=61,290)	Age≥65 & Inc<25K (n=23,744)
$< 75,000$	-0.032* (0.019)	0.122*** (0.036)		
$\leq 100,000$	0.011 (0.034)	0.147*** (0.052)		
$> 100,000$	0.073*** (0.026)	0.201*** (0.047)		
Region				
Northeast	Reference	Reference	Reference	Reference
Midwest	0.073*** (0.031)	0.023 (0.027)	0.072** (0.032)	0.295*** (0.041)
South	-0.046*** (0.005)	-0.124*** (0.014)	-0.102*** (0.009)	-0.058*** (0.018)
West	0.189*** (0.040)	-0.104*** (0.043)	-0.076* (0.045)	-0.265*** (0.54)
Post*Treated	-0.060* (0.051)	-0.121** (0.055)	-0.132** (0.058)	-0.144** (0.068)
No of Clusters(GLM)	368	367	368	367

In parentheses, robust standard errors reported, adjusted for clustering at the county level. In addition, (*) declares statistical significance at the 10% Significance Level, (**) at the 5% SL and lastly, (***) at the 1% SL. In all four regressions county and year effects are included

Table 6. "DiD" Treatment Effect for both OOP and Total ME

Marginal Effects	Full Sample	Age≥65	Inc<25K	Age≥65 & Inc<25K
$\frac{\partial y}{\partial x}\vert_{\text{OOPME}}$	-63.653	-180.045**	-173.769***	-211.165**
	(76.292)	(83.547)	(67.395)	(83.191)
$\frac{\partial y}{\partial x}\vert_{\text{TME}}$	-51.660	-123.095**	-104.868**	-122.141**
	(45.132)	(54.44)	(42.860)	(57.317)
$N_{(\text{OOP})}$	212,698	55,457	96,220	32,764
$N_{(\text{TME}}$	213,058	57,135	96,777	34,428

Standard errors reported in parentheses. Treatment effect in a Logit, Difference-in-Difference model, τ^{DiD}, is the incremental effect on the Prespecified interaction term, $(treat)_i * (time)_i$*, evaluated at $(treat)_i$=1 and $(time)_i$=1. The desired result is obtained by evaluating marginal effects and then measuring the instantaneous change in the interaction term

Table 7. "DiD" TE according to Specification Models presented in Sec. 6

Spec. Model	Est. Coef.	Age≥65	Inc<25K	Age≥65 & Inc<25K
Model (I)	$\frac{\partial y}{\partial x}\vert^{(I)}_{\text{OOP}}$	-194.226**	-152.320**	-205.536**
		(90.014)	(64.040)	(83.534)
	$\frac{\partial y}{\partial x}\vert^{(I)}_{\text{TME}}$	-123.148**	-94.596**	-104.868**
		(94.596)	(41.896)	(54.309)
Model (II)	$\frac{\partial y}{\partial x}\vert^{(II)}_{\text{OOP}}$	-189.712**	-159.658**	-206.127**
		(87.475)	(65.276)	(83.300)
	$\frac{\partial y}{\partial x}\vert^{(II)}_{\text{TME}}$	-119.134**	-99.571**	-110.615**
		(53.937)	(42.501)	(54.004)
Model (III)	$\frac{\partial y}{\partial x}\vert^{(III)}_{\text{OOP}}$	-187.594**	-174.799**	-221.064**
		(87.608)	(68.350)	(90.358)
	$\frac{\partial y}{\partial x}\vert^{(III)}_{\text{TME}}$	-117.975**	-107.059**	-114.435**
		(53.951)	(43.396)	(55.050)

In parentheses, Standard Errors reported. For Model (I), parameter vector has been excluded, for Model (II) the parameter vector consists of age, gender and income and in Model (III) year effects are not included. Number of Observations are the same as the ones reported on Table (6)

Table 8. Percent Reduction in Medical Expenditures, (ME).

Exp.	Age≥65	Inc.< 25K	Age≥65 & Inc<25K
(OOP)	8.03%	4.7%	11.6%
TME	12.9%	15.0%	14.8%

* Percentages are calculated based on results reported in Table (6) in the Appendix and mean values of Medical Expenditures according to the appropriate socioeconomic criterion

B Figures

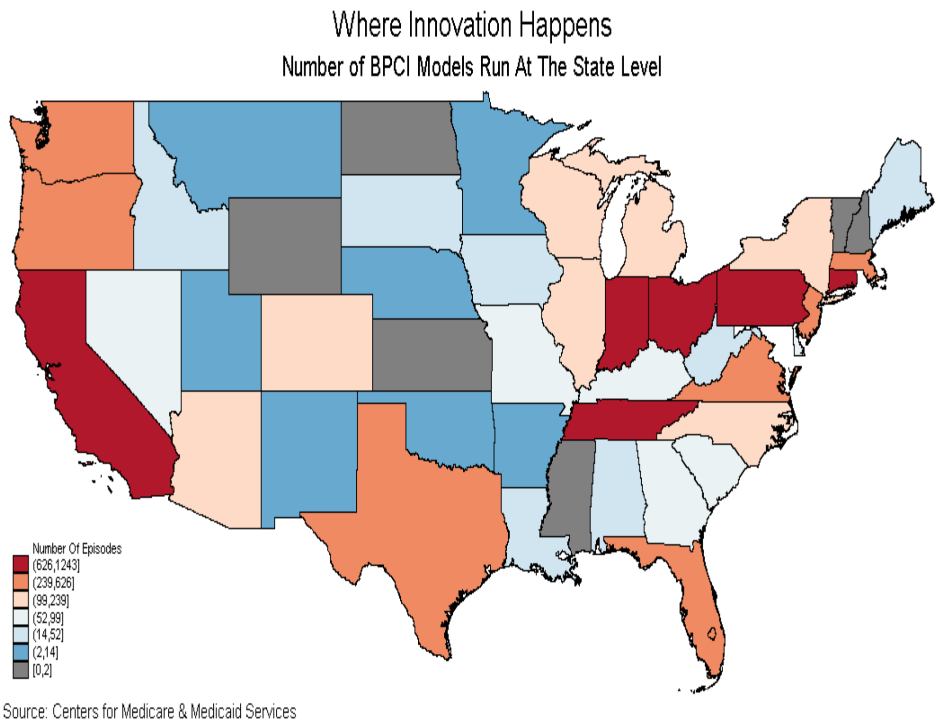

Figure 1. Number of BPCI Models Run at The State-Level