This work is distributed as a Discussion Paper by the

STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 15-030

Insurers Response to Selection Risk:

Evidence from Medicare Enrollment Reforms

By

Francesco Decarolis and Andrea Guglielmo

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Stanford Institute for Economic Policy Research at Stanford University supports research bearing on economic and public policy issues. The SIEPR Discussion Paper Series reports on research and policy analysis conducted by researchers affiliated with the Institute. Working papers in this series reflect the views of the authors and not necessarily those of the Stanford Institute for Economic Policy Research or Stanford University

Insurers Response to Selection Risk: Evidence from Medicare Enrollment Reforms

Francesco Decarolis and Andrea Guglielmo*

September 29, 2015

Abstract

Evidence on insurers behavior in environments with both risk selection and market power is largely missing. We fill this gap within the context of privatized Medicare providing one of the first empirical accounts of how insurers adjust plan features when faced with a potential change in selection. Our empirical strategy exploits the combined effects of a Medicare reform that altered the potential selection risk of the highest quality (5-star) Part C and D plans and the geographical dispersion of such plans over the US territory. Starting in 2012, exclusively for 5-star plans the open enrollment window was widened to allow enrollments at anytime during the year. We estimate that, due to the reform, the within-year enrollment of 5-star plans increases, but their risk pool does not worsen and actually slightly improves. Correspondingly, when estimating impacts on the market-level distribution of various plan features, we find lower premiums and decreased coverage generosity for 5-star plans relative to competing plans, leading us to argue that 5-star plans became more appealing for most beneficiary, but less so for those in worse health conditions.

JEL: I11, I18, L22, D44, H57.

Keywords: health insurance; risk selection; vendor rating; Medicare

*Decarolis, Boston University (fdc@bu.edu); Guglielmo, University of Wisconsin Madison (aguglielmo@wisc.edu). Decarolis is grateful to the Sloan Foundation (grant 2011-5-23 ECON) for financial support. We are also grateful for the comments received from Pierre Andre Chiappori, Randy Ellis, Amit Gandhi, Jesse Gregory, Kate Ho, Tim Layton, Mike Riordan, Alan Sorensen, Chris Taber and Bob Town and from the participants at the seminars at Boston University, Columbia University, Stanford University, Università Bocconi, University of Wisconsin Madison where earlier versions of this paper were presented.

The behavior of insurers is a crucial component of the functioning of any insurance market. Understanding such behavior is thus key to evaluate reforms like the creation of the healthcare marketplaces under the Patient Protection and Affordable Care Act (PPACA) and the growingly privatized provision of Medicare throughout the Part C and Part D programs. The question of how competition works in environments with potential risk selection (either advantageous or adverse) is, however, still unsettled from a theoretical perspective[1] and there is still much to be learned on the complex interaction between market power and selection.

This paper contributes to this understanding by providing one of the first empirical accounts of how insurers adjust plan features when faced with a potential change in selection. Evidence on this type of behavior is hard to collect because it is rare to observe changes in selection risk within a market. Furthermore, even when selection risk changes for a subset of plans, it is often impossible to consider the remaining plans as a valid comparison group since the equilibrium in the whole market is affected. Our analysis overcomes this difficulty by exploiting the combined effects of a Medicare reform that altered the potential selection risk of the highest quality Part C and D plans and the geographical dispersion of such plans over the US territory. This allows us to separately observe treated and control geographical markets both before and after this policy change, thus allowing a differences-in-differences approach. Our main finding is that the policy triggered a response by insurers that involved not only changing premiums, but also adjusting generosity of coverage and quality of service.

The starting point of our analysis is a Medicare reform changing the open enrollment period for a subset of plans. As in most insurance markets, beneficiaries select their Part C or D plan for coverage year t during a window of time in the fall of year $t-1$. However, starting with the enrollment year 2012, a reform allowed enrollees to switch to 5-star Part C or D plans at any point during the year. The Medicare plan rating system ranks plans from 1 to 5 stars and 5-star plans are the highest quality ones. Despite the official motivation offered to justify this new open enrollment policy (known as "5-star Special Enrollment Period" or "5-star SEP") was to foster enrollment into high quality plans, this reform exposes 5-star plans to an evident selection risk: enrollees could initially select cheap plans and then move to expensive 5-star plans with generous coverage only after being hit by health shocks.

[1] See, for instance, Mahoney and Weyl (2014), Azevedo and Gottlieb (2015) and Shourideh et al. (2015).

The impact of this reform is clearly linked to the presence of 5-star plans in the market. Due to regulatory reasons, the US territory is segmented into geographically separated markets both for Part C - where insurers offer plans at the county level - and for Part D - where insurers offer plans at regional level. Since not all geographical markets have 5-star plans, the heterogenous presence of these plans implies that some markets were affected by the reform while others were not. Our empirical strategy exploits this difference, together with the robustness to manipulations of the star rating in the first years after the policy change, to identify the causal effect of the policy on various features of the plans supplied.

The empirical analysis proceeds in two steps. First, we assess whether enrollees are responding to the 5-star SEP. The most direct effect that we seek to uncover is whether consumers move to 5-star plans during the year. We use Center for Medicare and Medicaid (CMS) data on monthly enrollment at the contract level to assess whether 5-star plans experience a change in their within-year enrollment (measured as the difference between the enrollment in December and in January of the same year) relative to comparable plans. For our baseline difference-in-difference models, the comparison plans are the 4 and 4.5 star plans offered in markets where no 5-star plans are offered. As explained below, this choice of control plans, aside from ensuring that both treatment and control plans are the top star-rated plans in their markets, also serves to limit the bias in identification that could result from a simultaneous reform of plan payments. Our main finding is that, for Part C plans, the 5-star SEP is associated with a positive and significant increase in the within-year change in enrollment ranging from 7 percent to 16 percent of the contract enrollment base.

We then look at enrollment changes across the years. While the previous results show that consumers respond to the most direct effect of the policy, a more sophisticated response would entail exiting 5-star plans during the open enrollment period and rejoining them during the year when hit by a health shock. The data, however, does not provide evidence in support of this behavior. Finally, the last element of the first part of our analysis looks at changes in plans risk pool across years. For both Part C and D risk score measures, we find clear evidence that the 5-star plans risk pool did not worsen in response to the policy. Under most model specifications, we estimate a positive, albeit small improvement in 5-star plans

risk pools. Hence, the first part of the analysis indicates that the 5-star SEP successfully achieved the goal of fostering 5-star plan enrollment, without worsening selection concerns.

The second part of the analysis explores the mechanisms through which this happened, emphasizing the role of insurers behavior. We begin by describing how two large insurers offering 5-star plans, Kaiser and Humana, modified features of the plans offered in terms of both premiums and coverage generosity. Motivated by this descriptive evidence, we then address the issue of causally estimating the effects of the 5-star SEP on a broad array of plan features. The methodology that we use is a quantile-based difference-in-differences analysis in the spirit of Chetverikov, Larsen and Palmer (2015). Relative to the first part of our analysis, this second part differs in terms of the unit of analysis: instead of looking at 5-star plans, here we analyze distributional changes in the whole market. Thus, we are able to assess how the distribution of premiums, generosity and quality measures in the treated geographical markets changes in response to the 5-star SEP relative to control markets.

We find a tendency for premiums to increase in the medium-low end of the premium distribution and to decrease in the medium-high end of the distribution, where 5-star plans are located. Similarly, plan generosity - measured, for instance, via the Part C maximum out of pocket (MOOP) - remains unchanged for plans in the high end of the MOOP distribution, but tends to worsen for plans at the low and medium end of the distribution. Since 5-star plans are among those with low MOOP, this result implies a worsening of their generosity. We find the same result when looking at the Part C plan out of pocket cost (OOPC) of enrollees in poor health. For enrollees in excellent health, instead, the 5-star SEP does not cause changes at any quintile of the Part C OOPC distribution. Interestingly, we observe that among the coverage generosity measures, the only one for which 5-star plans improve relative to competing plans is the deductible. Given the importance of the Part D deductible for beneficiaries switching to 5-star plans during the year, we argue that this is coherent with a strategic response by insurers.

We perform the same analysis on various other plan features, some entailing soft quality measures that are often hard to observe. For them we exploit the individual quality measures behind the star rating system and evaluate whether insurers also altered these dimensions

of plan quality. We find that the distribution of various quality measures (i.e. health care quality, customer service, drug access, etc.) widens up: plans at the higher end of the distribution experience an increase relative to plans at the lower end of the distribution. Thus, 5-star plans do not seem to worsen in terms of the soft quality measures determining the star rating. Overall, the evidence from the second part of our analysis indicates that the insurers response entailed making 5-star plans more appealing than competing plans for most consumers (by improving quality and lowering premiums and deductibles), but less so for the less healthy enrollees (by worsening coverage generosity).

Finally, to better understand the interaction between competition and the effects of the 5-star SEP, we repeat the analysis separately for markets where there is a monopolist insurer for 5-star plans and for markets where there is competition (duopoly) in the supply of 5-star plans. The most interesting result is that competition among 5-star insurers seems to exacerbate the extent to which these insures try to cream skim the market by worsening their plan generosity. Consumers in duopoly markets are more likely to be negatively affected by the 5-star SEP: in addition to a more substantial increase in the MOOP, they do not experience lower premiums or improvement in soft quality measures that accompany the 5-star SEP reform in monopoly markets.

From a policy perspective, our results offer several contributions. First, they show that insurers have the ability to design plan features even in the context of the tightly regulated Medicare market. Second, insurers' behavior involves not only changes to easily observable features - like premiums - that a regulator can target, but also harder to measure soft quality features. Third, the sophisticated reaction by insurers dramatically changes what a policy like the 5-star SEP could have produced. Insurers' sophisticated behavior was likely a key component of the success of the 5-star SEP reform, but it also underscores the complexity of designing rules capable of steering the market toward the goals set by the regulator.

Related literature

This study contributes to different strands of the literature on both demand and supply of health insurance, especially within the context of privatized Medicare. Within the broad literature that has looked at plan demand, our emphasis on plan switching is shared by a

few recent studies, like Ketcham et al. (2012), Ketcham, Lucarelli and Powers (2014), Ho, Hogan and Scott Morton (2014), for Part D and Nosal (2012) and Miller (2014) for Part C. Another closely related, albeit different, study is Madeira (2015) which exploits the 5-star SEP in the Part D market to study plan switching with regard to the presence of behavioral biases in enrollee choices. Finally, the relevance of the star rating system for plan choices has already been stressed by Abaluck and Gruber (2013), for Part D, and Reid et al. (2013) and Darden and McCarthy (2014), for Part C.[2]

On the supply side, our paper is one of the first studies providing empirical evidence directly relevant for the long standing, but still ongoing, theoretical debate on competition in selection markets.[3] Our focus on insurers response to the potential selection changes is related to Polyakova (2014) and Ho, Hogan and Scott Morton (2014). Both studies find evidence of selection in Part D and discuss how that interacted with the plan offerings by insurers. Self selection also entails a potential for strategic insurers to try to cream skim the market and, indeed, Carey (2014) finds evidence of this behavior in Part D. In Part C, older studies found evidence of this phenomenon (Cao and McGuire (2003) and Batata (2004)), but more recent studies have argued that risk adjustment drastically reduced it (McWilliams, Hsu and Newhouse (2012), Newhouse et al. (2013) and Brown et al. (2014).)

Our study also contributes to the analysis of how insurers respond to regulation. Thus, it is also related to other recent empirical studies that address this issue in the context of Medicare, like Decarolis (2015) for Part D and Geruso and Layton (2015) for Part C. Finally, our analysis of how insurers affect soft quality measures of the offered plans is related to the issue of the public disclosure of quality measures analyzed in Glazer and McGuire (2000).[4]

[2]In this respect, our paper is also related to a vast literature in health care that looks at whether public disclosure of quality measures has been effective in better matching patients with products and providers. See, for instance works on the impact of report cards on insurance plans (Dafny and Dranove (2008), Jin and Sorensen (2006)), fertility clinics (Bundorf et al. (2009)), hospitals (Cutler, Huckman and Landrum (2004)) and individual physicians (Wang et al. (2011)).

[3]This debate originates from the seminal studies of Akerlof (1970) and Rothschild and Stiglitz (1976). Several recent studies, Mahoney and Weyl (2014), Azevedo and Gottlieb (2015), Farinha Luz (2015) and Shourideh et al. (2015), exemplify well how the theoretical literature is still hotly debating this issue.

[4]Related applications involve the cases of how cardiac surgery report cards led to selection by providers David Dranove and Satterthwaite (2003) in New York and Pennsylvania and the similar evidence on the Nursing Home Quality Initiative by Werner et al. (2009) and Lu (2012).

I Baseline Framework

This section presents a baseline framework to discuss the potential effects of the enrollment reform in an environment with heterogenous consumers. While preference heterogeneity is a key motivation for the private delivery of Medicare, its presence does not necessarily imply risk selection. Indeed, we consider an environment where adverse selection emerges only after unrestricted enrollment into a subset of plans becomes feasible. We graphically describe through Figure 1 the equilibrium market shares of our simple model and leave for the web appendix the algebraic characterization.

Consider a market with two firms, A and B, each offering one insurance plan. Assume that firms can only set their plan premium. For each firm, the cost of enrolling a consumer is zero if the consumer is healthy and c if he is sick. Consumers choose between these plans or an outside option, Traditional Medicare (TM). For all consumers, let μ be the value of private insurance (A or B) relative to TM.[5] At the time of choosing, each consumer i also knows that he will be either sick, $h_i = 1$, or healthy, $h_i = 0$, and that, for all i, $h_i \sim Bernoulli(\gamma)$. Without loss of generality, assume A is preferable to B for sick enrollees and, in particular, let b be a vertical (i.e., commonly agreed) measure of the quality of plan A for sick enrollees. Finally, consumers are heterogeneous in how they value the benefit of insurance: let $\alpha_i \sim U[0,1]$ be such valuation and let it be known to consumers.

The two panels of Figure 1 describe the equilibrium market shares under two scenarios. In the first, consumers must choose between A, B or TB before learning their health status and plan switches are not allowed afterwards. In this case, the expected utility for consumer i before observing h_i is: $u_i = -h_i$ if in TM, $u_i = \mu - p_B + \alpha_i$ if in B, and $u_i = \mu - p_A + h_i(\alpha_i + b) + \alpha_i$ if in A.[6] The outside option, TM, is most appealing to those with low α and, as α increases, so does the value of A relative to B. As illustrated in the top panel of Figure 1, we have two indifference points: one separating consumers that choose B from those choosing TM ($\alpha_{B>TM}$) and the other separating consumers that choose A from those that choose B ($\alpha_{A>B}$). These cutoff points define the plans demand and their exact location

[5] A $\mu < 0$ captures the negative utility from the restricted network characterizing private insurance.

[6] The utility of TM is normalized to zero for sick enrollees and that of B is set to full insurance. Many alternative formulations leaving the plan ordering unchanged result in qualitatively similar results.

is an equilibrium outcome determined by the ensuing optimal premiums.

The second scenario that we consider entails the possibility of plan switching. To illustrate the effects of allowing consumers to switch to the high quality plan without entering the complexities of a fully dynamic model, consider now the setup above with the following modification of the timing of choices. Insurers set premiums aware that consumers in TM or B will be allowed to switch to A after observing the realization of h. Consumers choose a plan or the outside option aware of their own value, α_i, but unaware of their health status h or that they will be able to switch to A. Then h is realized and consumers learn they can switch to A by paying a switching cost $\phi_{TM\to A}$ or $\phi_{B\to A}$ respectively, plus any price differential to p_A. Switching occurs and, finally, market shares and profits are realized.[7]

The bottom panel of Figure 1 describes the equilibrium in this model. Compared to the case without the policy intervention, the $\alpha_{B>TM}$ and $\alpha_{A>B}$ cutoffs move due to the different equilibrium premiums. Moreover, two new cutoffs points exist determining which enrollees of TB and B will switch to A. The location of these two new cutoffs points, $\alpha_{TM\to A}$ and $\alpha_{B\to A}$, shows that among the enrollees of TM (or B) it is the subset with the highest values of α that will potentially move. Since switching is dominated for healthy enrollees, those switching are the sick ones, so a share of γ enrollees form both TM and B.

This simple framework allows us to illustrate several interesting effects of the policy. First, although the policy allows switches only to firm A, in equilibrium both A and B adjust their prices relative to the case without the policy. Depending on the model parameters, prices and profits can either tend to converge or diverge. Second, the policy creates an adverse selection problem since some of those who are sick switch to A. The average cost without the policy is $c\gamma$ for both A and B, while under the policy it becomes higher for A and lower for B.[8] Third, switching costs play an important role as, without them, major switches of sick enrollees to A could make the market unravel. Fourth, insurers have an incentive to engage in plan design manipulations: by altering b, firm A would be able to

[7]This model is likely more adequate to capture the initial response in the market after the introduction of the 5-star SEP, than to characterize its medium run impacts on consumer and insurer behavior.

[8]This can be illustrated through a numerical example. Suppose that $\gamma = .45$, $\mu = -.45$, $b = .6$, $\phi_{O\to A} = .7$, $\phi_{B\to A} = .3$ and $c = .1$. Then, without the policy each firm has an average cost per enrollee of $c\gamma$=0.045 and the two prices are $p_A^* = 0.420$ and $p_B^* = 0.074$. With the policy, prices are $p_A^* = 0.474$ and $p_B^* = 0.078$. At these prices, the average cost per enrollee in firm A increases to 0.052, while the one for B declines to 0.007.

better control the potential adverse selection. Finally, although not explicitly analyzed in this framework, it is evident that additional institutional features like a subsidy for the high quality plan or the usage of risk adjustment are potentially important elements capable of altering the equilibrium response of insurers. In particular, both a subsidy on plan A and a risk adjustment mechanism equalizing the costs between A and B could induce firm A to exploit plan switching behavior to bolster its market share without worring about selection.

II Institutions: Rating System and Policy Changes

The Medicare Part C and D programs share several organizational features. Both programs entail Medicare beneficiaries choosing a plan from a menu of plans offered by private insurers. Detailed regulations, mostly from the Center for Medicare and Medicaid Services (CMS), contribute to the determination of both the types of plans offered and their premiums. The two programs, however, differ along many dimensions: Part C is a privately provided alternative to TM. Thus, plans must cover Medicare Part A and Part B benefits (except hospice care), but can offer additional benefits.[9] Part D, instead, is a program with voluntary enrollment that provides coverage for prescription drugs. For Part C, nearly all Medicare Advantage (MA) plans also include Part D benefits.[10] However, enrollees of TM can obtain Part D benefits by enrolling in stand alone Part D plans know as Prescription Drug Plans (PDP). This section describes three key regulatory aspects for this study: plan rating systems and the reforms linking ratings with enrollment periods and subsidies, respectively.[11]

A. Rating Systems for Part C and D

To help beneficiaries select plans and to monitor the market, CMS rates plans on a 1 to 5 scale, with 5-stars indicating the highest quality. More precisely, CMS assigns ratings at the contract level and so every plan covered under the same contract receives the same rating.[12]

[9]Medicare Part A includes inpatient hospital, skilled nursing, and some home health services. Medicare Part B includes physicians' services, outpatient care, and durable medical equipment.

[10]The subset of plans offering both Pat C and D coverage are usually indicated as MA-PD plans. With a slight abuse of notation we will refer to all Part C plans as MA plans.

[11]Newhouse and McGuire (2014) and Duggan, Healy and Morton (2008) are recent studies discussing more broadly the institutional aspects of Part C and D respectively.

[12]In Part C, a contract is a particular product type (HMO, PPO or Private FFS) covering a specific service area (i.e county or group of counties), while a plan is finer specification of benefit package that include type

Information about plan performance has been collected since 1999, but the introduction of the star rating system started only in 2006 for Part D and in 2007 to Part C.

The details concerning the rating system are fairly complex and have changed over time. The essential aspect is that different data sources (enrollees surveys as well as CMS administrative data, and data from plans and other CMS contractors) are used to collect information on a broad set of indicators. The process through which CMS calculates the star rating involves several steps. At the most disaggregated level there is a large number of "individual measures," which are aggregated into a smaller number of "domain measures" and finally into the "summary rating" through a complex weighting system.[13] Table 1 reports the domain measures: for Part C, they cover features such as clinical quality, patient experience, and contractor performance; for Part D, they cover cover aspects such as call center hold time, members' ability to get prescriptions filled easily when using the drug plan, and plan fairness in denials to members' appeals. The overall rating, expressed in a 5-Star scale with increments of half a star, is released every year in October on the CMS Plan Finder web site.

A notable feature of the rating system is that it is hard to manipulate for insurers, especially in the short run. There are at least three reasons for this: first, CMS changes the system form year to year in terms of both which parameters are evaluated and how they are aggregated into the overall rating. This aspect is particularly salient given the large number of different measures that are evaluated, as shown in Table 1. Second, ratings on individual measures are assigned by comparing the *relative performance* of each contract to the entire population of contracts so that manipulations would require detailed information on all competing contracts. Third, and most crucially, the rating is based on lagged data: year t ratings (released on October of year $t-1$) use data for the period between January of year $t-2$ and June of year $t-1$. Thus, to ensure our results are not affected by rating manipulations, we will focus exclusively on the first two years after the enrollment reform.

Very few contracts obtain the 5-star maximum. In 2012 and 2013, for instance, out of the

of coverage, premium, copayment, etc. In Part D, a contract typically indicates a drug formulary and, then, each plan within the contract applies different conditions (for instance copays) to the same formulary.

[13]More precisely, for PDP and MA plans not offering Part D, the summary rating is also the overall rating. For MA plans, the Part C and D summary ratings are combined to obtain an overall rating. A more complete description of the process through which CMS calculates the star rating is detailed in the web appendix.

34 geographical regions into which Part D divides the United States, only 2 regions (region 3, New York, and region 25, formed by 7 midwest states) had a 5-star PDP. 5-star plans are more frequent among MA. However, while PDP must be offered to all counties within a region, Part C plans are offered at the county level. Figure 2 presents a heat map showing the offerings of MA plans. In 2012, 5-star plans are offered in 156 counties belonging to 17 different states and spanning almost all the U.S. geographical areas, with the relevant exception of the center-south area. This geographical dispersion of 5-star MA plans plays a fundamental role in our empirical strategy and we return to it in the next section.

B. Demand Side Reform: Plan Rating and Enrollment Periods

Generally, beneficiaries enroll in a plan from October to December of the the year before the coverage period (Open Enrollment Period, OEP) and must keep the same plan for the entire coverage year. Exceptions to the OEP, known as Special Enrollment Periods (SEPs), permit enrollees to change plans, but are typically confined to special circumstances.[14]

Starting with the 2012 coverage period, CMS introduced a new type of SEP linked to the star rating system. This reform allows all beneficiaries to enroll in a 5-star Part C or D plan at any point in time.[15] This SEP rule can only be used once per year and is available even to enrollees already in a 5-star plan, but who want to switch to another 5-star plan. Coverage with the new 5-star plan takes effect the first day of the month following the enrollment. Similar to any other enrollment request, 5-star plans must accept all applicants. The SEP is not available to enroll in a plan that does not have an overall 5-star rating, even if the plan receives 5-stars in some rating categories, or if the plan is in the same parent organization.[16] CMS has extensively advertised this new SEP rule in its communications to consumers. As regards insurers, they were publicly informed of the introduction of the 5-star SEP on November 2010. Since the next round of plan bids was in June 2011 for the menu of plans

[14]The most relevant SEPs are: (i) for change of residency, including moving to a nursing home; (ii) for low income people (dual eligible or qualifying for the LIS or for SPAPs); (iii) for people who enroll in a MA plan when they are first eligible at age 65 get a "trial period" (up to 12 months) to try out MA. This SEP allows them to disenroll from their first MA plan to go to TM.

[15]See the *2012 Newsletter* at `http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/downloads/Announcement2012final2.pdf`.

[16]There is also a special provision for which, if the enrollee uses the 5-Star SEP to enroll in either a 5-star PFFS plan or a 5-star Cost Plan, then he gets a "coordinating Part D SEP" allowing him to enroll in a stand-alone PDP, or in the Cost Plan's Part D optional benefit, if applicable.

to be offered in 2012, then we can consider 2012 as the first year from which we shall expect to see reactions in plan features driven by the policy change.

C. Supply Side Reform: Plan Rating and Insurers' Payments

Payments to insurers come mostly from various types of Medicare payments and, only in small part, from enrollees premiums, see Newhouse and McGuire (2014) and Decarolis (2014). The PPACA of 2010 reformed various aspects of the system and, crucially, introduced a link between the star rating system and payments.

This supply side reform affects exclusively Part C and, like the enrollment reform, became effective in 2012. Essentially, the reform wanted to reduce overall plan payments, but also to make payments relatively more generous for higher quality plans than for lower quality plans. For the purposes of our study, this reform implies that after 2012 per enrollee payments of 5-star plans are more comparable to those of 4 and 4.5 then to those of plans with lower ratings. In essence this is due to how this reform affects two features of the payment system.

The first is the benchmark. The benchmark is a function of what TM spends in the plan's service area. CMS determines the payment to an MA plan by comparing its "bid" (the amount the insurer requests to enroll a beneficiary in the plan) to the service area benchmark. Plans with a bid below benchmark (the typical case) receive their bid plus a rebate based on the difference between benchmark and the bid. The PPACA reform aligned benchmarks more closely with TM spending[17] and, instead of the flat 75% rebate used before 2012, introduced a variable rebate, ranging from 50% to 70%, linked to the plan star rating.[18]

The second is the bonus. Bonuses were introduced in 2012 to bolster payments for high-quality plans by proportionally increasing their benchmarks. For instance, in 2012 the bonus for 5-star plans is 5% of the benchmark. Thus, a 5-star plan with a bid below the benchmark receives a rebate equal to 73% of 1.05 times its service area benchmark. While under the PPACA bonuses were reserved for plans with 4 or more stars, CMS used its demonstration

[17] It ties the benchmarks to a percentage of mean TM cost in each county and caps them at the pre-PPACA level. These benchmarks are phased in from 2012 to 2017 by blending them with the old benchmarks.

[18] The new rebates are phased in from 2012 to 2014. In 2012, the rebate equals the sum of two-thirds of the old rebate amount and one-third of the new rebate amount. In 2013, the rebate equals the sum of one-third of the old rebate amount and two-thirds of the new rebate amount. From 2014 onward, the rebate is 70%for 5-4.5 star contracts, 65% for 4-3.5 contracts and 50% for the rest of the contracts.

authority to extend bonuses to plans with 3 or more stars. In the period that we study, benchmarks are increased by 4% for 4.5-4 star plans, by 3.5% for 3.5 star plans, by 3% for 3 star plans and plans that are too new or with too few enrollees to be rated.[19]

III Data

Our analysis is based on publicly available data released by CMS describing MA and PDP plan/contract characteristics. In addition to monthly enrollment, we observe characteristics such as Part C and D premiums, deductible, extra coverage in the gap, measures of drug generosity, risk scores for Part C and D and the star rating. For this latter variable, we have both the overall summary rating, as well as the score on each individual measure. We also use the Area Health Resource File released by the Health Resource Service Administration to assess a number of county-level demographic, economic and heath indicators.[20]

The empirical analysis in the next two sections looks separately at demand and supply effects. For the supply side, we focus on a broad spectrum of outcome measures ranging from premiums and other financial characteristics to various proxies of generosity of coverage. Among these proxies, the Part C maximum out of pocket (MOOP) is particularly relevant as it measures the maximum amount that an enrollees might spend to access in-network health care services through the plan (it includes all costs but the premium).[21] Another important and closely related variable is the out of pocket cost (OOPC) that we observe separately for the Part C and Part D components of the plan. This value, released by CMS, is obtained by simulating what would be out of pocket costs of representative beneficiaries and is available for enrollees with different health status ranging from poor to excellent health.

For the demand side, instead, we focus on three main outcome variables: the within-year change in enrollment, the across-year change in enrollment, and the risk score. The first variable, calculated as the difference in the contract enrollment on December of year t and

[19]The demonstration is expected to cost more than $8 billion, making it more costly than the combined cost of all 85 other Medicare demonstrations that have taken place since 1995. See Layton and Ryan (2014) for a first assessment of its effects.

[20]See the additional details on the datasets in the web appendix.

[21]We observe this measures starting from 2011.

the enrollment in January of year t, captures increased potential for plan switching during the year. This measure thus captures the most direct effect of the policy. We also consider the possibility of plan switching across years by calculating the difference in the contract enrollment on January of year t and the enrollment in December of year $t-1$. This latter variable can capture a strategic response by consumers: greater plan switching during the regular open enrollment period driven by the possibility of switching to a 5-star plan later. Regarding the risk score, this outcome variable is measured as the mean contract risk score, available from CMS at yearly level. Assessing changes in risk score is relevant to determining whether the composition of the enrollment pool of the contracts is affected by the 5-star SEP.

Table 3 reports summary statistics for the demand analysis sample: MA plans data aggregated at the level of contract, year and county. We conduct the analysis at contract and not at plan level both because the rating does not vary among plans under the same contract and because missing enrollment data are more common at plan than at contract level.[22] We focus on the period from 2009 to 2013 to assess the immediate response to the reforms implemented in 2012. The table reports statistics separately for the years 2009-2011 and 2012-2013, and for two subsets of contracts: contracts obtaining the 5-star rating in 2012 or 2013 (our treated group) and contracts obtaining the 4 or 4.5 rating in 2012 or 2013 and offered in counties that do not have 5-star contracts in the same years (our control group). On average 5-star contracts have higher enrollment, healthier enrollees and more generous coverage than the control group.

The summary statistics are suggestive that the within-year change in enrollment responds to the enrollment reform. The data show an increase in the within-year enrollment for 5-star contracts in the post 2011 period relative to the previous period, but not for the control group. Moreover, they suggest possible effects on the supply side offerings as well: Part C premiums tend to decline more for the treatment than for the control group, while MOOP increases for 5-star contracts relative to control contracts.

Finally, another crucial feature that the data reveal is that the 5-star SEP did not trigger

[22] A subset of our measures are available only at plan level. We aggregate them at contract level by weighting the plan characteristics by the enrollment of the plan. We tested the robustness of our results to aggregation (i.e. simple average), the results are reported in appendix.

any major entry/exit of plans. Table 2 reports (by year and insurer) the number of counties in which the plans achieving 5-star in 2012 or 2013 are offered. Comparing 2012 to 2013, it is clear that the 5-star plans did not reduce their presence. Indeed they seem to expand the number of counties served, regardless the parent organization. Our results below will offer an economic rational for why insurers were able to maintain their 5-star contracts. However, it is also relevant to point out that CMS poses limits to the exit of plans as it can impose a two year ban to a firms that retires all its contracts from MA.

IV Empirical Analysis I: Demand Effects

In this section, we provide evidence regarding the effect of the 5-star SEP on demand side responses related to beneficiaries enrollment and risk scores. We first present our empirical strategy and then discuss our main results, as well as the most relevant robustness checks. While the empirical strategies used to estimate demand and supply effects are closely related, they are not identical. We will discuss the strategy used for the supply analysis and the associated findings in the next section.

A. Empirical Strategy

To identify the effect of the 5-star SEP on demand side factors, we follow a difference-in-differences (DID) approach. For MA plans, this strategy exploits the fact, documented in Figure 2, that 5-star contracts are offered in only a subset of the US counties. We consider all contracts that achieve the 5-star rating in the period 2012-2013 as the DID treatment group (dark red areas in in Figure 2) and all contracts that achieve a 4 or 4.5 rating in the same period and are offered in counties without any 5-star contract as the control group (light red areas in in Figure 2). The regression model that we estimate is:

$$Y_{ict} = a_c + b_t + c_i + \beta D_{it}^{5S} + \varepsilon_{ict} \tag{1}$$

where i indicates the contract, c the county and t the year. The coefficient of interest is β, the effect on the dependent variable of a dummy equal to one for 5-star contracts after 2011, conditional on fixed effects for the county (a_c), time (b_t) and contract (c_i). Various

extensions of this baseline model are presented below.

There are challenges to interpret β as the causal effect of the policy change. As usual in any DID study, the first and foremost concern is to select an adequate control group. In our setting, 4 and 4.5 star contracts offered in counties that do not have any 5-star plan are a nearly ideal control group. Clearly, both the control and the treatment contracts are similar as they are the top quality contracts offered in their respective counties. Furthermore, as discussed above, contracts in the control group face similar financial incentives of those in the treatment group, thus allowing us to identify the effect of the 5-star SEP policy reform separately from any other effect produced by the simultaneous payment reform.

As shown in Table 3, however, treatment and control groups differ along several observable characteristics, like size of the enrollment base and features of the enrollment pool. Indeed, although Figure 2 reveals that the 5-star plans are scattered across many different counties, this does not ensure their assignment to counties is random. We have two arguments to address this concern, the first is that, for the three reasons explained in section 3, it is hard for insurers to perfectly control their rating so that the difference between a 4-4.5 and a 5-star plan is likely quasi-random, at least for the period object of analysis.[23] Second, to the extent that the selection into the treatment state is based on observable characteristics, we have a rich set of covariates that permits us to control for this threat. Thus, as a robustness check for our baseline estimates we use a matching DID strategy, where the control group observations are selected to match the characteristics of the treatment group.

Therefore, our identification strategy rests upon the fact that the assignment of the treatment relative to the control status is quasi-random within the union of the counties marked in dark and light red in Figure 2. Since the regulation separates the geographical markets, an additional benefit of this strategy is that, by selecting treatment and control groups from different counties, it avoids contamination issues.

B. Effect on Enrollment

The first outcome variable that we analyze is the contract-county within year enrollment

[23]We considered supplementing our DID strategy with a discontinuity design by restricting the analysis to treated and control plans with ratings close to the 4.75 star cutoff separating 4.5 star plans from 5-star plans. However, the paucity of plans around the cutoff renders this type of analysis infeasible.

change. The yearly trend in this variable is shown by Figure 3 separately for the treatment and control groups. There is a clear increase in the number of enrollees for the treatment group after the introduction of the SEP, as already highlighted by the statistics in Table 3.[24] Even before 2012, there is a growing trend for the treated group, relative to a declining path for the control group. Although for both groups these year-to-year changes are not statistically significant, thus limiting potential bias in the estimate of β, we will also report estimates including group-specific time trends in the DID model specification.

Panel A of Table 4 displays our baseline DID estimates. The dependent variable is the enrollment change between December and January both in levels (Columns 1-4) and in percentage terms (relative to the January enrollment base) (Columns 5-8). We estimate 4 specifications: the odd numbered columns include county and year fixed effects, the even numbered columns add contract fixed effects. Columns 3, 4, 7 and 8 add also a linear trend at state and treatment level. The 5-star SEP has a large and statistically significant effect on the within year change in enrollment. In our baseline specifications, columns 1 and 2, the number of enrollees increases on average by 225-235 enrollees. This effect is quite substantial, if, for instance, we compare it to an average value of the dependent variable in the pre treatment period of 386 enrollees. When including time trends, the effect is still present, but its magnitude is attenuated. Columns 5-8 report analogous estimates for the percentage enrollment change. This variable allows to normalize the enrollment changes by the existing enrollment base. The estimates that we obtain range from 7% to 9% in the baseline specifications and from 15 to 16% when including time trends.

It is informative to know in which month of the year enrollees use the SEP. Thus, we consider complementing the above estimates of the December minus January enrollment change with analogous estimates for the other months preceding December. In Figure 4, we plot the estimates obtained for the same specification as in model (2) of Table 4. The effect on enrollment of the SEP appears linearly increasing over time up until October and then it flattens out. Thus enrollees seem to use the new SEP uniformly over most of the year.

[24]The presence of an upward trend for the treatment group, can be explained by a number of factors. CMS has been strongly advertising to enrollees the Star rating as measure of quality and that could have affect the increase in the enrollment overtime.

We conclude this section by describing various robustness checks presented in the remaining panels of Table 4. To assess the sensitivity of our estimates to the choice of the control group, we use a twofold approach. First, we construct a sample of comparable contracts using propensity score matching. We use an extensive list of socio-economical, demographic and health indicators to predict the probability that a county has a 5-star contract in the 2012-13 period. Then, we restrict the control group to those contracts in counties belonging to the common support of the propensity score between the treatment and the control groups.[25] Second, we further restrict the control group to include only contracts that achieve at least the 4 star level in both 2012 and 2013, thus selecting contracts that are more likely to be comparable with the 5-star contracts. We report the findings in Panel B and C of Table 4. Overall, the policy change maintains its positive and statistically significant effect.[26]

To further assess the robustness of our estimates, Panel D of Table 4 reports the results of a placebo test. We repeat our analysis as if the 5-star SEP was introduced in 2011 instead of 2012. To avoid potential spillovers from the true SEP, we narrowed our exercise to the enrollment periods from 2009 to 2011. Panel D shows that, in our first two specifications, the simulated SEP has a positive and statistically significant effect on the within year enrollment change, but this effect vanishes once we control for time trends. Furthermore, we do not find a statistically significant effect of the placebo SEP on the percentage change in enrollment.

In Table 5, we repeat the whole analysis using as dependent variable the enrollment change across years. As explained earlier, a negative effect of the policy would be compatible with consumers acting strategically. Our estimates, however, fail to show the presence of such strategic behavior. The coefficient that we estimate is not statistically significant for most of the regression models and, when it is significant, it has a positive sign.

Finally, additional robustness checks for both the within and across years enrollment changes are reported in the web appendix. There we also report the analysis for Part D plans. While no supply side changes to the payment system occurred for PDP - thus making

[25]We tried various specification for the propensity score and results were broadly comparable to the reported specification. Further details as well as the probit estimates are reported in the web appendix.

[26]In the baseline model, the effect of the SEP ranges between 146 and 241 enrollees. The results for the percentage change in enrollment indicate an effect ranging between 8% and 22% in the matched sample. Once we restrict the control group to 4 star in both 2012 and 2013, we still observe a positive effect, between 5% and 12%, even if not statistically significant.

easier the selection of a control group - performing inference is problematic since only 2 out of the 34 regions are treated. With this caveats in mind, our Part D estimates are broadly in line with the findings of a positive and significant effect of the 5-star SEP on within year enrollment change. For Part D, we also find some evidence of a negative, although not statistically significant, effect of the 5-star SEP on enrollment switches across years.

C. Effect on Risk Score

The final piece of our demand analysis focuses on interactions between the 5-star SEP and the contracts risk pools. Here we analyze whether the 5-star SEP also causes a worsening of the risk pool of 5-star contracts. The two dependent variables on which we focus are the yearly average contract risk score that CMS releases separately for Part C and D. Each one of the two measures is normalized to 1 for the average risk of a TM enrollee, the higher the risk score the higher the risk (and the potential cost) of the enrollee.

Figure 5 shows the evolution over time of the risk score for 5 and 4-4.5 star contracts. For both risk score measures, there is a similar, descending trend in both the control and treatment groups. The decline in the latter, however, appears slightly more pronounced. This visual evidence is confirmed by the DID regression analysis reported in Table 6. The 5-star SEP has a negative and highly statistically significant effect on the risk score for both Part C and D. The effect, however, is small being in the order of 10 percent of a standard deviation of the dependent variable. To better quantify these effects, for Part C this is equivalent to reducing the expected average cost per enrollee by \$0.02 for each dollar spent.

This improvement in the risk pool is surprising given the significant increase in within-year enrollment. The next question is thus the robustness of this result. Robustness checks analogous to those performed for the enrollment outcomes broadly confirm the result.[27] However, a concern specific to the risk score variable is whether the timing with which it is recorded could confound our interpretation. The measure that we use is an yearly average. Could it be that this variable is unable to capture in a timely manner the high risk of those joining 5-star plans? The annual average risk score for a plan is built up by taking all of the individual-level risk scores and averaging them. So, when new enrollees join during year t,

[27] See results in the web appendix.

the risk scores of those enrollees will be factored into the year t average risk score. Moreover, we know from Geruso and Layton (2015) that insurers are extremely proactive in adjusting upward the risk score of their enrollees. This all suggests that our measure is adequate.

Nevertheless, the lag can be in how often the individual-level risk scores are updated. In 2013, an individual's risk score is based on his health status (diagnoses) from 2012. Thus, if an enrollee who used to be healthy switches to a 5-star plan immediately after becoming sick, our measure might be able to capture his higher risk only an year after the switch.[28] To account for this issue, we exploit the fact that we observe two years of data since the inception of the policy and repeat the DID estimates iteratively dropping from the sample one of the two post-policy years. Our expectation is that, if the negative estimate in the risk score regressions is driven by a lag in how the score is recorded, we will likely find that using exclusively 2013 as the post-policy year should lead us to find less negative, if not even positive estimates relative to when we use only 2012 as the post-policy year. The new estimates are reported in the latter two panels of Table 6. In Panel B we drop 2013, while in Panel C we drop 2012. Both sets of estimates confirm that the negative sign of the coefficient. Moreover, although the magnitudes are similar, there is a tendency for the Panel C estimates to be larger in magnitude than those in Panel B. Hence, these results confirm that the risk pool of 5-star plans improved and it is not a spurious correlation driven by lagged a response in the risk score measures.

D. Discussion

Taken together, the findings on enrollment and risk score offer a nuanced picture of how the market responded to the 5-star SEP. Enrollees switch to 5-star plans during the year, but the risk pool of 5-star plans, instead of worsening, slightly improves. This fact could be explained through a combination of high risk consumers already being enrolled in 5-star plans (i.e., before the SEP reform) and sufficiently high switching costs that lock in enrollees

[28] A more subtle problem could, in principle, involve new Medicare enrollees. Enrollees who are enrolling in Medicare for the first time (either FFS or MA) have no diagnoses, so their risk scores are based on age/gender only and are not particularly indicative of health status. After they have been in Medicare for a full calendar year, their risk scores switch to being based on diagnoses instead. However, since new Medicare enrollees aren't actually affected by the reform we are studying since they could join any plan during any month of the year (as long as it is the first month they enroll in Medicare), so this should not be a concern for our analysis.

to their plans during the OEP. Hence, although the enrollees that switch have higher risk relative to the ones that stay in their plan, these switchers have nevertheless a lower risk than the consumers already enrolled in 5-star plans.

Figure 6 shows evidence compatible with this argument. The figure is constructed by separating contracts between those that lose and those that gain enrollees during the year and then, separately for the two subsets of contracts, calculating the average risk score (weighting contracts by their share of switchers in-flow or out-flow). We find that the out-flow tends to be from lower risk plans, while the inflow is toward higher risk plans.[29]

A different, but not mutually exclusive explanation is that 5-star plans are attracting enrollees that are not the worst risk ones in their original plans. An interesting finding in this respect is shown by Figure 7 reporting the sources of the within-year flows: TM without Part D, TM with Part D or other MA plans. The plot on the right illustrates that for the counties with 5-star plans, it is TM without Part D to suffer the largest outflow of enrollees during the year. Although this could be reconciled with the explanation above if the switching cost from TM to MA is lower than that between different MA plans, this seems rather unlikely. Indeed, what is more likely happening is that the presence of a flow of low risk enrollees from TM is the result of the strategic response of insurers to the 5-star SEP, considering that on average MA plans tend to have a lower risk score than TM (see Curto et al. (2014)).[30] This is the object of the following section.

V Empirical Analysis II: Supply Effects

The firms active on the supply side of Part C and D are many and heterogeneous. They range from large scale, nation-wide insurers like United Healthcare and Humana, to a plethora of small local companies. Almost all insurers offering Part C also offer Part D, but some major Part D insurers, like CVS Caremark, are not present in Part C. As documented in Table 2, there are seven insurers offering 5-star plans in 2012-2013. Among them, Group Health,

[29]The fact that both for out-flow and in-flow the average risk score is below 1 is explained by the fact that our analysis excludes the southern US regions, as illustrated in Figure 2, where risk scores tend to be higher.

[30]This is coherent with the findings of Aizawa and Kim (2013). MA plans are able to use advertising to attract and select, according to their risk level, new enrollees.

Humana and Kaiser Foundation are the largest insurers. However, while the 5-star plans of Group Health and Humana are offered only in a limited geographical area (Wisconsin for Humana and Oregon-Washington for Group Health), Kaiser has 5-star plans in various states: California, Colorado, Hawaii, Oregon and Washington. Kaiser's 5-star contracts have large market shares in all of these states, ranging from 12 to 48 percent of the relative markets. For Group Health and Humana, the market shares of their 5-star plans are smaller but in both cases greater than 5 percent.

The relevance and peculiarity of Kaiser, together with the presence of small local insurers on the supply side of 5-star plans suggest assessing the robustness of our previous findings to the identity of the firms involved. Nevertheless, when repeating the previous demand side analysis by iteratively eliminating each one of the seven firms offering 5-star contracts, we broadly confirm the findings described above: within year enrollment grows.

Thus, our next step is to look at what strategies these insurers implement to prevent adverse selection while, at the same time, expanding their enrollment base. For both Humana and Kaiser, the fact that both insurers also offer non-5 star plans in counties where no 5-star plan is offered by any company allows some descriptive comparisons. The most interesting aspect we find is that Humana and Kaiser seem to follow different strategies. Comparing the periods before and after the 5-star SEP, Humana's 5-star plans offered in Wisconsin lower their generosity (the average MOOP grows from \$3,400 to \$6,260), substantially more than what done by both the 4.5 star plans also offered in Wisconsin (the average MOOP grows from \$4,500 to \$6,331) and the 4.5 star plans offered in other Midwest counties (the average MOOP grows from \$3,952 to \$4,431). In the same period, the average premium of 5-star plans registers a small increase, but in line with that of the 4.5 plans. For Kaiser, instead, we can compare its 5-star plans with the 4.5 star plans it offers in Georgia. We observe that generosity remains nearly identical for both the 5-star plans (the average MOOP goes from \$3,200 to \$3,230) and 4.5 star plans (the average MOOP remains identical at \$3,400). Average premiums, however, decline slightly more for 5-star plans than for 4.5 star plans (Part D premiums decline from \$11 to \$9 for 5-star plans, while they increase from \$1.5 to \$2 for 4.5 plans; Part C premiums, instead, remain almost identical).

This descriptive evidence is suggestive that insurers response to the increased selection risk involves both premium and generosity dimensions. To draw more coherent conclusions about such responses, however, it is strictly necessary to take into account how not only 5-star insurers, but also their competitors reacted to the policy change. Non 5-star insurers operating in markets with 5-star plans are at risk of losing enrollees during the year. Moreover, they might face a worsening of selection if 5-star plans increase their cream skimming activity to limit the potential risk worsening. This type of equilibrium responses are likely the most interesting aspect induced by the SEP reform and to study these effects we describe below an empirical strategy that aims to detect them.

A. Empirical Strategy

The empirical strategy that we pursue in this part of the study is a form of DID, but it differs in two crucial dimension from the previous demand side analysis. First, while before the unit of analysis were the contracts, here the unit of analysis is the county. The key insight from the previous discussion is that all contracts in a county with a 5-star contract can respond to the SEP reform. Thus, we label counties with at least one 5-star plan in either 2012 or 2013 as treated. We label as control counties those having highest starred plans that are 4 or 4.5 stars.

The second difference is that, to capture the changes in how the overall market readjusts, we pursue a quantile-based DID analysis. This allows us to evaluate changes along the whole distribution of each one of the dependent variables that we will consider (premium, deductible, etc.). The goal is to understand how the SEP affects the nature of competition within a market. For example in the case of the premium, a 3 star contract with a low premium and a 5-star contract with an high premium would probably have a different reaction to the SEP, and analyzing different percentiles of the premium distribution within a market can be more informative than just focusing on the mere average effect. Following Chetverikov, Larsen and Palmer (2015), we implement this strategy by estimating the model:

$$Y_{ct}(\tau) = a_c(\tau) + b_t(\tau) + \beta(\tau) \times 5StarCounty_{ct} + \varepsilon(\eta_{ct}, \tau) \tag{2}$$

where c is the county, t the year and τ the quantile. Y_{ct} are the deciles of the various contracts

characteristics we observe. The coefficient of interest is β, the effect on the dependent variable of a dummy equal to one after 2011 and only for counties with 5-star contracts, conditional on fixed effects for county (a_c), and time (b_t). As shown in Chetverikov, Larsen and Palmer (2015), this approach permits us to estimate distributional effects when a group (i.e., county) level treatment is correlated with a group unobservable factor. The assumptions required for the validity of this strategy are the same of the standard DID framework.

B. Baseline Results

The plots of Figure 8 summarize our findings for each of the plan characteristics analyzed. Plot (a), for instance, reports the effect of the policy change on the Part C premium. The plot contains a great deal of information: The solid, dark line is drawn using the 19 regression coefficients, $\beta(\tau)$, estimated separately for each one of the quintiles of the Part C premium distribution. The two slid lines around it show the 95 percent confidence interval. This plot reveals that the policy change is associated with a premium increase at the lower end of premiums (up until the third decile) and with a premium decrease in the top end of the premiums (starting from the seventh decile). The decline is about $20 for plans at the 90th percentile of the distribution. The plot also describes where 5-star plans are located within the Part C premium distribution. Small squares and circles are used to mark the fraction of 5-star plans present at each decile of the distribution: squares measure the share of 5-star plans in the pre-policy period, while circles measure them in the post-policy period. In terms of the Part C premium distribution, 5-star plans are mostly concentrated in the top 50 percent of the distribution, both pre and post policy.

Finally, to illustrate the usefulness of a distributional analysis, the plots also report the average effect. The dark, horizontal, dashed line shows the mean effect (with the associated surrounding lines denoting the 95 percent confidence interval) that is estimated by applying a conventional DID method, like the one used for the demand analysis. For Part C premium, this mean effect is negative but not statistically significant. The mean effect is unable to reveal the nature of the market readjustment uncovered by the distributional analysis.

Using the same logic to interpret the evidence in the remaining plots, we find a number of interesting results. First, consistently with the behavior of Part C premiums, also for Part

D we observe a slight tendency of premium increases for plans in the medium-low end of the distribution and decreases for plans in the medium-high end of the distribution (where 5-star plans are mostly located). Second, and most crucially, plan generosity - as summarized by the Part C MOOP - tends to worsen for plans at the low and medium end of the MOOP distribution, while it remains unchanged for plans in the high end of the MOOP. 5-star plans, that are disproportionately concentrated in the lowest end of the MOOP distribution, seem to respond by reducing their generosity and so do the plans closest to them in terms of MOOP.

The following plots, (d)-(g), report additional results in terms of the OOPC. It is particularly interesting to compare the estimates for the Part C OOPC of beneficiaries in poor health and excellent health. For enrollees in poor health, the evidence in Plot (d) is once again of an increase in costs for the plans at the low end of the OOPC distribution and a decline in costs for the high OOPC plans. This is not surprising given the close connection between this OOPC measure and the MOOP. For enrollees in excellent health, however, Plot (e) shows that for all deciles there is no effect. For the Part D OOPC, the results are rather different and we see an improvement of generosity for the plans that, like the 5-star ones, were already low in terms of their OOPC and a worsening of generosity for high OOPC plans. These features involve both the case of poor health beneficiaries, Plot (f), and of excellent health beneficiaries, Plot (g). A likely explanation for the different behavior of the Part C and D OOPC measures is based on what happens to the Part D deductible.

For the Part D deductible, the estimates in Plot (h) indicate that low deductible plans (like 5-star plans) reduce their deductible even further, while the deductible increases further for high deductible plans. This evidence, is likely explained by the very peculiar role played by the deductible under the 5-star SEP. If 5-star plans were to ask for high deductibles, this would reduce their appeal for every consumer considering a within year switch. On the other hand, for non 5-star plans increasing the deductible might not trigger a major loss of enrollees under the 5-star SEP since these enrollees are aware of the possibility of switching to 5-star plans.

The decline in generosity of 5-star plans is also confirmed by Plot (i) and (j) for two Part

D plan characteristics: the share of most frequently used drugs that the plan covers and the number of drugs that the plan covers without placing any utilization restrictions. For both variables, generosity improves for plans in the low end of the distribution, while it declines for plans in the medium-high end (where 5-star plans are located).

In addition to the plan characteristics considered above, Plot (k)-(m) report the effects for the individual measures composing the summary rating.[31] An interesting result revealed by these estimates is that, while the distribution of premiums and MOOP tend to converge toward the middle, the distribution of various quality measures like *health care quality*, *customer service* and *drug access* widens: plans at the higher end of the distribution experience an increase relative to plans at the lower end of the distribution. There is an apparent heterogeneity, however, across the various measures: while for *health care quality* plans at the high end of the distribution experience a positive and statistically significant effect, for *customer service* the the effect is negative essentially throughout the entire distribution.

Observing the presence of such heterogenous responses is particularly interesting as they indicate the need, stressed by Glazer and McGuire (2000), to broaden the view of the margins along which insurers compete. The fact that, relative to non 5-star plans, the financial generosity of 5-star plans worsens, but their soft quality measures improve indicates that a sophisticated type of cream skimming might be happening. These soft quality measures might indeed be positively associated with advantageously selected consumers who care about both being healthy and obtaining high quality services from their plan. This can further help to explain the previous evidence in terms of risk scores slightly improving for 5-star plans. Thus, it is informative for descriptive purposes to apply the quantile based DID also to the Part C and D risk scores measures. These results are reported in Plots (n) and (o). For Part C, we observe that risk scores in the middle-upper end of the distribution tend to slightly decline, while they remain unchanged in the lower end. For Part D, the effect is mostly negative for the portion of the distribution where 5-star plans tend to concentrate, but the effect is typically non significant for most of the percentiles.[32]

[31] As stated earlier, the summary rating uses lagged individual measures. Thus, to perform our analysis on the response of individual measures up to 2013, we use the individual measures released through 2015.

[32] In the web appendix, we report the quantile analysis for matched samples. We use the same procedure - matching on county characteristics - described before. The results are similar to those discussed above.

C. Markets with 5-Star Contracts Monopoly or Duopoly

As discussed at the beginning of this section, counties where 5-star plans are present have either one or two insurers offering these plans.[33] The distinction between markets with 5-star plan monopoly and duopoly is potentially informative of the interactions between competition and the 5-star SEP reform. Indeed, the reform is such that even enrollees of a 5-star plan can switch plan within the year, provided they move to another 5-star plan. While irrelevant in monopoly markets, this provision can exacerbate the downward pressure on plan generosity in duopoly markets. Since the existing pool of 5-star plans typically contains high risk enrollees, for a 5-star plan receiving the riskiest enrollees of some other 5-star plan can be particularly costly.

To evaluate differences in market responses to the policy between monopoly and duopoly markets, we repeat the previous analysis on two subsamples. The six top panels of Figure 9 report the distributional effects for the monopoly case, while the latter six report the effect for the duopoly cases. The comparison of the two environments reveals that, while the decline in premiums is roughly similar, the increase in the MOOP for the portion of the distribution where 5-star plans are located is higher for duopoly than for monopoly markets. Similarly, for the customer service variable (as well as for most of the soft quality measures not reported here), duopoly markets reveal a more pronounced worsening of quality for plans that, like 5-star plans, are located in the high end of the quality distribution.

This evidence is further supported by the results involving the risk score. Both Part C and D risk scores experience a clear decline for 5-star plans in duopoly markets, but there is no statistically significant decline for the case of monopoly markets. Altogether, this evidence is suggestive that 5-star plans in duopoly markets decreased their generosity and quality more than 5-star plans in monopoly markets. On the other hand, these reductions are not accompanied by a more pronounced premium decline. Thus, relative to the pre-policy period, the effect of the 5-star SEP appears to have been more beneficial for consumers located in counties with a single firm offering 5-star plans than in areas with competition between 5-star plans. This potentially problematic effect of competition is an interesting manifestation of the complexity of making competition work in healthcare markets.

[33] We observe 7 counties for which there were more than one 5-star plan in either 2012 or 2013.

VI Conclusions

The reform that, starting in 2012, allowed consumers to switch at any point in time to the highest quality, 5-star plans could have backfired. By undermining the use of rigid open enrollment periods, a pillar of most insurance markets, this policy could have exacerbated the adverse selection faced by 5-star plans, potentially triggering premium spikes or even plan exit. We find that, although enrollees responded to the policy and 5-star plans enrollment grew, the naive prediction of a worsening of selection for these plans did not materialize.

We argue that a likely source of this result is the sophisticated response adopted by suppliers. Both 5-star insurers and their competitors responded to the new policy. The 5-star insurers lowered their premiums, while, at the same time, worsening the amount of coverage offered by their plans. This contributed to expand their enrollment base, without worsening their risk pool. The overall adjustments in the market suggest that areas where 5-star plans were offered experienced a compression in the characteristics of the available plans, with greater convergence in terms of both premiums and financial characteristics of the plans. Soft measures of plan quality also reveal a potential response along subtle dimensions that are harder for the regulator to monitor in real time.

These results, based on a clean identification strategy, empirically document key features of the Part C and D markets. There are various implications for both research and policy. In terms of research, our findings suggest the relevance of three main avenues for future research. First, when modeling insures behavior it is necessary to consider that competition extends well beyond premium competition and entails subtle aspects of plan design. Second, enrollees inertia in plan choices makes prominent the need to better understand the drivers of plan switching behavior. Third, the presence of risk adjustment and subsidies can push firms to compete for market shares, making adverse selection a second order concern.

Finally, in terms of policy, our results are both encouraging and problematic. On the one hand, the flexibility in product design that insurers retain in Medicare Pact C and D has allowed the 5-star SEP to achieve the goal of bolstering enrollment into 5-star. More generally, such flexibility is likely to help making the market sustainable for insurers. On the other hand, however, the very presence of such flexibility implies difficulties in designing

rules capable of steering the market toward any public goal. In the context of the 5-star SEP, the reduced generosity of 5-star plans could negatively affect the well being of the weakest beneficiaries and could also represent a diminished allocative efficiency in the market.

References

Abaluck, Jason, and Jonathan Gruber. 2013. "Evolving choice inconsistencies in choice of prescription drug insurance." National Bureau of Economic Research NBER Working Papers.

Aizawa, Naoki, and You Suk Kim. 2013. "Advertising Competition and Risk Selection in Health Insurance Markets: Evidence from Medicare Advantage." *Unpublished manuscript, University of Pennsylvania, Philadelphia.*

Akerlof, George. 1970. "The Market for Lemons: Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 488Ð500.

Azevedo, Eduardo, and Daniel Gottlieb. 2015. "Perfect Competition in Markets with Adverse Selection." mimeo.

Batata, Amber. 2004. "The effect of HMOs on fee-for-service health care expenditures: evidence from Medicare revisited." *Journal of health economics*, 23(5): 951–963.

Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston. 2014. "How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program." *American Economic Review*, 104(10): 3335–64.

Bundorf, M Kate, Natalie Chun, Gopi Shah Goda, and Daniel P Kessler. 2009. "Do markets respond to quality information? The case of fertility clinics." *Journal of health economics*, 28(3): 718–727.

Cao, Zhun, and Thomas G McGuire. 2003. "Service-level selection by HMOs in Medicare." *Journal of Health Economics*, 22(6): 915–931.

Carey, Colleen. 2014. "Government Payments and Insurer Benefit Design in Medicare Part D." mimeo.

Chetverikov, Denis, Bradley Larsen, and Christopher Palmer. 2015. "IV Quantile Regression for Group-level Treatments, with an Application to the Effects of Trade on the Distribution of Wages." *Econometrica*, forthcoming.

Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya. 2014. "Can Health Insurance Competition Work? Evidence from Medicare Advantage." National Bureau of Economic Research Working Paper 20818.

Cutler, David M., Robert S. Huckman, and Mary Beth Landrum. 2004. "The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery." *The American Economic Review*, 94(2): pp. 342–346.

Dafny, Leemore, and David Dranove. 2008. "Do report cards tell consumers anything they don't already know? The case of Medicare HMOs." *The Rand journal of economics*, 39(3): 790–821.

Darden, Michael, and Ian M McCarthy. 2014. "The Star Treatment: Estimating the Impact of Star Ratings on Medicare Advantage Enrollments." *Journal of Human Resources*, Forthcoming.

David Dranove, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is more information better? The effects of Òreport cardsÓ on health care providers." *Journal of Political Economy*, 111(3): 555–588.

Decarolis, Francesco. 2015. "Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design?" *American Economic Review*, 105(4): 1547–80.

Duggan, Mark, Patrick Healy, and Fiona Scott Morton. 2008. "Providing prescription drug coverage to the elderly: America's experiment with Medicare Part D." *The Journal of Economic Perspectives*, 69–92.

Farinha Luz, Vitor. 2015. "Dynamic Competitive Insurance." mimeo.

Geruso, Michael, and Timothy Layton. 2015. "Upcoding: Evidence from Medicare on Squishy Risk Adjustment." National Bureau of Economic Research.

Glazer, Jacob, and Thomas McGuire. 2000. "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care." *American Economic Review*, 1055Ð1071.

Handel, Ben. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton. 2014. "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program." mimeo.

Jin, Ginger Zhe, and Alan T Sorensen. 2006. "Information and consumer choice: the value of publicized health plan ratings." *Journal of Health Economics*, 25(2): 248–275.

Ketcham, Jonathan, Claudio Lucarelli, and Christopher A. Powers. 2014. "Paying Attention or Paying Too Much in Medicare Part D." *American Economic Review*, Forthcoming.

Ketcham, Jonathan D., Claudio Lucarelli, Eugenio J. Miravete, and M. Christopher Roebuck. 2012. "Sinking, Swimming, or Learning to Swim in Medicare Part D." *American Economic Review*, 102(6): 2639–73.

Layton, Timothy J., and Andrew M. Ryan. 2014. "The Effect of the Medicare Advantage Quality Based Payment Demonstration on Quality of Care in Medicare."

Lu, SF. 2012. "Information disclosure, multitasking and product quality: Evidence from nursing homes." *J. Econom. Management Strategy, v21*, 673–705.

Madeira, Tarso. 2015. "The Cost of Removing Deadlines: Evidence from Medicare Part D." mimeo.

Mahoney, Neale, and Glen Weyl. 2014. "Imperfect Competition in Selection Markets." mimeo.

McWilliams, J Michael, John Hsu, and Joseph P Newhouse. 2012. "New risk-adjustment system was associated with reduced favorable selection in Medicare Advantage." *Health Affairs*, 31(12): 2630–2640.

Miller, Keaton. 2014. "Do private Medicare firms have lower costs?" Mimeo.

Newhouse, Joseph P, and Thomas G McGuire. 2014. "How Successful Is Medicare Advantage?" *Milbank Quarterly*, 92(2): 351–394.

Newhouse, Joseph P, J Michael McWilliams, Mary Price, Jie Huang, Bruce Fireman, and John Hsu. 2013. "Do Medicare Advantage plans select enrollees in higher margin clinical categories?" *Journal of health economics*, 32(6): 1278–1288.

Nosal, Kathleen. 2012. "Estimating switching costs for medicare advantage plans." *Mimeo*.

Polyakova, Maria. 2014. "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D." mimeo.

Reid, Rachel O, Partha Deb, Benjamin L Howell, and William H Shrank. 2013. "Association Between Medicare Advantage Plan Star Ratings and EnrollmentStar Ratings for Medicare Advantage Plan." *Journal of American Medical Association*, 309(3): 267–274.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*, 90(4): 629Ð649.

Shourideh, Ali, Ben Lester, Venky Venkateswaran, and Ariel Zetlin-Jones. 2015. "Screening and Adverse Selection in a Frictional Market." mimeo.

Wang, Justin, Jason Hockenberry, Shin-Yi Chou, and Muzhe Yang. 2011. "Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania." *Journal of health economics*, 30(2): 392–407.

Werner, Rachel M, R Tamara Konetzka, Elizabeth A Stuart, Edward C Norton, Daniel Polsky, and Jeongyoung Park. 2009. "Impact of public reporting on quality of postacute care." *Health services research*, 44(4): 1169–1187.

Table 1: Domain Measures for Part C and D - Year 2012

Managed Care		**Prescription Drugs**	
Staying Healthy: screenings, tests, vaccines	12	Drug Plan Customer Service	3
Managing Chronic (long-term) Conditions	9	Member Complaints, problems getting services, and improvement in the drug plan's performance	3
Member Experience with the Health Plan	5	Member Experience with the Drug Plan	3
Member Complaints, problems getting services, and improvement in the health plan's performance	3	Patient safety and accuracy of drug pricing	6
Health Plan Customer Service	2		

Notes: The table reports the list of the domain measures used to calculate the Part C and D summary ratings in 2012. There are 5 domain measures for part C and 4 for Part D. The numbers in the table that follow the description of each domain measure indicate the number of underlying individual measures.

Table 2: Number of Counties with Treated Contracts by Insurer

Year	2009	2010	2011	2012	2013
Baystate Health, Inc.	3	3	3	3	3
Group Health Cooperative	13	13	13	13	13
Gundersen Lutheran Health System Inc.	11	11	11	16	16
Humana Inc.	0	0	11	30	30
Kaiser Foundation Health Plan, Inc.	63	63	64	64	64
Marshfield Clinic.	32	32	32	32	36
Martin's Point Health Care, Inc.	12	15	16	16	18

Notes: The table shows the number of counties in which the treated contracts where offered. Treated contracts are contracts that achieve the 5-star rating in 2012 or 2013.

Table 3: Descriptive Statistics for Part C

	2009-2011							
	Control				Treament			
	Mean	s.d.	Median	N	Mean	s.d.	Median	N
Tot. Enrollment	1338.7	4176.5	196.3	4796	7129.7	17910.4	888	409
Change Enrollment Dec.-Jan.	92.38	378.3	27	4796	386.0	863.7	117.5	409
% Change Enrollment Dec.-Jan.	0.350	0.743	0.147	4796	0.301	0.721	0.068	409
Premium Part C	41.44	38.91	36.29	4796	62.91	34.05	69.91	409
Premium Part D	27.82	17.58	29.07	4796	19.41	11.73	21.30	409
In Network MOOP	3838	1084.3	3400	1696	2781.4	604.8	2682	148
N. Top Drugs	95.20	5.973	94	4765	83.17	14.92	90	409
N. Unrestricted Drug	532.6	130.5	520	4765	641.4	102.4	641	409
Deductible Part D	44.59	94.41	0	4796	21.34	61.12	0	409
Risk Score Part C	0.965	0.229	0.908	4796	0.925	0.109	0.965	409
Risk Score Part D	0.934	0.111	0.915	4796	0.882	0.044	0.880	409
Part C OOPC Excellent	823.2	197.7	807.9	4425	800.2	110.8	801.2	409
Part C OOPC Poor	1763.5	529.9	1730.2	4425	1632.6	393.2	1643.3	409
Drug OOPC - Excellent	592.2	145.8	597.2	4425	720.7	151.0	777.3	409
Drug OOPC - Poor	1974.9	645.2	1972.9	4425	2455.9	687.5	2552	409
Health Care Quality	4.048	0.788	4	4658	4.748	0.435	5	397
Customer Service	3.809	1.128	4	3660	4.698	0.492	5	397
Drug Access	4.163	0.838	4	4654	4.952	0.214	5	397

	2012-2013							
	Control				Treament			
	Mean	s.d.	Median	N	Mean	s.d.	Median	N
Tot. Enrollment	1265.5	3753.6	236	4300	8636.0	21040.4	1320	263
Change Enrollment Dec.-Jan.	55.68	228.7	13	4300	569.6	1364.1	122.1	263
% Change Enrollment Dec.-Jan.	0.133	0.327	0.066	4300	0.101	0.110	0.0674	263
Premium Part C	35.64	35.26	31.20	4300	52.67	29.15	53.93	263
Premium Part D	25.86	18.65	25.50	4300	17.76	13.81	17.53	263
In Network MOOP	3755.6	991.6	3400	4026	3362.9	1124.3	3400.0	263
N. Top Drugs	87.05	3.757	88	4274	89.31	3.132	88	263
N. Unrestricted Drug	415.2	123.5	409.4	4274	415.6	75.30	389	263
Deductible Part D	40.54	89.19	0	4300	30.68	73.59	0	263
Risk Score Part C	0.953	0.196	0.900	4299	0.907	0.0913	0.930	263
Risk Score Part D	0.909	0.0967	0.893	4299	0.857	0.043	0.854	263
Part C OOPC Excellent	979.0	192.5	998.2	4033	989.8	121.2	1009.2	263
Part C OOPC Poor	2225.2	412.7	2286.9	4033	2172.4	372.3	2121.5	263
Drug OOPC - Excellent	624.8	130.9	618.0	4033	629.7	207.5	524.8	263
Drug OOPC - Poor	2399.0	546.6	2367.9	4033	2312.6	989.2	2163.6	263
Health Care Quality	4.236	0.622	4	4267	4.817	0.387	5	263
Customer Service	3.926	1.033	4	4219	4.319	1.225	5	263
Drug Access	3.908	1.015	4	4272	4.669	0.929	5	263

Notes: The table reports the mean, standard deviation, median and number of observations. The unit of observation is Contract/County/Year. The top panel include observation from 2009 to 2011. The bottom panel include observations from 2012 to 2013. The "Treatment" sample includes observation from Contract with 5 Star Rating in either 2012 or 2013. The "Control" sample include contracts with either 4 Star in either 2012 or 2013 in counties without 5 Star Contract. "Tot. Enrollment" is the contract enrollment measures as January. "Change Enrollment Dec.-Jan" is the change in enrollment from January to December. "% Change Enrollment Dec.-Jan" is the percentage change in enrollment from January to December. "Premium Part C" is the monthly Premium for Part C. "Premium Part D" is the monthly Premium for Part D. "In Network MOOP" is the maximum outside of pocket expenditure for in network service, exclude Part D drugs. "Deductible Part D" is the maximum annual amount of initial out of pocket expenses for Part D drugs. "N. Top Drugs" is the number of top drugs (out of 117 most frequently purchased) included in the plan formulary. "N. Unrestricted Drug" is the number of drugs without restriction on utilization included in the plan formulary. "Risk Score Part C" is the average risk score measure for Part C coverage. "Risk Score Part D" is the average risk score measure for Part D. "Part C OOPC Excellent (Poor)" is the average yearly out-of-pocket for individuals with Excellent (Poor) heath status for Part C coverage. "Drug OOPC Excellent (Poor)" is the average yearly out-of-pocket for individuals with Excellent (Poor) heath status for Part coveerage. "Health Care Quality" is a star rating (1-5), over member's evaluation of health care quality (CAHPS Survey). "Customer Service" is a star rating (1-5), over ability of the health plan to provide information or help when members need it (CAHPS Survey). "Drug Access" is a star rating (1-5) over the ease of getting prescriptions filled when using the plan (CAHPS Survey). "Tot. Enrollment", "Change Enrollment Dec.-Jan", "% Change Enrollment Dec.-Jan', "Health Care Quality", "Customer Service" and "Drug Access" are measured at contract level. "Premium Part C", "Premium Part D", "In Network MOOP", "Deductible Part D", "N. Top Drugs", "N. Unrestricted Drug", "Part C OOPC Excellent (Poor)", " Drug OOPC Excellent (Poor)", "Risk Score Part C" and "Risk Score Part D" are measured at plan level and aggregated at contract level as weighted average, with enrollment as weights. Plan with less than 10 enrollees are imputed 5 enrollees.

Table 4: MA Contracts - Within Year Enrollment Change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Baseline Sample							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
5 Star	224.327***	235.741***	86.860**	86.131**	0.074*	0.089**	0.165**	0.155**
	(50.125)	(48.533)	(39.527)	(37.405)	(0.044)	(0.042)	(0.075)	(0.070)
Observations	9,768	9,768	9,768	9,768	9,768	9,768	9,768	9,768
R-squared	0.553	0.620	0.564	0.630	0.196	0.281	0.229	0.313
	Panel B Matched Sample							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
5 Star	145.972***	153.032***	63.519**	60.888**	0.089*	0.099**	0.219***	0.202***
	(25.732)	(25.236)	(25.683)	(24.662)	(0.046)	(0.046)	(0.079)	(0.075)
Observations	7,616	7,616	7,616	7,616	7,616	7,616	7,616	7,616
R-squared	0.461	0.548	0.475	0.562	0.185	0.272	0.220	0.308
	Panel C: 4-4.5 Both Years Sample							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
Star 5	227.579***	234.389***	82.132**	79.196**	0.046	0.058	0.119	0.107
	(50.612)	(48.910)	(39.792)	(37.513)	(0.045)	(0.043)	(0.078)	(0.073)
Observations	5,674	5,674	5,674	5,674	5,674	5,674	5,674	5,674
R-squared	0.704	0.731	0.719	0.745	0.231	0.297	0.285	0.352
	Panel D: Placebo							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
5 Star	108.113***	116.613***	15.924	15.130	-0.038	0.015	0.155	0.102
	(33.168)	(30.406)	(46.024)	(42.162)	(0.065)	(0.059)	(0.110)	(0.099)
Observations	5,205	5,205	5,205	5,205	5,205	5,205	5,205	5,205
R-squared	0.469	0.618	0.478	0.630	0.277	0.428	0.311	0.464
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Contract FE	NO	YES	NO	YES	NO	YES	NO	YES
Time Trend	NO	NO	YES	YES	NO	NO	YES	YES

Notes: The table reports the DID estimates of the effect of the 5-star SEP. The outcome variable is the difference in the contract enrollment between December and January (of the same year) calculated either in levels (first four columns) or in percentage (latter four columns). The four model specifications considered for each dependent variable differ in the set of controls used, as reported in the block at the very end of the table. Panel A reports the estimates for the baseline sample: treatment group contracts with 5-star in 2012 or 2013; control group contracts with more 4 or 4.5 star in 2012 or 2013 in counties without 5-star contracts. Panel B reports the estimates for a sample matched using a propensity score. The probability that a county has a 5-star contract is estimated over a range of socio-economical, demographic and health indicators of the counties. Only the county on common support of the propensity score between the treatment and the control groups are included. Panel C, treatment group contracts with 5-star in 2012 or 2013; control group contracts with 4 or 4.5 star in both 2012 and 2013 in counties without 5-star contracts. Panel D, placebo test, over the year 2009-2011 with a simulated policy introduced in 2011 (same sample as Panel A). Standard errors in parentheses clustered at county level *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: MA Contracts - Across Years Enrollment Change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Baseline Sample							
	Jan.-Dec. Enrollment Change				Jan.-Dec. Enrollment % Change			
5 Star	-2.072	0.272	21.254	22.616	0.044	0.039	0.186***	0.204***
	(15.362)	(15.002)	(26.370)	(24.777)	(0.037)	(0.033)	(0.054)	(0.052)
Observations	8,823	8,823	8,823	8,823	8,823	8,823	8,823	8,823
R-squared	0.079	0.121	0.088	0.130	0.143	0.219	0.148	0.225
	Panel B: Matched Sample							
	Jan.-Dec. Enrollment Change				Jan.-Dec. Enrollment % Change			
Star 5	8.495	10.458	8.776	8.914	0.065	0.057	0.243***	0.261***
	(13.164)	(12.988)	(19.117)	(18.275)	(0.040)	(0.036)	(0.056)	(0.055)
Observations	7,094	7,094	7,094	7,094	7,094	7,094	7,094	7,094
R-squared	0.138	0.190	0.167	0.220	0.118	0.204	0.124	0.212
	Panel C: 4-4.5 Both Years Sample							
	Jan.-Dec. Enrollment Change				Jan.-Dec. Enrollment % Change			
5 Star	2.032	1.267	5.136	3.585	-0.027	-0.041	0.019	0.017
	(17.416)	(16.864)	(39.290)	(37.351)	(0.031)	(0.025)	(0.035)	(0.032)
Observations	5,407	5,407	5,407	5,407	5,407	5,407	5,407	5,407
R-squared	0.093	0.127	0.100	0.134	0.191	0.250	0.206	0.263
	Panel D: Placebo							
	Jan.-Dec. Enrollment Change				Jan.-Dec. Enrollment % Change			
Star 5	-16.327	-13.821	-95.281**	-90.711***	-0.176***	-0.094**	0.009	0.027
	(19.684)	(18.303)	(37.429)	(32.715)	(0.054)	(0.048)	(0.106)	(0.087)
Observations	4,636	4,636	4,636	4,636	4,636	4,636	4,636	4,636
R-squared	0.090	0.172	0.092	0.174	0.197	0.391	0.204	0.395
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Contract FE	NO	YES	NO	YES	NO	YES	NO	YES
Time Trend	NO	NO	YES	YES	NO	NO	YES	YES

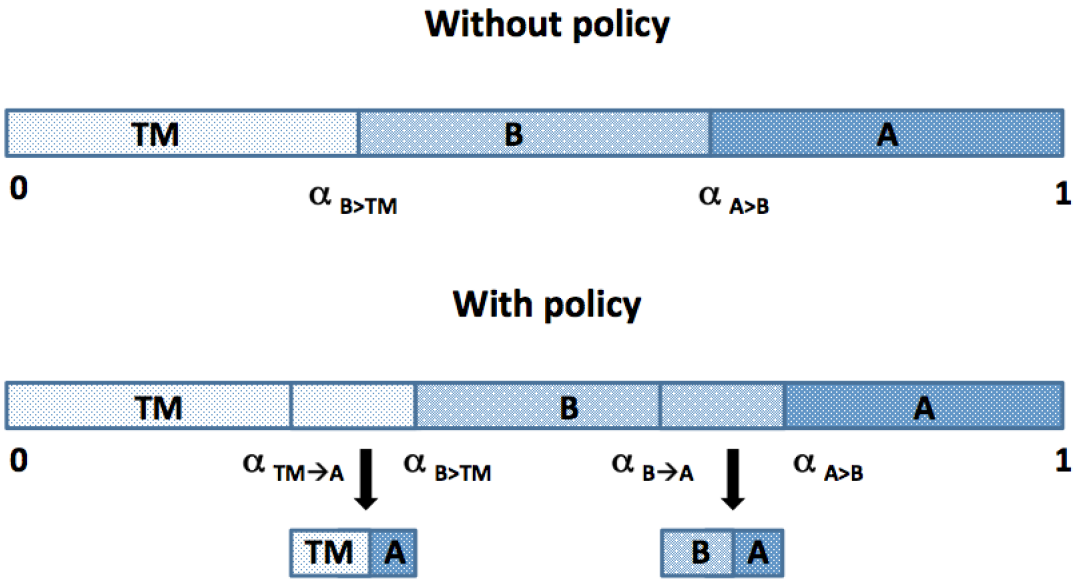
Notes: The table reports the DID estimates of the effect of the 5-star SEP. The outcome variable is the difference in the contract enrollment between January and December (of consecutive years) calculated either in levels (first four columns) or in percentage (latter four columns). Panel A reports the estimates for the baseline sample: treatment group contracts with 5-star in 2012 or 2013; control group contracts with more 4 or 4.5 star in 2012 or 2013 in counties without 5-star contracts. Panel B reports the estimates for a sample matched using a propensity score. The probability that a county has a 5-star contract is estimated over a range of socio-economical, demographic and health indicators of the counties. Only the county on common support of the propensity score between the treatment and the control groups are included. Panel C, treatment group contracts with 5-star in 2012 or 2013; control group contracts with 4 or 4.5 star in both 2012 and 2013 in counties without 5-star contracts. Panel D, placebo test, over the year 2009-2011 with a simulated policy introduced in 2011 (same sample as Panel A). Standard errors in parentheses clustered at county level *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: MA Contracts - Risk Score Part C and D

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Baseline Sample				
	Risk Score Part C				Risk Score Part D			
5 Star	-0.024***	-0.029***	-0.016***	-0.014***	-0.007***	-0.010***	-0.010***	-0.008***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)
Observations	9,767	9,767	9,767	9,767	9,767	9,767	9,767	9,767
R-squared	0.349	0.949	0.354	0.953	0.349	0.930	0.354	0.935
				Panel B: 2012 Effect				
	Risk Score Part C				Risk Score Part D			
5 Star	-0.014***	-0.023***	-0.025***	-0.022***	-0.004**	-0.009***	-0.015***	-0.013***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.002)	(0.001)	(0.003)	(0.002)
Observations	7,372	7,372	7,372	7,372	7,372	7,372	7,372	7,372
R-squared	0.355	0.954	0.361	0.959	0.363	0.937	0.368	0.942
				Panel C: 2013 Effect				
	Risk Score Part C				Risk Score Part D			
5 Star	-0.042***	-0.042***	-0.032***	-0.013	-0.013***	-0.013***	-0.020***	-0.012**
	(0.007)	(0.006)	(0.011)	(0.010)	(0.003)	(0.003)	(0.005)	(0.005)
Observations	7,600	7,600	7,600	7,600	7,600	7,600	7,600	7,600
R-squared	0.356	0.951	0.361	0.955	0.366	0.928	0.371	0.934
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Contract FE	NO	YES	NO	YES	NO	YES	NO	YES
Time Trend	NO	NO	YES	YES	NO	NO	YES	YES

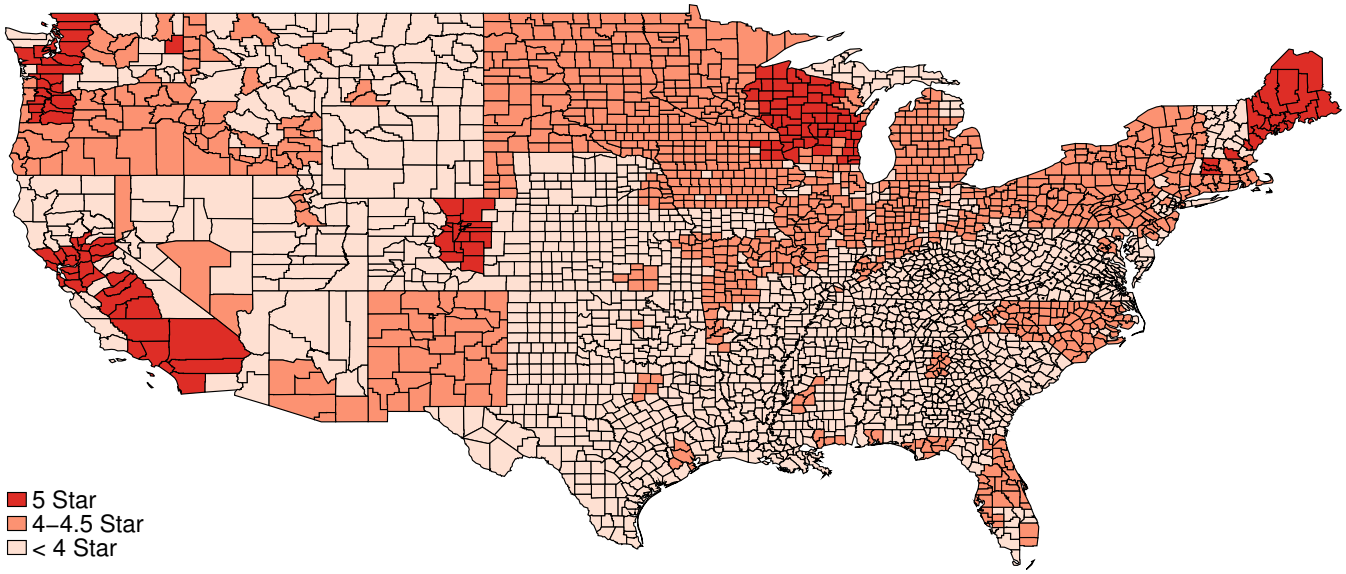
Notes: The table reports the DID estimates of the effect of the 5-star SEP. The outcome variable is the risk score for Part C (first four columns) and Part D (latter four columns). The four model specifications considered for each dependent variable differ in the set of controls used, as reported in the block at the very end of the table. Panel A reports the estimates for the baseline sample: treatment group contracts with 5-star in 2012 or 2013; control group contracts with more 4 or 4.5 star in 2012 or 2013 in counties without 5-star contracts. Panel B reports estimates from a sample without observation from 2013. Panel C reports estimates from a sample without observation from 2012. Standard errors in parentheses clustered at county level *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 1: Enrollment Shares with and without Policy



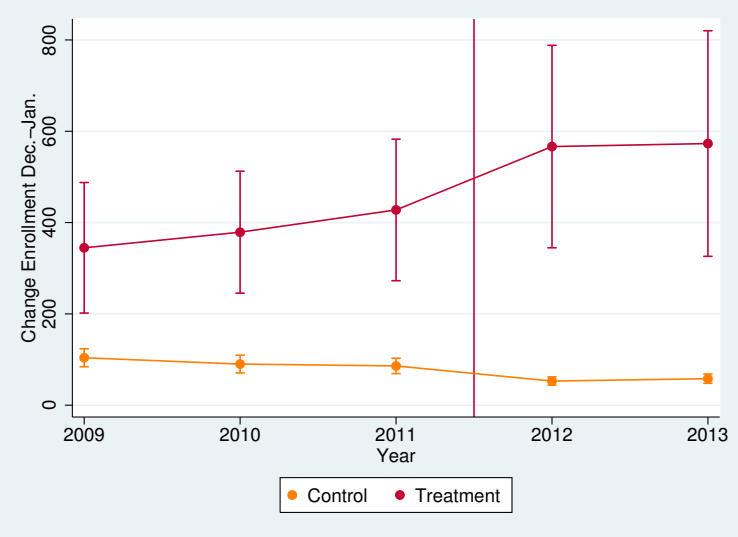
Notes: The two figures show the allocation of consumers to A, B and the outside option TM. There is a unit mass of consumers who are sorted in the figure by their value of α, from the lowest (zero) to the highest (one).

Figure 2: Maps of 5-Star Counties



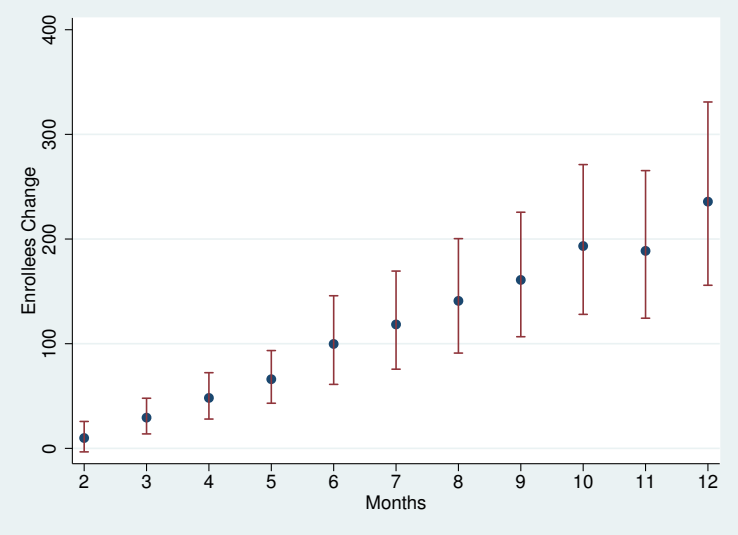
Notes: The heat map reports with the darkest color the set of counties where at least one 5-star plan was offered in 2012 or 2013. The lightest color counties are those where in the same period no plan got a score of 4 or higher. The remaining counties have at least one plan with a score of al least 4, but no plan with a score of 5.

Figure 3: MA Contracts - Within Year Enrollment Change



Notes: Evolution of the Dec. minus Jan. contract enrollment variable for both treatment and control contracts.

Figure 4: MA Contracts - Monthly Enrollment Change Relative to January



Notes: Estimate of the effect of the 5-star SEP on within year enrollment change, calculated at all months. The last value on the horizontal axis (12) represents the Dec. minus Jan. enrollment, the next value (11) represents the Nov. minus Jan. enrollment, and so on until (2) that represent the Feb. minus Jan. enrollment. The value for the Dec. minus Jan. enrollment is the same reported in the second column of Panel A in Table 4. All other estimates are obtained using the same specification.

Figure 5: MA Contracts - Evolution of the Risk Score Measures

(a) Risk Score Part C (b) Risk Score Part D

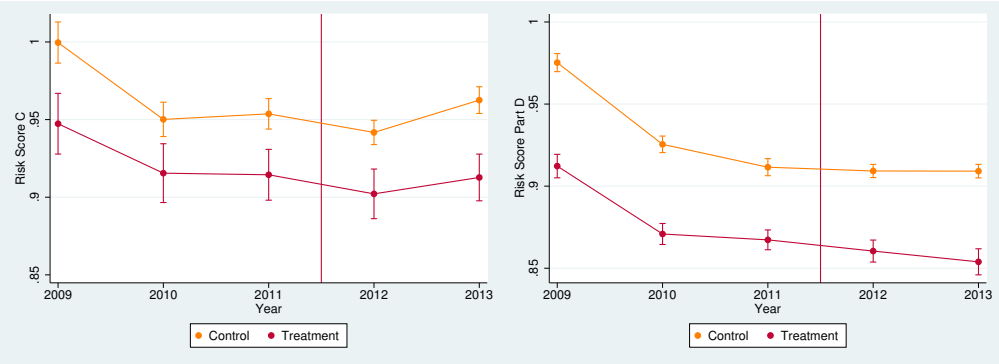


Figure 6: Average Risk Score of Contracts with Net Inflow or Outflow of Enrollees

(a) Risk Score Part C (b) Risk Score Part D

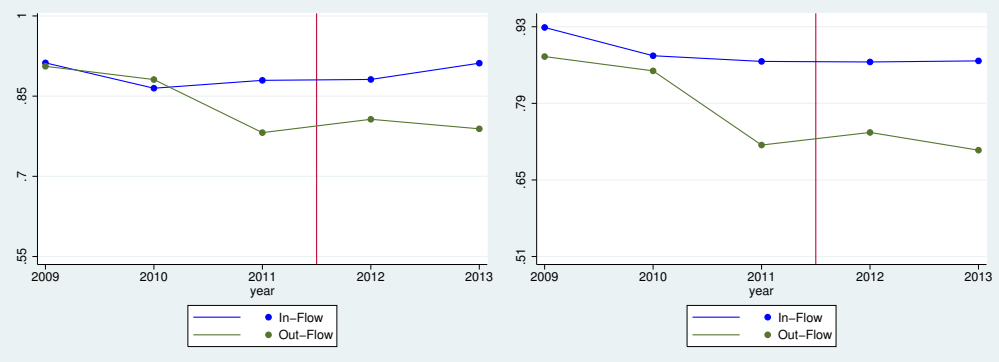


Figure 7: Sources of the Enrollees Inflow/Outflow by Program

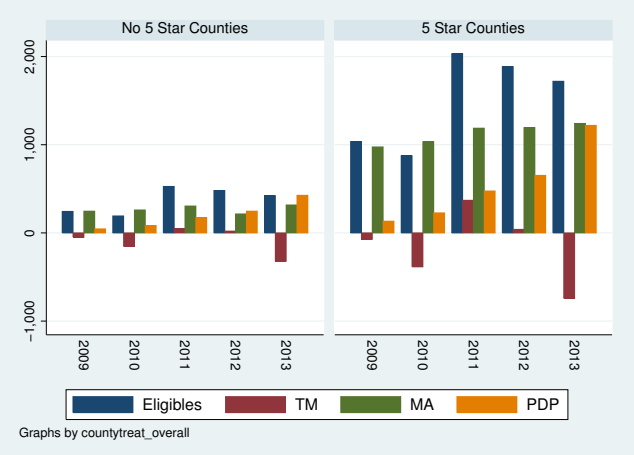
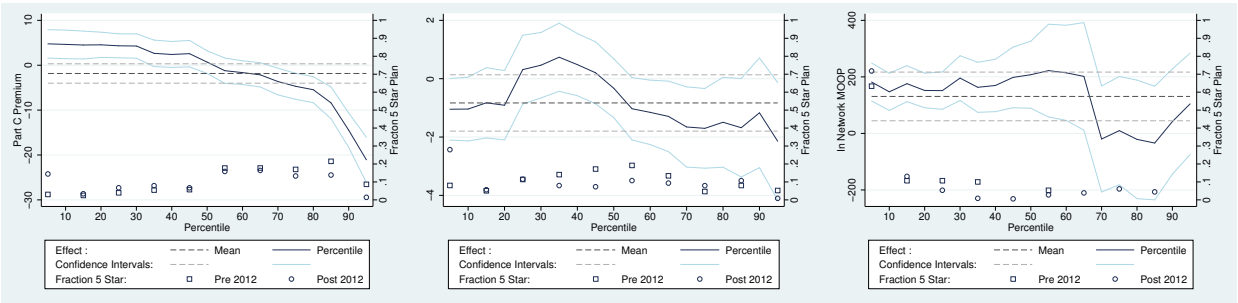
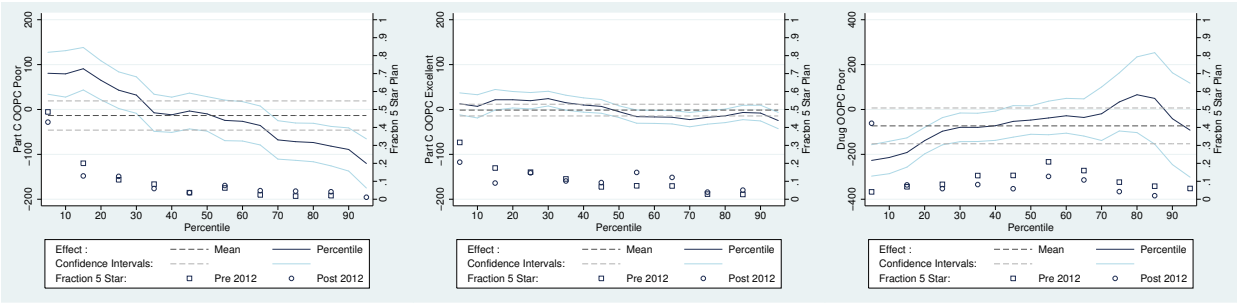


Figure 8: Quantile Regression Estimates for Plan Characteristics

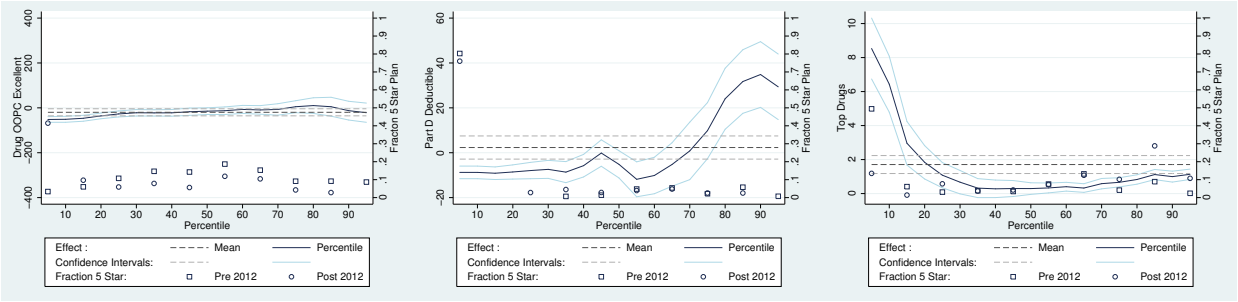
(a) Part C Premium (b) Part D Premium (c) Maximum OOP Part C



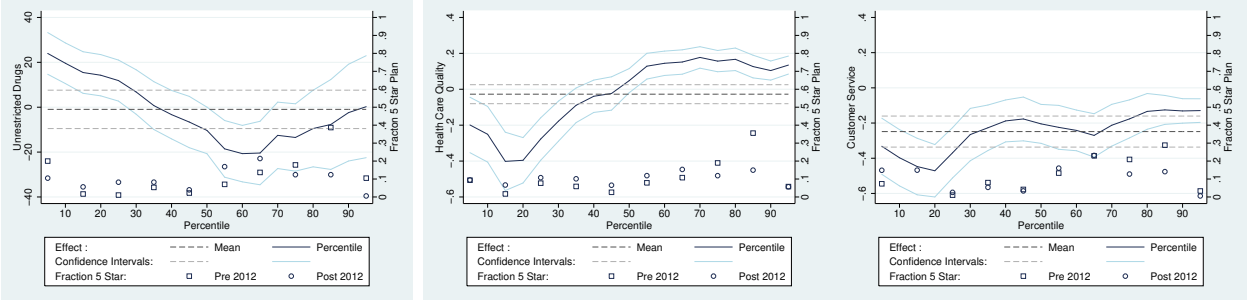
(d) Part C OOPC - Poor Health (e) Part C OOPC - Excellent Health (f) Drug OOPC - Poor Health



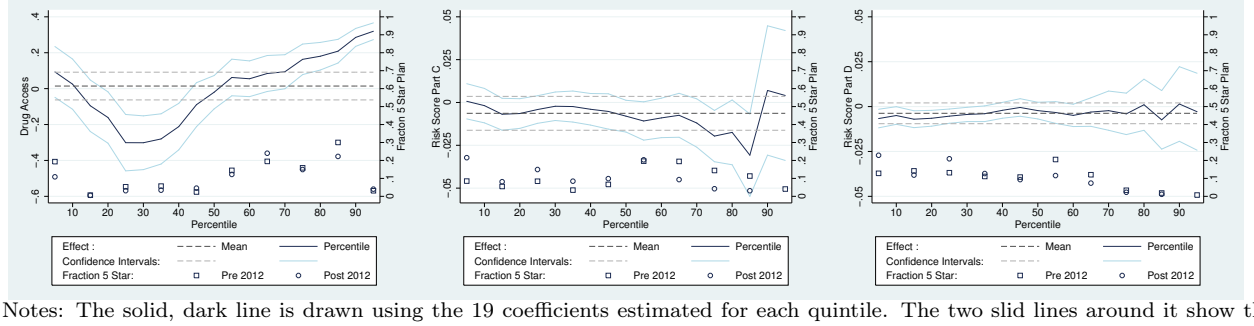
(g) Drug OOPC - Excellent Health (h) Part D Deductible (i) N. Top Drugs



(j) N. Unrestricted Drugs (k) Health Care Quality (l) Customer Service



(m) Drug Access (n) Risk Score Part C (o) Risk Score Part D

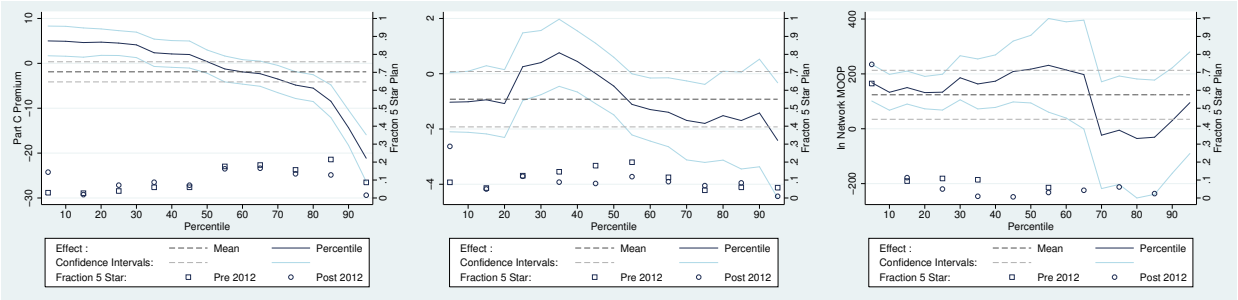


Notes: The solid, dark line is drawn using the 19 coefficients estimated for each quintile. The two slid lines around it show the 95% confidence interval. The dark, horizontal, dashed line shows the mean effect, the lighter lines denotes the 95% confidence interval. Squares measure the share of 5-star plans in the pre-policy period, while circles measure them in the post-policy period.

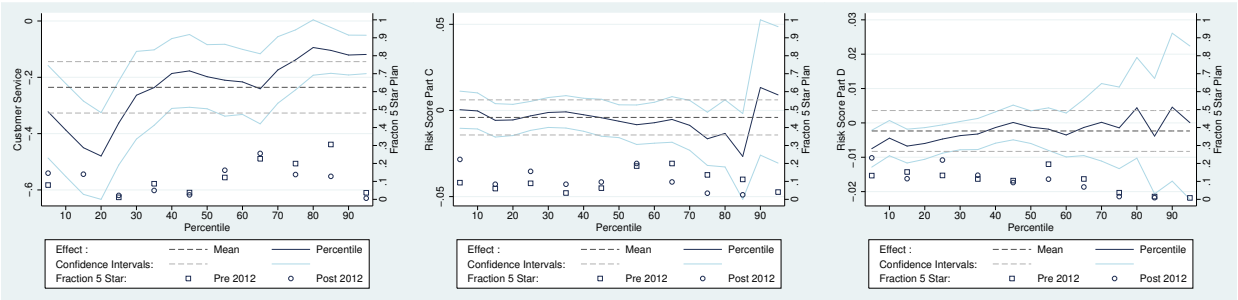
Figure 9: Quantile Regression Estimates - Monopoly and Duopoly Counties

Part I: 5-Star Monopoly Counties

(a) Part C Premium (b) Part D Premium (c) Maximum OOP Part C

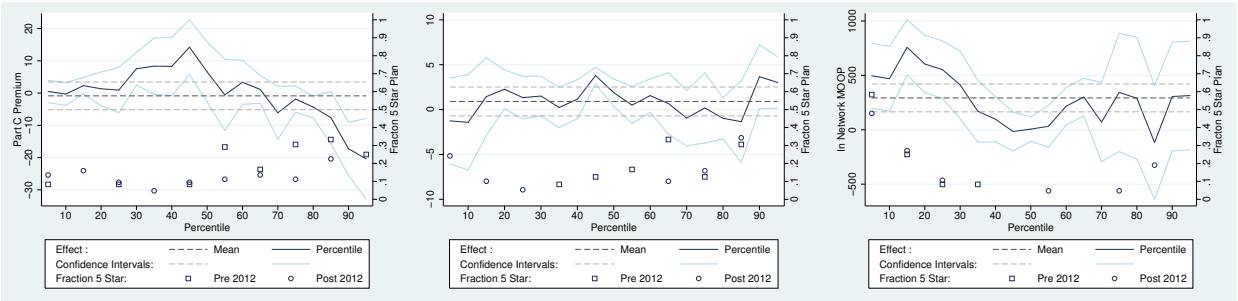


(d) Customer Service (e) Risk Score Part C (f) Risk Score Part D

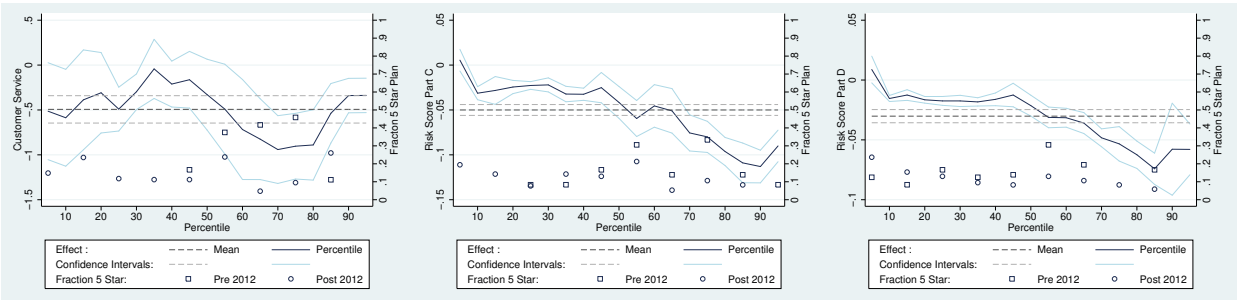


Part II: 5-Star Duopoly Counties

(g) Part C Premium (h) Part D Premium (i) Maximum OOP Part C



(j) Customer Service (k) Risk Score Part C (l) Risk Score Part D



Notes: See note to previous table. Top panel includes as treated counties only those with 1 insurer offering all 5-star plans. Bottom panel includes as treated counties only those with 2 insurers offering all 5-star plans.

For Publication on the Author' Web Page

Insurers Response to Selection Risk:
Evidence from Medicare Enrollment Reforms

Web Appendix

A. Data and Institutions

The dataset was assembled from data made publicly available by CMS (Center for Medicare and Medicaid Services). In particular, data on monthly enrollment for the years 2009-2013 at plan level was downloaded from:

`http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/index.html`.

The *Crosswalk Files* available from the same web site were used to link plans through the years. Premiums and plan financial characteristics are from the *Premium Files*:

`http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/index.html`.

Plans formulary and pharmacy network are from the FRF (*Formulary Reference Files*):

`https://www.cms.gov/PrescriptionDrugCovContra/03_RxContracting_FormularyGuidance.asp`

Part C and D performance data determining the star ratings were obtained from:

`https://www.cms.gov/medicare/prescription-drug-coverage/prescriptiondrugcovgenin/performancedata.html`

Demographic characteristics for the geographic areas are the only ancillary data source and were obtained from:

`http://ahrf.hrsa.gov/download.htm`.

The calculation of the star rating described in the main text is illustrated below in greater details for the case of the Part D rating for year 2012. A weighted average of the scores earned on each of the individual measures determines the final score.

Table A.1: Rating Calculation for Part D - Year 2012

Individual Measures			Domain Measures	Summary Measures
Definition	Type of Data	Weights		
D01 Call Center - Hold Time	Call Center Monitored by CMS	1.5	Domain 1 Drug Plan Customer Service	Summary Rating
D02 Call Center - Foreign Language Interpreter	Call Center Monitored by CMS	1.5		
D03 Appeals Auto-Forward	Independent Review Entity	1.5		
D04 Appeals Upheld	Independent Review Entity	1.5		
D05 Enrollment Timeliness	Medicare Advantage Prescription Drug System (CMS)	1		
D06 - Complaints about the Drug Plan	Complaint Tracking System (CMS)	1.5	Domain 2 Member Complaints, Problems Getting Services, and Choosing to Leave the Plan	
D07 - Beneficiary Access and Performance Problems	CMS Administrative Data	1.5		
D08 - Members Choosing to Leave the Plan	Medicare Beneficiary Database Suite of Systems (CMS)	1.5		
D09 - Getting Information From Drug Plan	CAHPS Survey	1.5	Domain 3 Experience with Drug Plan	
D10 - Rating of Drug Plan	CAHPS Survey	1.5		
D11 - Getting Needed Prescription Drugs	CAHPS Survey	1.5		
D12 - MPF Composite	Prescription Drug Event, Medicare Plan Finder, Health Management Plan System and Medispan	1	Domain 4 Drug Pricing and Patient Safety	
D13 - High Risk Medication	Prescription Drug Event	3		
D14 - Diabetes Treatment	Prescription Drug Event	3		
D15 - Part D Medication Adherence for Oral Diabetes Medications	Prescription Drug Event	3		
D16 - Part D Medication Adherence for Hypertension (ACEI or ARB)	Prescription Drug Event	3		
D17 - Part D Medication Adherence for Cholesterol (Statins)	Prescription Drug Event	3		

Notes: The table reports the details of how the 2012 summary rating is calculated for Part D. There are three sets of measures: individual measures (17 measures, reported in the first column), domain measures (4 measures, reported in the fourth column) and the final summary rating (fifth column). The third column describes the weights associated to each of 17 the individual measures in the calculation of the corresponding domain measures. The 4 domain measures are equally weighted in the calculation of the summary rating.

B. Baseline Framework: Details

This section reports the details of the baseline framework presented in the text. Given the assumptions on consumers utility stated in the text, in the pre-policy period, a consumer will choose B over the TM when $\mu - p_B + \alpha_i > -\gamma$, inducing a cutoff point $\alpha_{B>TM} = p_B - \mu - \gamma$. A consumer will choose A over B when $\mu - p_A + \gamma(\alpha_i + b) > \mu - p_B$, inducing a cutoff point $\alpha_{A>B} = -b + \frac{p_A - p_B}{\gamma}$. As regards insurers, we assume that for each firm the cost of enrolling a consumer is zero if he turns out to be healthy and c if sick. Firms set premiums to solve:

$$\max_{p_B} \pi_B = (\alpha_{A>B} - \alpha_{B>TM})(p_B - \gamma c) \quad \text{and} \quad \max_{p_A} \pi_A = (1 - \alpha_{A>B})(p_A - \gamma c)$$

Denoting the equilibrium prices as (p_B^*, p_A^*), the equilibrium cutoffs are $\alpha_{B>TM} = p_B^* - \mu - \gamma$ and $\alpha_{A>B} = -b + \frac{p_A^* - p_B^*}{\gamma}$. The top panel of Figure 1 describes the resulting market shares.

In the post-policy period, given that consumers are initially unaware of the policy change, the initial choice cutoffs $\alpha_{A>B}$ and $\alpha_{B>TM}$ are the same functions as above. However, once the policy is revealed consumers from the outside option will switch to A if $-\alpha_i - h_i < \mu - p_A + h_i(\alpha_i + b) - \phi_{TM \to A}$. Switching to A is a dominated choice for healthy consumers and, hence, the subset of TM enrollees switching to A is composed by those that turn out to be sick and who have $-\alpha_i - 1 < \mu - p_A + \alpha_i + b - \phi_{O \to A}$, inducing a cutoff $\alpha_{TM \to A} = \frac{p_A - b - \mu + \phi_{O \to A} - 1}{2}$.

Similarly, consumers from B find switching to A suboptimal when healthy, but sick consumers switch when their α is such that: $\mu - p_B < \mu - p_A - \phi_{B \to A} + \alpha + b$, inducing a cutoff point $\alpha_{B \to A} = p_A - p_B - b + \phi_{B \to A}$. Given this demand, firms set premiums to solve:

$$\max_{p_B} (\alpha_{B \to A} - \alpha_{B>TM})(p_B - \gamma c) + (1 - \gamma)(\alpha_{A>B} - \alpha_{B \to A})(p_B) \text{ and}$$
$$\max_{p_A} (1 - \alpha_{A>B})(p_A - \gamma c) + \gamma(\alpha_{A>B} - \alpha_{B \to A})(p_A - c) + \gamma(\alpha_{B>TM} - \alpha_{TM \to A})(p_A - c).$$

The ensuing equilibrium market shares can be found by inserting the resulting equilibrium prices into the four cutoff functions: $\alpha_{TM \to A}$, $\alpha_{B>TM}$, $\alpha_{B \to A}$ and $\alpha_{A>B}$.

C. Additional Results and Robustness Checks

- The first set of additional results reported concerns the probit estimates used for the construction of the matched DID estimates in the demand analysis. Table A.2 reports the estimates for four model specifications (i.e., columns 1-2, 3-4, 5 and 6) where we gradually increase the set of controls. All controls are county-level demographic characteristics collected from the AHRF files of the Health Resources and Services Administration. The estimates reported in column 2 and 4 differ from those in columns 1 and 3, respectively, for the sample of counties included: due to missing data for some characteristics, for columns 2 and 4 we use a smaller sample than that used for columns 1 and 3. The sample used for columns 2 and 4 is the same used for columns 5 and 6. The matched DID reported in the main text are based on the estimates in column 6 of Table A.2. Although this table clearly shows that estimates are fairly stable across models, to further assess the robustness of the DID in the main text we report in Table A.3 matched DID estimates based on the outcomes of the three other probit models (i.e. model 1, 3 and 5). Overall, the results are broadly in line with what reported in the main text.

- Table A.4 complements Table 6 in the main text by reporting for the Part C and D risk score measures the three types of robustness checks performed for the enrollment measures (see panels B, C and D of Table 4 and 5). The negative and significant effect reported in the main text is robust to these robustness checks. Regarding the placebo test, we observe diverging effect on the risk score for Part C and D, that could be explain by other underlying trend in the market.

- Table A.5 reports the estimates of the demand analysis for Part D: within-year and across years demand for PDP. As mentioned in the text, for PDP a major difficulty to apply the DID approach is that only 2 regions out of 34 are treated. As result, we face a limit our capacity to conduct inference given the small number of treated units.[34] Even

[34]See Timothy G. Conley and Christopher R. Taber, "Inference with 'Difference in Differences' with a Small Number of Policy Changes," *Review of Economics and Statistics*, 2011, 93 (1): 113-25.

under this caveat it is interesting to note that the Part D findings resemble the Pact C ones: within-year enrollment increases (columns 1-4) and by a magnitude similar to what found for Part C (roughly by 10 percent of the enrollment base). The across-years enrollment of 5 star contracts declines, but in a way that is not statistically significant.

- Finally, we present two sets of robustness checks for our supply side analysis. The first one entails using a control group that matches the treatment group on observable characteristics. Thus, we repeat for the quantile DID what done for the demand analysis when implementing the matched DID. The results, reported in Figure A.1, show patterns nearly identical to what reported in the main text. The second set of robustness checks involves the way plan features are aggregated at contract level. Indeed, while we perform our analysis at contract level, certain features, like the Part D deductible are plan-specific and will differ among plans within the same contract. In the main text, we presented results where the aggregation method used an enrollment-weighted average of the plans. In this Figure A.2, instead, we report the results from equally weighted plans. We consider only the subset of characteristics varying at plan level. The findings are broadly in line with what reported in the main text.

Table A.2: Probit Results - Probability of County Having 5 Star Plan

	(1) 5 Star County	(2) 5 Star County	(3) 5 Star County	(4) 5 Star County	(5) 5 Star County	(6) 5 Star County
MA Enrollees	2.981***	2.334***	2.858***	2.268***	2.234***	2.255***
	(0.448)	(0.484)	(0.454)	(0.487)	(0.513)	(0.518)
Pop. Male > 65	0.000951***	0.00126***	0.000896***	0.00120***	0.00100*	0.00105*
	(0.000333)	(0.000461)	(0.000317)	(0.000456)	(0.000555)	(0.000600)
Pop. Female > 65	-0.000787***	-0.000973***	-0.000747***	-0.000921***	-0.000836**	-0.000878**
	(0.000245)	(0.000328)	(0.000236)	(0.000324)	(0.000392)	(0.000430)
Pop. White-Male > 65	-0.000890**	-0.00119**	-0.000851**	-0.00114**	-0.00111*	-0.00118*
	(0.000361)	(0.000489)	(0.000344)	(0.000484)	(0.000592)	(0.000645)
Pop. White-Female > 65	0.000573**	0.000780**	0.000542**	0.000739**	0.000653	0.000705
	(0.000255)	(0.000348)	(0.000242)	(0.000344)	(0.000413)	(0.000451)
Medicare Eligibles	8.13e-05***	6.55e-05***	8.25e-05***	6.47e-05**	0.000149***	0.000150***
	(2.38e-05)	(2.53e-05)	(2.46e-05)	(2.62e-05)	(3.80e-05)	(4.09e-05)
Unemployment			0.0519**	0.0488*	0.0305	0.0289
			(0.0254)	(0.0267)	(0.0285)	(0.0289)
Poverty Rate			-0.0321**	-0.0241	-0.0110	-0.0104
			(0.0148)	(0.0155)	(0.0159)	(0.0162)
# Medicare Cert Hosp.					0.216***	0.110
					(0.0660)	(0.256)
# Hosp. Med Patients					-2.32e-05***	-2.63e-05***
					(4.15e-06)	(4.87e-06)
# Outpatients Visits					1.50e-07	1.03e-07
					(2.17e-07)	(2.41e-07)
Hosp. Util. Rate 0-39						-0.0999
						(0.270)
Hosp. Util. Rate 40-59						0.144
						(0.262)
Hosp. Util. Rate 60-79						0.296
						(0.262)
Hosp. Util. Rate >80						0.330
						(0.283)
Constant	-1.762***	-1.588***	-1.756***	-1.681***	-1.960***	-1.922***
	(0.109)	(0.120)	(0.241)	(0.268)	(0.291)	(0.295)
Observations	987	841	987	841	841	841

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.3: MA Contracts - Within Year Enrollment Change - Matched Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Model 1							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
5 Star	212.267***	222.275***	79.771**	78.749**	0.078*	0.089**	0.154**	0.140*
	(49.064)	(48.172)	(38.894)	(37.235)	(0.044)	(0.043)	(0.075)	(0.072)
Observations	8,486	8,486	8,486	8,486	8,486	8,486	8,486	8,486
R-squared	0.635	0.686	0.647	0.697	0.193	0.272	0.224	0.305
	Panel B: Model 3							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
5 Star	210.904***	221.579***	80.953**	80.095**	0.073*	0.087**	0.156**	0.144**
	(49.086)	(48.160)	(38.891)	(37.213)	(0.044)	(0.043)	(0.075)	(0.072)
Observations	8,734	8,734	8,734	8,734	8,734	8,734	8,734	8,734
R-squared	0.628	0.682	0.640	0.694	0.188	0.273	0.219	0.305
	Panel C: Model 5							
	Dec.-Jan. Enrollment Change				Dec.-Jan. Enrollment % Change			
treat_overall	154.346***	161.143***	66.955**	66.349***	0.089*	0.100**	0.222***	0.205***
	(26.869)	(26.381)	(26.612)	(25.548)	(0.046)	(0.046)	(0.079)	(0.076)
Observations	7,533	7,533	7,533	7,533	7,533	7,533	7,533	7,533
R-squared	0.440	0.523	0.453	0.536	0.183	0.271	0.219	0.307
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Contract FE	NO	YES	NO	YES	NO	YES	NO	YES
Time Trend	NO	NO	YES	YES	NO	NO	YES	YES

Notes: The table reports the DID estimates of the effect of the 5-star SEP. The outcome variable is the difference in the contract enrollment between December and January (of the same year) calculated either in levels (first four columns) or in percentage (latter four columns). The four model specifications considered for each dependent variable differ in the set of controls used, as reported in the block at the very end of the table.

Table A.4: MA Contracts - Risk Score Part C and D - Additional Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 4-4.5 Both Years								
	Risk Score Part C				Risk Score Part D			
5 Star	-0.019***	-0.018***	-0.007	-0.003	-0.005**	-0.005**	-0.006**	-0.005*
	(0.005)	(0.004)	(0.005)	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)
Observations	5,674	5,674	5,674	5,674	5,674	5,674	5,674	5,674
R-squared	0.419	0.948	0.424	0.953	0.388	0.916	0.397	0.923
Panel B: Matched Sample								
	Risk Score Part C				Risk Score Part D			
5 Star	-0.035***	-0.033***	-0.029***	-0.027***	-0.012***	-0.012***	-0.018***	-0.015***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	8,234	8,234	8,234	8,234	8,234	8,234	8,234	8,234
R-squared	0.290	0.939	0.295	0.942	0.277	0.912	0.282	0.917
Panel C: Placebo								
	Risk Score Part C				Risk Score Part D			
Star 5	-0.003	-0.003	-0.018***	-0.006	0.008***	0.006***	0.004	0.008**
	(0.004)	(0.004)	(0.007)	(0.005)	(0.002)	(0.002)	(0.004)	(0.003)
Observations	6,114	6,114	6,114	6,114	6,114	6,114	6,114	6,114
R-squared	0.381	0.949	0.389	0.955	0.410	0.925	0.415	0.930
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Contract FE	NO	YES	NO	YES	NO	YES	NO	YES
Time Trend	NO	NO	YES	YES	NO	NO	YES	YES

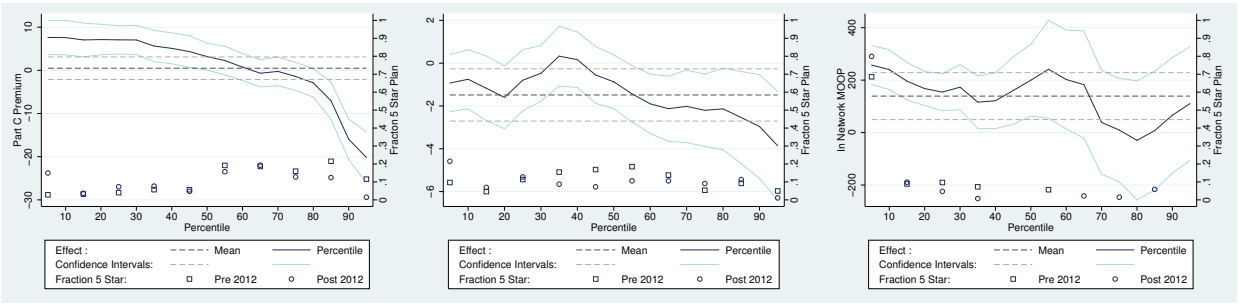
Notes: The table reports the DID estimates of the effect of the 5-star SEP. The outcome variable is the risk score for Part C (first four columns) and Part D (latter four columns). The four model specifications considered for each dependent variable differ in the set of controls used, as reported in the block at the very end of the table.

Table A.5: PDP Plans - Within and Across Year Enrollment Changes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Within-Year				Across-Year			
	Change		% Change		Change		% Change	
5 Star	2,419**	2,416**	0.128**	0.130**	-17,835	-17,582	-0.0786	-0.0803
	(919.3)	(900.4)	(0.0567)	(0.0550)	(12,233)	(11,985)	(0.0896)	(0.0873)
Observations	499	499	499	499	497	497	372	372
R-squared	0.018	0.026	0.204	0.251	0.186	0.202	0.074	0.097
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES	YES	YES
Contract FE	NO	YES	NO	YES	NO	YES	NO	YES

Figure A.1: Quantile Regression Estimates for Plan Characteristics - Matched Samples

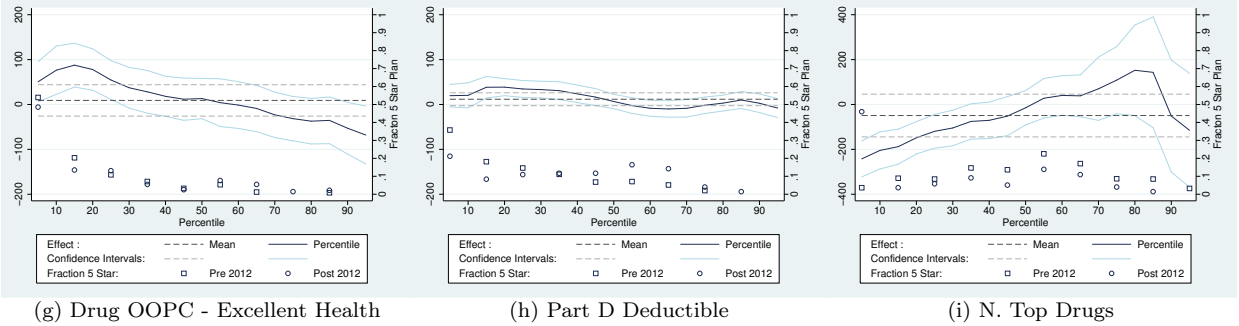
(a) Part C Premium



(b) Part D Premium

(c) Maximum OOP Part C

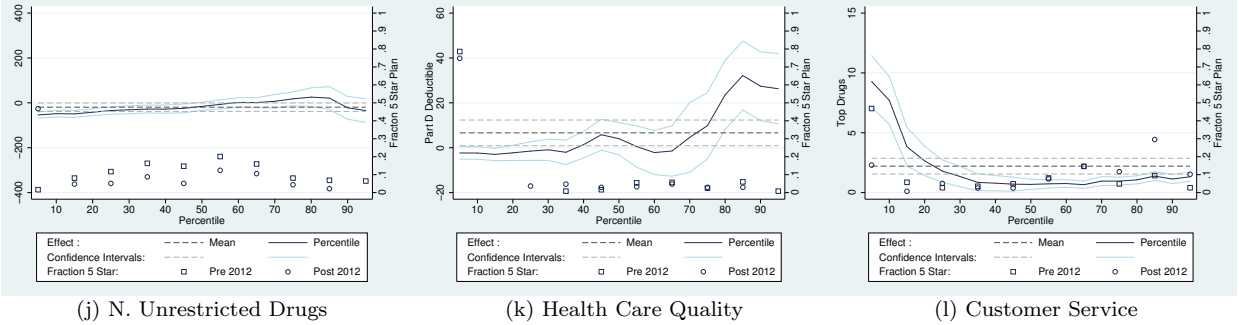
(d) Part C OOPC - Poor Health



(e) Part C OOPC - Excellent Health

(f) Drug OOPC - Poor Health

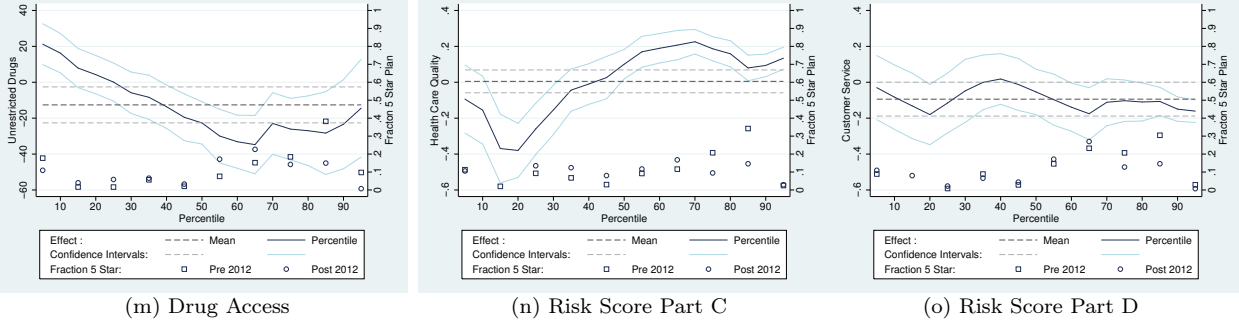
(g) Drug OOPC - Excellent Health



(h) Part D Deductible

(i) N. Top Drugs

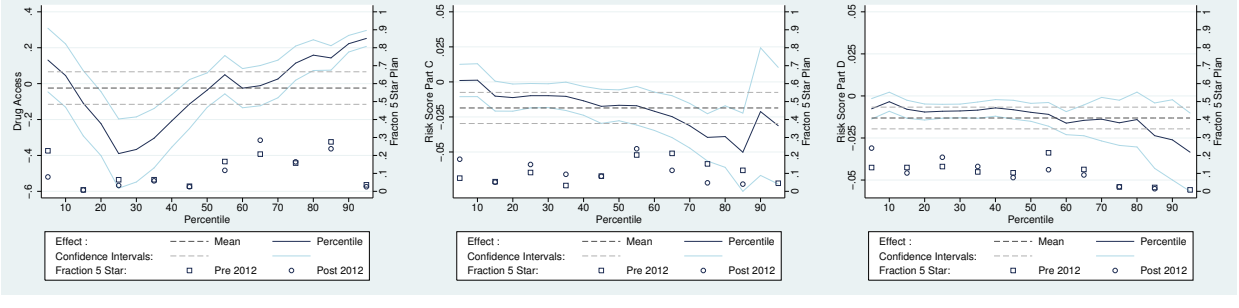
(j) N. Unrestricted Drugs



(k) Health Care Quality

(l) Customer Service

(m) Drug Access

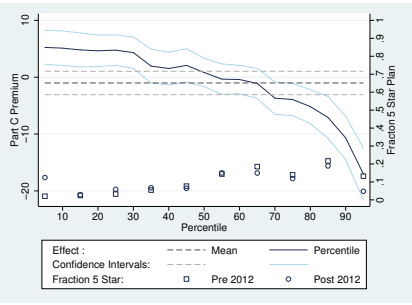


(n) Risk Score Part C

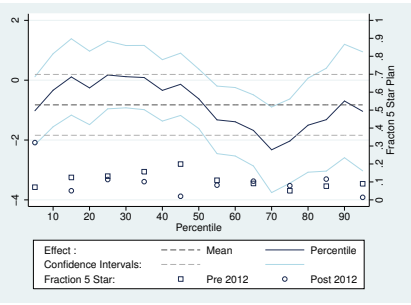
(o) Risk Score Part D

Figure A.2: Quantile Regression Estimates for Plan Characteristics - Mean Characteristics

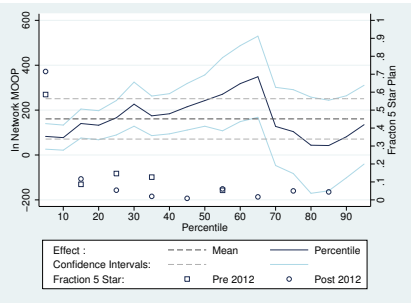
(a) Part C Premium



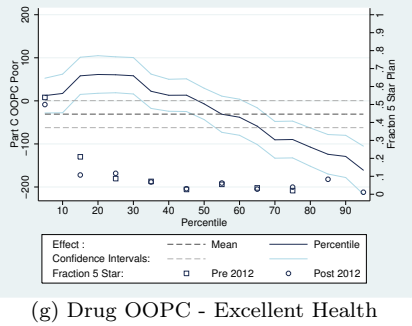
(b) Part D Premium



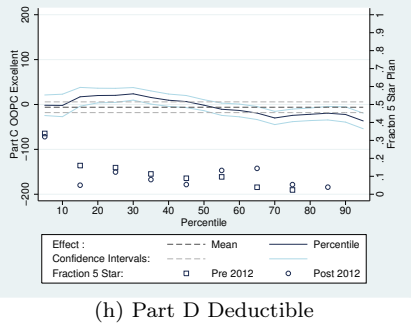
(c) Maximum OOP Part C



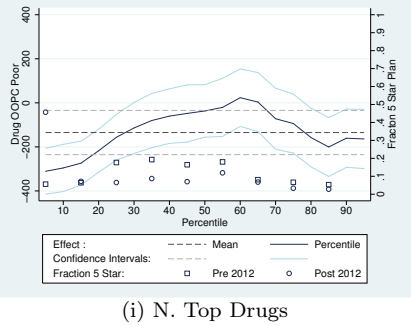
(d) Part C OOPC - Poor Health



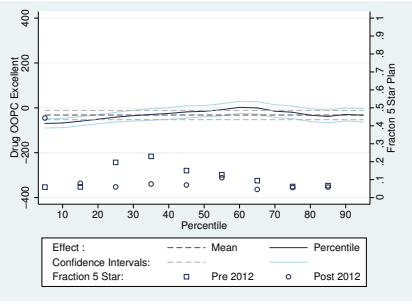
(e) Part C OOPC - Excellent Health



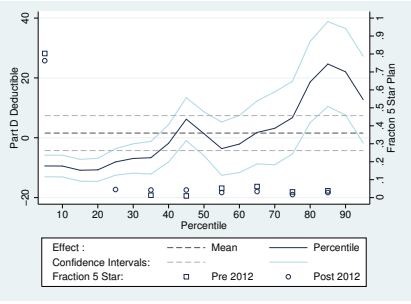
(f) Drug OOPC - Poor Health



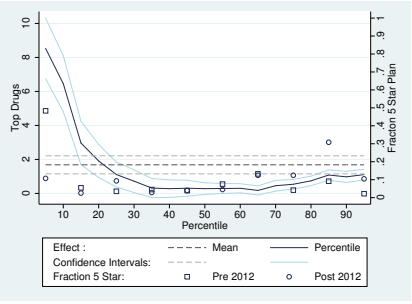
(g) Drug OOPC - Excellent Health



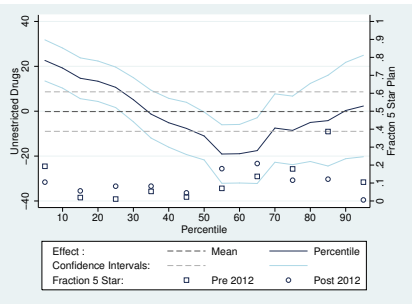
(h) Part D Deductible



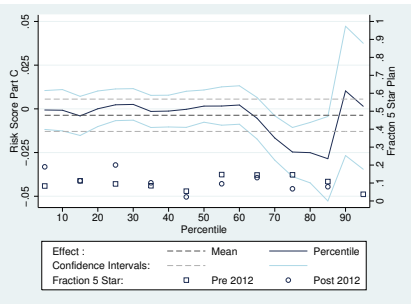
(i) N. Top Drugs



(j) N. Unrestricted Drugs



(k) Risk Score Part C



(l) Risk Score Part D