Mandatory Quality Disclosure and Forward-looking Firm Behavior

Ian M. McCarthy*
Emory University

November 2016

Abstract

Mandatory quality disclosure is pervasive across several industries and often includes a period over which the quality of new entrants is unreported. This provides the opportunity for forward-looking firms to adjust product characteristics in advance of disclosure. Using comprehensive data on Medicare Advantage from 2007-2014, I first demonstrate empirically that there exists a consumer response to quality disclosure and persistence in market shares over time. I then investigate the presence of forward-looking behavior, where I find that low-quality firms benefit from nondisclosure of quality by charging higher premiums and offering less variety across markets.

1 Introduction

Consumers have increasing access to a variety of quality measures when making purchasing decisions. Such quality measures derive from several sources, including self-disclosed quality via advertising and other brand management strategies, customer word-of-mouth and aggregated reviews from individual users (as published on Google,

*Correspondence to: Ian McCarthy, Emory University, Department of Economics, 1602 Fishburne Dr., Atlanta, GA 30322, U.S.A. E-mail: ian.mccarthy@emory.edu

Yelp, Rotten Tomatoes, etc.), third-party quality disclosure such as *Consumer Reports*, *U.S. News & World Report*, and other rating entities, and direct government regulation in the form of mandated disclosure or licensing. Consumers may also rely on their own personal experience in gauging product quality. How these sources of information influence consumer and firm decision making is the subject of a large theoretical and empirical literature (Dranove & Jin, 2010).

Third-party or mandatory rating systems often require a minimum amount of data available or a minimum period of data collection before a firm can be assigned a quality measure. This is the case in many healthcare applications, including hospital report cards and Medicare Advantage (MA) quality ratings issued by the Centers for Medicare and Medicaid Services (CMS). In these settings, quality disclosure is involuntary and fully anticipated by the firms, which introduces the possibility that forward-looking firms adjust their plan characteristics and plan offerings today in anticipation of quality disclosure in the future.

In this paper, I exploit the timing of quality disclosure in the MA rating system and examine the effect of anticipated quality disclosure on firm behaviors. I first demonstrate a demand-side response to nondisclosure of quality and then consider firm behaviors leading up to the publication of their quality ratings. The MA market is well-suited to examine these questions due to the complexity of health insurance plans, the subsequent importance of quality information to consumer decisions (Hibbard *et al.*, 1998; Abaluck & Gruber, 2011), and recent changes in this market to better disclose plan quality. The MA market is also a large and growing component of the U.S. healthcare system, with nearly 16 million individuals (30% of the Medicare population) currently enrolled in an MA plan for their health insurance benefits.[1] In a broader healthcare context, understanding the influence of quality measures on consumer and firm behaviors is critical as we move increasingly from a "volume-based" to a "value-based" healthcare

[1]This reflects a three-fold increase since the Medicare Modernization Act of 2003. Kaiser Family Foundation MA Update, available at http://kff.org/medicare/fact-sheet/medicare-advantage-fact-sheet/.

system. In this value-based system, reimbursement is tied at least in-part to quality, and this necessarily requires systematic and mandatory quality measures. The type of quality rating system currently in use for MA contracts is also used in many other healthcare markets, including nursing homes, dialysis clinics, hospital and physician report cards, and potentially to health insurance plans operating on the exchanges as part of the Affordable Care Act (ACA). Results based on the MA market may therefore inform policy in these other areas.

Using market-level data on MA enrollments, county demographics, and characteristics of the local hospital market, I estimate demand-side responses to quality disclosure using a nested logit demand model of differentiated products following Berry (1994). The dependent variable in these models is a plan's log market share relative to the log market share of traditional Medicare fee-for-service (FFS), which serves as a common outside option for all Medicare eligibles in all markets. Within this structure, I estimate the effects of nondisclosure in two ways. First, I consider a standard difference-in-difference (DD) approach, where the control group consists of the MA contracts receiving a star rating, the treatment group consists of contracts without a star rating, and the pre-post periods are delineated by the introduction of the overall MA star rating program in 2009. This analysis relies on MA enrollment data from 2007 through 2014, therefore covering a period before and after the introduction of the current overall rating system. Second, I estimate fixed effects models in which contracts without quality ratings in the current period but who ultimately received a star rating, s, are compared to contracts with a disclosed rating of s in the current period. In this second analysis, MA contracts are therefore compared based on quality, but contracts in one group have not yet had their quality rating revealed to the market. This analysis necessarily excludes years 2007 and 2008, during which no contracts received an overall star rating. The details of the identification strategy and sensitivity analysis exploit several unique aspects of the MA quality rating program, which I discuss in more detail in Section 2.

Consistent with the findings in Reid *et al.* (2013) and Darden & McCarthy (2015), I find a significant effect of quality ratings on enrollment, with low-quality plans benefitting from nondisclosure. Specifically, plans with an undisclosed star rating of less than 3-stars tend to enroll 43 additional beneficiaries per month due to nondisclosure of quality, while plans with an undisclosed star rating of 4 or above receive 256 fewer enrollments per month on average due to nondisclosure. Note also that my estimates speak to the effect of quality disclosure for an otherwise identical contract, whereas the estimates in Darden & McCarthy (2015) consider the enrollment effect from changes in reported quality for the same contract.

As discussed in more detail in Section 5, observing a response to *anticipated* quality disclosure requires not only that consumers are responsive to quality disclosure, but also that there exists some persistence in market shares over time. The health insurance market, and particularly the complexity of the Medicare Advantage market, is a natural setting in which to expect these mechanisms are at play. I examine the presence of share persistence in more detail in Section 5. I then examine changes in plan premiums, number of plans offered, and plan mix just prior to quality disclosure. Each of these supply-side responses speak to a firm's behavior in anticipation of their quality being revealed to the market (i.e., do contracts adjust their plan offerings or change premiums leading up to the disclosure of their quality ratings?).

The results reveal clear differences between low- versus high-quality contracts with regard to their response to quality disclosure. Low-quality contracts (below 3-stars) appear to take advantage of nondisclosure of quality with higher premiums in periods prior to quality disclosure, while higher quality contracts raise premiums after quality is fully disclosed, if at all. This behavior of high-quality firms is consistent with Hirth & Huang (2016), who find that the publication of quality star ratings for nursing home facilities caused highly rated nursing homes to raise their prices by over $3. Conversely, higher quality contracts more actively adjust their plan offerings leading up to quality disclosure.

I also find that insurers raise premiums in advance of quality disclosure when competing against a larger proportion of lower quality contracts. Similarly, in anticipation of quality disclosure, contracts offer fewer plans and a more homogeneous mix of plans across counties when confronted with a larger proportion of lower quality contracts. These results are consistent with a forward-looking firm who anticipates a reduction in share persistence over time due to the disclosure of quality.

My analysis contributes broadly to the literature on quality disclosure and consumer/firm behavior, and specifically to the growing study of quality disclosure in health insurance markets.[2] In a hypothetical scenario, Spranca *et al.* (2000) found that consumers with access to quality ratings were more likely to choose higher rated but less expensive, less comprehensive health insurance plans. Those without access to quality ratings were more likely to choose more expensive plans offering more comprehensive coverage. The study therefore examined the behavior of some individuals with access to quality ratings versus another set of individuals with no such access.

Beaulieu (2002) analyzed plan choice and switching behavior using data on Harvard University health insurance plans from 1995 through 1997. Comparing switching behaviors from 1995 to 1996 (when health plan quality information was not made available to enrollees) to switching behavior in 1996 to 1997 (when quality information was available), she found that enrollees responded (albeit modestly) to quality reports by switching away from lesser quality plans and concluded that quality reports provided additional information beyond what consumers independently obtained from experience. Several other studies have examined similar questions with different empirical techniques and datasets, including Scanlon *et al.* (2002), Wedig & Tai-Seale (2002), Jin & Sorensen (2006), Chernew *et al.* (2008), Dafny & Dranove (2008), Reid *et al.* (2013),

[2]The potential for supply-side responses to MA policy has received relatively little attention from researchers. One recent exception is Stockley *et al.* (2014), who examine how MA plan premiums and benefits respond to variation in MA benchmark payments. The authors find that contracts do not adjust premiums directly as a result of changes in benchmark payment rates but instead adjust the generosity of plan benefits. McCarthy & Darden (2016) also consider the supply-side response to quality ratings, with a focus on changes in premiums and plan entry/exit following a change in the reported quality of a given contract.

Darden & McCarthy (2015), and McCarthy & Darden (2016). These papers focus either exclusively on demand-side considerations, or they examine supply-side responses to changes in reported quality. Meanwhile, this paper acknowledges that firms may be more forward-looking, anticipating their future quality disclosure, rather than purely responding to existing ratings.

The remainder of the paper is organized as follows. In Section 2, I discuss the institutional details of the MA star rating program and relate this to my identification strategy. I discuss my data sources and overall summary statistics in Section 3. My analyses of the enrollment effects of quality disclosure and supply-side responses are presented in Sections 4 and 5, respectively. Section 6 considers the sensitivity of my analysis to the construction of my comparison groups, and Section 7 concludes.

2 The Medicare Advantage Star Rating Program

Since the passing of the Balanced Budget Act (BBA) of 1997, CMS has undergone a significant effort to better inform Medicare beneficiaries of the quality of health insurance plans available in their area. This quality information was initially limited to specific attributes. For example, an MA plan would be scored based on the percentage of women ages 50 to 69 who received a mammography within the past two years. The percentages for each plan in a beneficiary's area would then be included in the *Medicare and You* booklet. In 2007, CMS introduced a star rating system that provided a rating of one to five stars in each of five quality domains. This rating system essentially aggregated the ratings of specific plan attributes into each of the following domains: 1) "helping you stay healthy;" 2) "getting care from your doctors and specialists;" 3) "getting timely information and care from your health plan;" 4) "managing chronic conditions;" and 5) "your rights to appeal." These ratings were first reported for the 2008 open enrollment period, and were available in the *Medicare and You* booklet in addition to the Medicare Plan Finder website and through the Medicare helpline, 1-800-MEDICARE.

As part of the 2009 open enrollment period, CMS began aggregating the scores for individual attributes into an overall star rating for each MA contract. This overall star rating ranges from one to five stars in half-star increments. Although the underlying calculations have changed over time, this overall star rating system is still in place today. These ratings are clearly presented alongside other plan characteristics on the Medicare Plan Finder website and remain available in the *Medicare and You* booklet and through 1-800-MEDICARE.

Star ratings are calculated based on data collected from a variety of sources, including the Healthcare Effectiveness Data and Information Set (HEDIS), the Consumer Assessment of Healthcare Providers and Systems (CAHPS), the Health Outcomes Survey (HOS), the Independent Review Entity (IRE), the Complaints Tracking Module (CTM), and CMS administrative data. From these raw data, specific plan attributes are assigned a star rating typically based on the plan's percentile performance in the respective attribute, where the percentile thresholds delineating 1 through 5 stars differ across attributes. The star values for each attribute are then averaged and rounded to the nearest half-star to generate an overall star rating, after additional adjustments by CMS intended to reward consistency across individual attributes.

The MA market has a relatively unique structure which is critical to understanding the role of quality ratings. In particular, it is important to note the difference between an MA contract versus an MA plan. An MA contract is an agreement between a private insurance company and CMS whereby the company agrees to insure Medicare beneficiaries in exchange for some risk-adjusted payment per person from CMS. A contract is approved by CMS to operate in specific counties, and an approved contract typically offers a menu of MA plans that are differentiated by premium, prescription drug coverage, and if covered, the prescription drug deductible. Most MA contracts are required to offer at least one plan that includes prescription drug coverage. Consistent with this process, I use the term "contract" to refer to the private health insurance product that is approved by CMS to provide Medicare services through Medicare Advantage, and I

use the term "plan" to refer to specific products within a given MA contract.[3]

The CMS star ratings are calculated at the MA contract level so that plans operating under the same MA contract will receive the same star rating. This may not be clearly evident to a given Medicare beneficiary, as information is generally presented for each plan available in their area. For example, since the introduction of the overall star rating program, a beneficiary comparing plans on the Medicare Plan Finder website will be presented with information on several different plans meeting their search criteria, including premium, out-of-pocket limits, deductibles, copay/coinsurance rates, formulary restrictions (if applicable), and an overall star rating. To a given Medicare beneficiary, the star ratings may therefore appear as if they are plan specific, while in fact, all plans operating under a given contract will receive the same star rating in all counties in which the contract operates.[4]

By construction, star ratings will not be reported under two scenarios: 1) the contract has insufficient enrollments; or 2) the contract is too new to receive a quality rating.[5] Of the 17,100 contract-county observations indicated as "too new" to receive a rating from 2009 through 2012, over 12,000 (or 70%) ultimately received a star rating by 2014. Since most new contracts ultimately stay in the market long enough to receive a star rating, I estimate the quality of these new contracts at time t using the first observed star rating (at time $t+1$ or $t+2$). By construction, the MA star rating program relies on one- or two-year lagged measures when calculating a contract's star

[3]Insurers tend to operate multiple contracts in a given county. These contracts are often (thought not always) differentiated by network structure. For example, Aetna may offer one contract structured as an HMO, one as a PPO, and a third contract as a FFS. Each contract may then have a few plans within that contract, which may differ in terms of premiums and other out-of-pocket expenditures, prescription drug coverage, and other covered services.

[4]Beginning in 2012, plans offering prescription drug coverage were rated based on a larger set of underlying measures compared to plans without prescription drug coverage. As such, a given contract will tend to receive one star rating for its plans that do not participate in Part D and a potentially different star rating for its prescription drug plans. However, only the overall star rating (including or excluding Part D measures where relevant) is readily visible to a given beneficiary.

[5]For example, when calculating the star ratings for the 2009 open enrollment period, the "breast cancer screening" metric was based on data collected from January 2007 through December 2007. Contracts not yet approved during that time period would not have the necessary data available to calculate a star rating for this measure.

rating, and as such, the star rating at time $t+1$ or $t+2$ is intuitively reflective of the contract's underlying quality at time t. For contracts with disclosed quality, the lagged nature of the star rating is less relevant because enrollees will intuitively act on the information presented to them at that time, regardless of whether the underlying data were collected in prior periods. I exploit these features of the star rating program in order to estimate a contract's underlying (undisclosed) quality and compare outcomes among these contracts to those with identical underlying (disclosed) quality.[6]

3 Data

I collect data on MA market shares, contract/plan characteristics, and market area characteristics from several publicly available sources from 2007 through 2014. First, the set of all MA contracts in a given county are constructed from the Medicare Service Area files, which list all approved MA contracts in a county/month/year.[7] To these records, I merge enrollment and plan information at the contract/plan level from the MA enrollment files. I also merge county level MA penetration information to control for the prevalence of MA enrollment. Note that enrollment data are available monthly; however, there is little variation in enrollments across months due to the nature of the open enrollment process. I therefore take the average enrollment of each plan across months in a given year. The resulting unit of observation is the contract/plan/county/year.

Next, I merge quality information at the contract/year level, which includes star ratings for different domains of quality (e.g., helping you stay healthy), star ratings and continuous summary scores for each individual metric (e.g., percentage of women receiving breast cancer screening and an associated star rating), and an overall sum-

[6]Nonetheless, my analysis ultimately involves a form of imputation for contracts with undisclosed quality ratings, and I consider the sensitivity of my findings in Section 6.

[7]I use the Service Area files because the CMS enrollment files include individuals that move and keep their MA coverage despite the fact that a particular MA contract may not be approved in the new market area, and thus, not part of a potential enrollee's choice set. Data are available for download at www.cms.gov.

mary star measure beginning in 2009. I then merge plan premium information at the contract/plan/county/year level, county-level census demographic and socioeconomic information from the American Community Survey (ACS), and Medicare Advantage county benchmark rates from CMS. In addition, I collected hospital discharge data from the annual Hospital Cost Report Information System (HCRIS).

I present summary statistics for every other year of my data in Table 1. The measure of "Plan Mix" in Table 1 reflects the Euclidean distance between the vector of plan offerings in a given county relative to the average plan offerings across all other counties. Specifically, denote by y_{cm} a $1 \times J_{cm}$ vector of indicator variables, with each variable set to 1 if plan j is offered in market m. Similarly denote by $\bar{y}_{c,-m}$ the $1 \times J_{cm}$ vector of percentages of all other markets in which plan j is offered. Denoting the jth element of y_{cm} by y^j_{cm} and similarly for $\bar{y}^j_{c,-m}$, the distance function reflecting plan mix for contract c in market m is

$$d\left(y_{cm}, \bar{y}_{c,-m}\right) = \sqrt{\sum_{j=1}^{J_{cm}} \left(y^j_{cm} - \bar{y}^j_{c,-m}\right)^2}. \tag{1}$$

Higher values of $d\left(y_{cm}, \bar{y}_{c,-m}\right)$ therefore reflect larger variation in plan offerings in market m relative to other markets in which a given contract operates.[8]

TABLE 1

At least two salient features of the MA market emerge from these summary statistics. First, the MA market has become increasingly concentrated in recent years, with a spike in the total number of plan/county observations in 2009 and quickly dropping down to 95,505 and 62,031 in 2010 and 2011, respectively, with similar trends in the total number of plans per county. Consistent with these trends, average plan market

[8]I ultimately use plan mix as an outcome in the supply-side analysis. Since enrollments are also affected by quality disclosure, I consider the unweighted measure of plan mix rather than weighting by enrollments.

share decreased from 8.3% in 2007 to 6.1% in 2009, and increased back over 8% as the number of plans dropped. Enrollment per plan similarly dropped from 283 in 2007 to 246 in 2009, increasing to over 400 beneficiaries per plan per county in 2013. Monthly premiums (in excess of the Part B premium) remained relatively stable at around $43 until recently increasing to over $50.

Second, the types of plans available have become more homogeneous in many respects. For example, in 2007, less than 30% of plans were managed care and around 68% offered prescription drugs. In 2013, 80% of plans offered prescription drug coverage and over 75% of plans were managed care. At the contract/county level, there has also been a shift in average contract quality such that the large majority of contracts offered across the country are now 3 or 3.5-stars.[9] Similarly, contracts offer fewer plans per county (3.2 in 2007 versus 2.2 in 2013) and tend to offer the same plans across all counties in which the contract operates, with my measure of plan mix decreasing from 1.2 in 2009 to 0.89 in 2013. Put another way, around 6% of contracts in 2009 offered fully homogeneous plans in all counties (i.e., contracts with a plan mix of 0). This percentage increased to over 20% in 2012, subsequently decreasing to 14% and 12% in 2013 and 2014, respectively.

4 Consumer Response to Quality Disclosure

A small but significant beneficiary response to quality ratings has been shown in several studies using MA data, including Dafny & Dranove (2008), Reid *et al.* (2013), and Darden & McCarthy (2015). The difference in the current paper is that I am explicitly comparing contracts with disclosed quality versus contracts with (involuntary) undisclosed quality. Since previous studies have not explicitly made this comparison using the MA star rating system, I first demonstrate a demand-side response to published

[9]Note that these average star ratings are at the contract/county level rather than just the contract level, reflecting an average star rating weighted by prevalence across counties.

quality ratings.[10]

4.1 Methods

Following Berry (1994), Town & Liu (2003), Dafny & Dranove (2008), and others, I consider a discrete choice model in which a Medicare eligible individual maximizes her utility over a menu of Medicare options available in her market area. In all markets, an individual may opt for traditional Medicare FFS, which I define as the outside option $j = 0$. Alternatively, an individual in market area m may select a contract(plan), $c(j)$, from the set $J^m(i)$. Denote the utility of individual i from selecting Medicare option $c(j)$ in market area m at time t by

$$U_{ic(j)mt} = \delta_{c(j)mt} + \xi_{c(j)mt} + \zeta_{ig} + (1 - \sigma)\epsilon_{ic(j)mt}, \tag{2}$$

where $\delta_{c(j)mt}$ and $\xi_{c(j)mt}$ represent the mean level of utility derived from observed and unobserved contract-plan-market area characteristics, respectively. Following the nested logit structure of Berry (1994), I partition the set of Medicare options into four groups:1) MA managed care plans that offer prescription drug coverage (MC-PD plans); 2) MA managed care plans that do not offer prescription drug coverage (MC-Only plans); 3) MA fee-for-service plans that offer prescription drug coverage (FFS-PD plans); and 4) MA fee-for-service plans that do not offer prescription drug coverage (FFS-Only plans).[11] In addition to the i.i.d. extreme value error $\epsilon_{ic(j)mt}$, individual preferences are allowed to vary through group dummies ζ_{ig}. This nested logit structure relaxes the independence of irrelevant alternatives assumption and allows for differential substitution patterns between nests. The nesting parameter, σ, captures the within-group

[10] I also allow for heterogeneous effects of quality disclosure depending on the underlying (but unreported) rating of existing contracts. As such, the comparison group in some specifications consists not just of contracts with undisclosed quality, but contracts of specific quality levels which are as yet undisclosed.

[11] Although I present results based on this four-nest structure, my demand-side findings are unchanged when instead considering a two-nest structure delineated by plans offering prescription drug coverage versus plans without prescription drug coverage.

correlation of utility levels.

Berry (1994) shows how to consistently estimate the parameters of utility function (2) by integrating out the individual level variation in preferences. If we assume that $\epsilon_{ic(j)mt}$ follows a multivariate extreme value distribution, then from Cardell (1997), it follows that $\zeta_{ig} + (1-\sigma)\epsilon_{ic(j)mt}$ is also an extreme value random variable. The relative probability that an individual in market area m will select option $c(j)$, as compared to Medicare FFS, therefore has the following closed-form:

$$ln(P_{c(j)mt}) - ln(P_{0mt}) = \delta_{c(j)mt} + \sigma ln(P_{c(j)mt|g}) + \xi_{c(j)mt}.$$

Here, $P_{c(j)mt|g}$ is the conditional probability of an individual enrolling in option $c(j)$ within group g at time t. Applying market share data as empirical estimates of the probabilities yields our final estimation equation

$$ln(S_{c(j)mt}) - ln(S_{0mt}) = \delta_{c(j)mt} + \sigma ln(S_{c(j)mt|g}) + \xi_{c(j)mt}, \tag{3}$$

where $S_{c(j)mt}$ denotes the share of individuals (relative to all Medicare eligibles) enrolling in option $c(j)$ in market area m at time t, $S_{c(j)mt|g}$ denotes the within-group market share of option $c(j)$ at time t, and $\xi_{c(j)mt}$ denotes the mean utility derived from unobserved plan characteristics. I follow Town & Liu (2003) in treating observed product characteristics as exogenous after product fixed effects, and I instrument for premium and within-group shares using characteristics of the local hospital market as well as a contract's premium characteristics (minimum, maximum, and mean premiums) in other markets within the same state.

Within this econometric framework, I consider two alternative parameterizations of $\delta_{c(j)mt}$. The first,

$$\delta_{c(j)mt} = \beta x_{c(j)mt} + \nu_{c(j)m} + \gamma_p Post_t + \gamma_n New_{c(j)t} + \gamma_{np} Post_t New_{c(j)t}, \tag{4}$$

is a difference-in-difference (DD) model with plan-county fixed effects where $Post_t$ indicates the presence of the overall star rating system (beginning in 2009), $New_{c(j)t}$ denotes whether the contract was in operation for less than 2 years, and the final term is an interaction between $New_{c(j)t}$ and $Post_t$. Finally, $x_{c(j)mt}$ denotes a vector of county characteristics, plan premiums, as well as a count of the number of other plans in the county and the number of counties in which the current contract operates.

The second specification allows for differential effects of quality disclosure across the quality distribution, with

$$\delta_{c(j)mt} = \beta x_{c(j)mt} + \nu_{c(j)m} + \sum_{r \in 1,2} \left[\gamma_r Star_{c(j)rt} + \gamma_{rd} Disclosed_{c(j)t} Star_{c(j)rt} \right]. \quad (5)$$

This amounts to a standard fixed effects (FE) model with a series of indicator variables for different quality levels, interacted with indicators for whether star ratings were disclosed, $Disclosed_{c(j)t}$. Due to relatively small numbers of contracts receiving lower star ratings, I condense the star rating scale to $\{0, 1, 2\}$, where contracts with below 3 stars are assigned $r = 0$, contracts with 3 or 3.5 stars are assigned $r = 1$, and contracts with 4 stars or higher are assigned $r = 2$. This essentially divides the distribution of star ratings into low, average, and high-quality. For contracts with undisclosed ratings, I estimate $Star_{c(j)rt}$ with the the first observed star rating for that contract in future years. A contract that is too new in 2009 but ultimately receives a 3.5-star rating in 2011 is therefore estimated to be a 3.5-star contract in 2009 ($Star_{c(j)rt} = 1$).[12] This analysis only applies beginning in 2009 when the star rating system is in effect.

4.2 Results

Figure 1 presents kernel density estimates for the change in log relative market share before and after quality disclosure. The solid line reflects the kernel density for contracts disclosed as less than 3-stars, the dotted line reflects 3 and 3.5-star contracts, and the

[12] I consider the sensitivity of the results to the predicted star rating in Section 6.

dashed line presents kernel density estimates for 4 to 5-star contracts. The figure reflects a clear shift in the distribution of share changes by contract quality, with lower (higher) quality contracts seeing a reduction (increase) in market share following quality disclosure.

FIGURE 1

Regression results based on the specification in equations 4 and 5 are summarized in columns 1 and 2 of Table 2, respectively. For comparison with equation 4, my excluded comparison group consists of contracts with less than a 3-star rating, and I include an additional dummy variable to indicate if the contract's rating was undisclosed to the market. This indicator therefore measures the overall effect of undisclosed ratings, and the individual star rating indicators measure relative changes to the overall effect according to a contract's quality. With this adjustment, the $Undisclosed_{c(j)t}$ variable reflects the overall effect of nondisclosure on a plan's log market share (relative to traditional FFS) in both specifications.

TABLE 2

The results are very similar, with both specifications indicating a positive and significant effect of nondisclosure. Column 2 reveals that these positive effects are isolated among low-quality contracts. Specifically, contracts with underlying quality of less than 3-stars, but whose quality is not reported, receive a positive and significant increase in relative market share. This effect is smaller among undisclosed 3-star contracts, with a net negative effect among undisclosed 4, 4.5, and 5-star contracts. Consistent with the existing literature, column 2 also shows that higher rated contracts receive higher relative market shares once quality is disclosed (the "Disclosed Rating" panel). Finally, as expected, plan premiums have a significant negative effect on market shares, and there is a positive and significant correlation for within-group shares.

To better interpret the results, I translate the estimates in column 2 of Table 2 into effects on overall market shares and ultimately on predicted enrollments. Specifically, I estimate the mean observed utility, $\hat{\delta}_{c(j)m}$, setting the "Undisclosed Rating" indicator to 1 and again setting the indicator to 0 (with the appropriate switching of the star rating indicator variables as well). The estimated market shares in each scenario are then derived as follows (Berry, 1994):[13]

$$\begin{aligned}\hat{s}_{c(j)} &= \hat{s}_{c(j)|g} \times \hat{s}_g \\ &= \frac{\exp\left(\frac{\hat{\delta}_{c(j)}}{1-\hat{\sigma}}\right)}{\hat{D}_g} \times \frac{\hat{D}_g^{1-\hat{\sigma}}}{\sum_g \hat{D}_g^{1-\hat{\sigma}}} \\ &= \frac{\exp\left(\frac{\hat{\delta}_{c(j)}}{1-\hat{\sigma}}\right)}{\hat{D}_g^{\hat{\sigma}} \sum_g \hat{D}_g^{1-\hat{\sigma}}}, \end{aligned} \tag{6}$$

where

$$\hat{D}_g = \sum_{c(j) \in J^m} \exp\left(\frac{\hat{\delta}_{c(j)}}{1-\hat{\sigma}}\right) \text{ and } D_0 = 1.$$

Denoting by $\hat{s}_{c(j)}^{T=1}$ the predicted shares with undisclosed ratings and by $\hat{s}_{c(j)}^{T=0}$ the predicted shares with disclosed ratings, the estimated effect of disclosure on overall market shares is estimated by the average difference in these predicted values across all observations,

$$\triangle \hat{s}_{c(j)} = \frac{1}{N} \sum_j \left(\hat{s}_{c(j)}^{T=1} - \hat{s}_{c(j)}^{T=0}\right).$$

This is translated to effects on enrollments based on the total number of Medicare eligibles in the market. Consistent with the direction and size of the estimates in Table 2, I estimate an **increase** of 43 enrollments per month due to nondisclosure of quality for plans with an underlying star rating of 2.5 or below. Meanwhile, plans with an undisclosed star rating of 4 or above see a **decrease** of 256 enrollments per month due

[13] All share calculations are specific to a given market area, m, but I suppress the notation for simplicity.

to nondisclosure of quality.

Also included in the bottom panel of Table 2 are summary statistics from the first-stage regressions. These first-stage regressions yield high and significant global F-statistics, and a test of overidentifying restrictions yields a low and insignificant Hansen's J-statistic. The first-stage results therefore suggest that the instruments are highly correlated with premiums and within-group market shares and appropriate for this analysis. Nonetheless, the effects of quality disclosure do not appear to be sensitive to the endogeneity of premium or within-group shares, as the estimates from a standard linear fixed effects regression are similar to those from a fixed effects instrumental variables regression. These results are summarized in Appendix Table A.1.

5 Firm Response to Anticipated Quality Disclosure

In addition to the consumer response examined in Section 4, the presence of a supply-side response to *anticipated* quality disclosure requires some persistence in market shares over time, such that a firm's decisions in one period also influence shares in future periods. A large and growing empirical literature suggests that such persistence exists in a variety of differentiated product markets, including health insurance and Medicare in particular (Farrell & Klemperer, 2007; Abaluck & Gruber, 2011; Ketcham *et al.*, 2012; Handel, 2013; Ericson, 2014). My data are consistent with these findings as well. For example, including the lagged relative share as an additional covariate in my demand-side analysis yields a positive and significant coefficient of 0.263 (p-value $<$ 0.001), so that a 1% increase in market share relative to Medicare FFS in the prior year persists with a 0.3% increase in relative share in the current enrollment period.[14]

More formally, consider an existing plan j seeking to maximize the expected discounted present value of its profits in market m, which I assume is additively separable

[14]Other estimates in the demand-side specification are qualitatively unchanged when including lagged shares as an additional covariate, as summarized in Appendix Table A.2. I also consider a dynamic panel estimation using the Arellano-Bond estimator (Holtz-Eakin *et al.*, 1988; Arellano & Bond, 1991).

across geographic markets (Bresnahan & Reiss, 1991; Cawley *et al.*, 2005; Abraham *et al.*, 2007; Ericson, 2014):

$$V_{c(j)mt} = \left(P_{c(j)mt} + B_{mt} - AVC_{c(j)mt}\right) \times s_{c(j)mt} + \delta V_{c(j)m,t+1}\left(s_{c(j)mt}\right), \tag{7}$$

where $P_{c(j)mt}$ denotes plan j's premium (within contract c), B_{mt} denotes the benchmark payment rate from CMS in market m, $AVC_{c(j)mt}$ denotes the plan's average variable cost of enrolling and covering its beneficiaries in market m, $s_{c(j)mt}$ denotes the plan's expected quantity of Medicare beneficiaries in market m, δ denotes the insurer's discount factor, and $V_{c(j)m,t+1}(s_{c(j)mt})$ reflects the dependence of the contract's future profits on current shares. Plan premiums are then determined by the first order condition, with

$$P_{c(j)t} + B_{mt} - AVC_{c(j)mt} = -\frac{s_{c(j)mt}}{\frac{\partial s_{c(j)mt}}{\partial p_{c(j)mt}}} - \delta \frac{\partial V_{c(j)m,t+1}}{\partial s_{c(j)mt}}. \tag{8}$$

With price information fixed, disclosure of product quality will intuitively alter the degree of share persistence in the MA market. In the context of equation 8, persistence in market shares implies $\frac{\partial V_{c(j)m,t+1}}{\partial s_{c(j)mt}} > 0$, which will tend to reduce price-cost margins relative to a market with no such persistence. This is reflective of a standard investment motive in the "invest-then-harvest" literature (Farrell & Klemperer, 2007; Ericson, 2014). Anticipated quality disclosure in period $t+1$ should then reduce $\frac{\partial V_{c(j)m,t+1}}{\partial s_{c(j)mt}}$ for low-quality plans but possibly increase $\frac{\partial V_{c(j)m,t+1}}{\partial s_{c(j)mt}}$ for high-quality plans. This suggests that the investment motive from future quality disclosure (i.e., downward pressure on concurrent prices) is strongest for high-quality plans but relatively weak for low-quality plans, and premiums should then be lower for high-quality plans during the pre-disclosure period.

Importantly, contracts may also respond to anticipated quality disclosure in ways other than direct changes to plan premiums. For example, out of all unique plans offered throughout the U.S. from 2009 through 2014, 98% charge the same premium in

all counties in which the plan operates. Variation in premiums for the same plan across regions is therefore extremely low. Instead, the cross-sectional variation in premiums (within the same contract) derives from variation in a contract's plan offerings across markets. This suggests that plan mix, in addition to direct premium changes over time, is an important strategic variable for MA insurers. I therefore examine the effect of anticipated quality ratings on premiums at the plan level and on plan mix at the contract level, as well as the effect of anticipated ratings on the simple count of plans offered by a given contract.

5.1 Methods

I investigate the *supply-side* response to anticipated quality disclosure with a series of linear fixed effects regressions of the form

$$y_{c(j)mt} = \beta x_{mt} + \nu_{c(j)m} + \tau_t + \theta_{c(j)mt} + \varepsilon_{c(j)mt}, \tag{9}$$

where x_{mt} denotes a vector of market (county) characteristics, $\nu_{c(j)m}$ denotes plan or contract fixed effects (depending on the nature of the outcome variable), τ_t denotes year fixed effects, and $\theta_{c(j)mt}$ captures several terms relevant to plan/contract quality and quality disclosure. I specify θ as

$$\begin{aligned}\theta_{c(j)mt} =& \gamma_n New_{c(j)t} + \gamma_e Expect_{c(j)t} + \sum_{r\in 1,2}\Big[\gamma_r Star_{c(j)rt} + \gamma_{rd} Disclosed_{c(j)t} Star_{c(j)r} \\ &+ \gamma_{re} Expect_{c(j)t} Star_{c(j)rt}\Big] + \gamma_s Share_{c(j)mt} + \gamma_{se} Share_{c(j)mt} Expect_{c(j)t},\end{aligned} \tag{10}$$

where $New_{c(j)t}$ is an indicator for whether the contract is too new to receive a quality rating, $Expect_{c(j)t}$ is an indicator for whether contract/plan $c(j)$ will have its quality reported in the next period, $Star_{c(j)r}$ is an indicator set to 1 if contract $c(j)$ has a rating of r (based on the 0 to 2 scale discussed previously), $Disclosed_{c(j)t}$ is an indicator for whether the contract's rating has been disclosed, and $Share_{c(j)mt}$ reflects the percentage

of competing contracts in market m at the same or lesser quality rating to contract c. The latter two terms therefore account for the overall distribution of quality in the market, allowing a contract's behavior regarding their own quality to also depend on the quality of its competitors.

In this specification, $New_{c(j)t}$ captures any effect on $y_{c(j)mt}$ from initial entry into the market, with heterogeneities in these effects across underlying quality ratings captured by $Star_{c(j)rt}$. $Expect_{c(j)t}$ measures effects at time $t = 1$ relative to time $t = 0$, again with heterogeneous effects across quality ratings captured by $Expect_{c(j)t} \times Star_{c(j)rt}$. Finally, once the contract's quality is disclosed, the $New_{c(j)t}$ and $Expect_{c(j)t}$ indicator variables are set to 0, and heterogeneous effects of contract quality are captured by $Disclosed_{c(j)t} \times Star_{c(j)r}$.[15]

5.2 Results

Before proceeding to the regression analysis, I first examine overall trends in my outcomes of interest as contracts enter the market (at time $t = 0$) and ultimately have their quality disclosed (at time $t = 2$). Note that overall differences in levels of premiums, plan count, and plan mix are predominantly attributed to underlying differences in plan and contract characteristics. For example, a simple pooled linear regression of premiums on star ratings and year fixed effects yields a positive and significant effect of $17 per month for 3 or 3.5-star contracts relative to contracts of 2-stars or below, with a relative increase of $49 per month for contracts with 4-stars or more. This same regression allowing for plan/county fixed effects reveals a negative coefficient on the star rating indicators. Differences in premium levels by plan quality can therefore be explained by underlying differences in plan characteristics, which are inherently removed in my fixed effects analysis.

To provide a more appropriate initial comparison of my supply-side outcomes across

[15]Once a contract is rated, I also control for contract age with additional indicators for whether the contract has operated for between 4 and 8 years or 8+ years. Results are unchanged when instead including contract age and age squared as covariates rather than the indicator variables.

star ratings, I first predict the residuals from a linear fixed effects regression,

$$y_{c(j)mt} = \alpha + \nu_{c(j)m} + \varepsilon_{c(j)mt}.$$

I then plot the predicted residuals, $\hat{\varepsilon}_{c(j)mt}$, against the age of the contract, separately by the underlying star rating. The resulting trends of the residuals for premiums, number of plans offered per county, and plan mix per county are illustrated graphically in Figures 2, 3, and 4, respectively. Each figure presents a kernel-weighted local linear regression of $\hat{\varepsilon}_{c(j)mt}$ against the age of the contract, focusing on the first five years of the contract.

FIGURES 2-4

Figure 2 reveals a large decrease in premiums (relative to expected premium levels based on plan fixed effects) among higher rated contracts prior to quality disclosure, with a subsequent increase once quality is revealed. Conversely, low-quality contracts appear to temporarily increase their premiums relative to expected levels in anticipation of quality disclosure and slightly pull premiums back down once quality is fully disclosed. The results for number of plans (Figure 3) and plan mix (Figure 4) similarly illustrate changes in plan offerings prior to quality disclosure, particularly among high-quality contracts. Meanwhile, once quality is disclosed, it is the lower rated contracts that appear to more actively adjust their plan offerings. In all figures, residuals for 3 and 3.5-star contracts appear relatively stable both before and after quality disclosure, with much more variation over time for the low and high-quality contracts.

Recall that these figures reflect differences in the outcome of interest versus what would be expected based solely on time-invariant plan or contract characteristics. Therefore, although the figures are consistent with forward-looking behavior in anticipation of quality disclosure and heterogeneous responses across contracts, these summary-level results say little about the magnitude of the effects of quality and antic-

ipated quality disclosure, much less the effects after controlling for demographic characteristics of the county. To examine these effects more formally, my regression results are presented in Table 3 and divided into two specifications for each of three different outcomes (premiums, number of plans offered, and plan mix as measured by equation 1). Results based on the full specification in equation 10 are presented in column 2 for each outcome. The first column for each outcome instead focuses on the overall effect of anticipated quality disclosure, excluding the interaction between the contract's underlying rating and the indicator for anticipated disclosure.

TABLE 3

With regard to premiums, the results show that new contracts initially charge lower premiums (about $2 per month on average), particularly for higher quality plans as evident by the large negative effects in the "Underlying Quality" panel. Low-quality plans then increase their premiums by approximately $4.43 per month in advance of quality disclosure. Conversely, higher quality plans tend to decrease premiums in the period just before quality disclosure, as reflected in the bottom panel of the table. Once quality is disclosed, there is some evidence that high-quality plans (4-stars or more) increase premiums relative to lower quality plans, although this effect is insignificant in the full specification in column 2. These effects are generally consistent with the role of share persistence discussed previously.

The results for number of plans and plan mix are less clear, particularly once quality is disclosed. However, I consistently find large and significant effects of anticipated disclosure. Specifically, new contracts on average offer more plans than older contracts, with a significant increase of 0.15 plans per county in the period prior to quality disclosure. This increase in plan offerings is driven by 3 and 3.5-star contracts, with an additional positive (but insignificant) change among 4- to 5-star contracts. I also find a significant increase in plan mix leading up to quality disclosure, again with much larger effects among higher quality contracts. Collectively, these results suggest that

average and high-quality contracts tend to expand their plan offerings and provide a more heterogeneous mix of plans across counties in anticipation of quality disclosure. Changes in plan offerings and plan mix are less pronounced once quality is disclosed.

Finally, the coefficients for "% Low Quality" reveal differential effects according to the distribution of existing quality in the market. For example, the overall effect of \$4.74 for % low quality on premiums means that contracts tend to raise premiums by just over \$1 per month following a one standard deviation increase in the percentage of equal or lesser quality contracts in the market. This differential effect is most pronounced just prior to a contract's quality being disclosed and driven by contracts with underlying ratings of less than 3 stars. Similar results emerge for plan count and plan mix, where contracts are particularly responsive to the existing distribution of quality just prior to their quality being disclosed to the market.

6 Sensitivity to Quality Comparisons

Allowing for differential responses to quality disclosure according to a contract's underlying quality rating necessarily requires some estimate of otherwise unobserved quality. As discussed in Section 2, I estimate a contract's underlying rating based on that contract's first observed star rating in the data. This assignment is supported by two empirical facts in the MA market. First, relatively few contracts (less than 30%) entirely exit the MA market before ever receiving a star rating. Second, star ratings are calculated based on lagged values of underlying metrics, so that the final disclosed rating at time t is essentially based on underlying quality at time $t-1$ and $t-2$. In this sense, looking one or two periods into the future is reflective of a contract's current underlying quality.

However, the lagged nature of the star rating system then calls into question whether the current star rating (once disclosed) is appropriate as a measure of current quality. For example, consider a contract with a disclosed 3-star rating in 2009 and 4-star rating

in 2010. This means that the contract's underlying quality measures in 2008 and 2009 were higher than the measures underlying its initial 3-star rating, and it is unclear to what extent consumers already knew of the improved quality of the plan or if they acted more on information reflected by the current star rating.

This lagged structure ultimately speaks to the appropriateness of my control group (i.e., the contracts with disclosed ratings) when allowing for differential effects by star rating. To assess the sensitivity of my results to this issue, I consider two additional analyses. First, I re-estimate my supply-side analysis from Table 3, excluding any measures of predicted quality for plans with undisclosed ratings. This analysis avoids splitting the treatment and control groups based on quality, and essentially identifies a weighted effect of anticipated quality disclosure across all contracts. The results are summarized in Table 4. Consistent with the initial findings, I still find a significant effect of anticipated quality disclosure on premiums, the number of plans offered, and plan mix.

TABLE 4

Second, I limit my analysis only to those contracts with stable ratings once disclosed (i.e., contracts whose ratings do not change over a two-year period). For such contracts, their star rating at time t is the same as their star rating at time $t+1$, suggesting that the underlying quality measures are also sufficiently similar over the prior two years. In this way, a 3-star contract with a disclosed rating at time t remains similar in terms of underlying quality to a contract whose 3-star rating is not disclosed until $t+1$. Results based only on these contracts are summarized in Table 5. The results for the anticipated disclosure effects in the bottom panel of the table are qualitatively similar to my initial findings in Table 5. Coefficients on the quality rating indicators are generally larger in magnitude relative to the initial findings. This is not surprising since the sample is constructed of plans that are consistently in the same overall quality range. For example, the excluded group of contracts receiving less than a 3-star rating

must have consistently received such a rating over a two-year period, in which case this comparison group is of lesser quality on average compared to the initial analysis which included contracts that moved from low-quality to average quality over the same time period. As such, quality effects should be more pronounced in this analysis as the quality rankings are more strongly delineated. Indeed, I estimate that higher quality plans in this sample charge significantly higher premiums once quality is disclosed, with \$6-\$9 more per month for 3 to 3.5-star contracts and \$17-\$20 more per month for plans with 4 stars or more.

TABLE 5

7 Conclusion

Quality ratings are available across a variety of industries in the U.S. and are increasingly available throughout different areas of the healthcare sector. A common system in healthcare employs some form of star rating in which a series of individual measures are aggregated into an overall score assigned to a given provider or insurer. For participants in Medicare, CMS has pursued this form of star rating system for nursing homes, dialysis clinics, hospitals and physicians, and insurance plans operating through Medicare Advantage. Yet despite its prevalence, relatively little is known about how firms may respond to these rating systems.

In this paper, I am particularly interested in firm behaviors *prior* to quality disclosure. I have in mind a theoretical structure where market shares are persistent over time, and where a forward-looking firm anticipates a reduction in this share persistence due to the disclosure of product quality. This framework predicts that firms will adjust product characteristics in anticipation of quality being revealed, although the direction and magnitude of these adjustments are ultimately empirical questions.

Applied to the MA market, my results suggest that high-quality plans tend to reduce

premiums while quality remains undisclosed, during which they also more actively adjust their plan offerings. Low-quality plans, meanwhile, increase premiums in advance of quality disclosure. Regarding a contract's relative quality, insurers are most responsive to competitors' quality *before* their own quality is disclosed, with smaller and often insignificant effects once quality is disclosed. Essentially, by the time their quality is disclosed, higher quality firms have already incorporated the effects of competitor quality on their own plan characteristics and plan offerings.

These adjustments due to future quality disclosure may or may not improve consumer welfare, depending on the alignment of plan offerings with underlying beneficiary preferences for product variety. The welfare effects from changes in plan premiums are similarly ambiguous, as any welfare reduction due to an increase in premiums for low (undisclosed) quality plans is somewhat offset by the decrease in premiums among higher (undisclosed) quality plans. Given the increasing prevalence and use of quality rating systems throughout the U.S. healthcare system, examining these welfare effects more formally is an important topic of future research.

There are a few interrelated mechanisms that may be driving my estimates. During the period of nondisclosure, firms are learning about their underlying quality based on observed patterns of care and customer behavior. Firms are also learning about how their product characteristics align with product quality (e.g., is the product overpriced relative to the quality of care provided, or is the contract offering excessive variety?). To the extent that firms know their underlying quality and have adjusted plan characteristics accordingly, my estimates can be interpreted as strategic behavior due to future quality disclosure; however, to the extent that firms do not fully know their underlying quality, my estimates reflect the combination of continued learning of quality and the disclosure of this (uncertain) quality in the future period. In this case, the estimates also speak to a firm's attitude toward risk, where uncertainty in the future quality rating drives changes to plan characteristics in the current period.

In reality, my estimates reflect a combination of learning and strategic behavior. For

example, after one year on the market, firms will have access to standard process of care measures collected throughout their first year of operation. These same data will be used in calculating the firm's star rating in the following year. Meanwhile, consumer survey data underlying the star rating will not be collected until the year prior to disclosure. Star ratings are also based on the distribution of other contract's measures across the country. As such, there is inherent uncertainty regarding a contract's star rating built into the CMS calculations. This uncertainty is reduced after one year of operation, at which point firms may better predict their future star rating but remain undisclosed. The magnitude of effects in advance of quality disclosure relative to estimates after quality disclosure suggests that, even though firms may still be learning about their underlying quality, they are particularly aggressive in adjusting plan characteristics in the year just prior quality disclosure. This appears more consistent with strategic behavior versus learning.

My findings have at least two important policy implications. First, U.S. healthcare policy has increasingly relied on the private provision of public health insurance benefits, highlighted by the growing prevalence of Medicaid managed care and Medicare Advantage. Similar policies are regularly debated in areas of social security benefits and public education. One of many ways that commercial products may behave differently, as highlighted in the current paper, is by incorporating future information into current product offerings. Policy evaluation in these markets based only on a pre-post analysis may therefore be misleading. Second, there are attempts in many areas of healthcare and education to tie funding to some measure of quality. In addition to directly affecting firms' current profits, these policies will also tend to reinforce the role of learning and strategic behavior prior to quality assessment.

References

Abaluck, Jason, & Gruber, Jonathan. 2011. Heterogeneity in choice inconsistencies among the elderly: evidence from prescription drug plan choice. *The American Economic Review*, **101**(3), 377–381.

Abraham, Jean, Gaynor, Martin, & Vogt, William B. 2007. Entry and Competition in Local Hospital Markets. *The Journal of Industrial Economics*, **55**(2), 265–288.

Arellano, Manuel, & Bond, Stephen. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, **58**(2), 277–297.

Beaulieu, N.D. 2002. Quality information and consumer health plan choices. *Journal of Health Economics*, **21**(1), 43–63.

Berry, Steven T. 1994. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Bresnahan, Timothy F, & Reiss, Peter C. 1991. Entry and competition in concentrated markets. *Journal of Political Economy*, 977–1009.

Cardell, N.S. 1997. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, **13**(02), 185–213.

Cawley, John, Chernew, Michael, & McLaughlin, Catherine. 2005. HMO participation in Medicare+ Choice. *Journal of Economics & Management Strategy*, **14**(3), 543–574.

Chernew, M., Gowrisankaran, G., & Scanlon, D.P. 2008. Learning and the value of information: Evidence from health plan report cards. *Journal of Econometrics*, **144**(1), 156–174.

Dafny, L., & Dranove, D. 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, M., & McCarthy, I. 2015. The Star Treatment: Estimating the Impact of Star Ratings on Medicare Advantage Enrollments. *Journal of Human Resources*, **50**(4), 980–1008.

Dranove, David, & Jin, Ginger Zhe. 2010. Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature*, **48**(4), 935–963.

Ericson, Keith M. 2014. Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange. *American Economic Journal: Economic Policy*, **6**(1), 38–64.

Farrell, Joseph, & Klemperer, Paul. 2007. Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, **3**, 1967–2072.

Handel, Benjamin R. 2013. Adverse selection and inertia in health insurance markets: When nudging hurts. *The American Economic Review*, **103**(7), 2643–2682.

Hibbard, J.H., Jewett, J.J., Engelmann, S., & Tusler, M. 1998. Can Medicare beneficiaries make informed choices? *Health Affairs*, **17**(6), 181–193.

Hirth, Richard, & Huang, Sean. 2016. *Quality Rating and Private-Prices: Evidence from the Nursing Home Industry.* Working Paper. University of Michigan.

Holtz-Eakin, Douglas, Newey, Whitney, & Rosen, Harvey S. 1988. Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, 1371–1395.

Jin, G.Z., & Sorensen, A.T. 2006. Information and consumer choice: the value of publicized health plan ratings. *Journal of Health Economics*, **25**(2), 248–275.

Ketcham, Jonathan D, Lucarelli, Claudio, Miravete, Eugenio J, & Roebuck, M Christopher. 2012. Sinking, swimming, or learning to swim in Medicare Part D. *The American Economic Review*, **102**(6), 2639–2673.

McCarthy, I., & Darden, M. 2016. Supply-side Responses to Public Quality Ratings: Evidence from Medicare Advantage. *American Journal of Health Economics*, **forthcoming**.

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

Scanlon, D.P., Chernew, M., McLaughlin, C., & Solon, G. 2002. The impact of health plan report cards on managed care enrollment. *Journal of Health Economics*, **21**(1), 19–41.

Spranca, M., Kanouse, D.E., Elliott, M., Short, P.F., Farley, D.O., & Hays, R.D. 2000. Do consumer reports of health plan quality affect health plan selection? *Health Services Research*, **35**(5 Pt 1), 933.

Stockley, Karen, McGuire, Thomas, Afendulis, Christopher, & Chernew, Michael E. 2014. *Premium Transparency in the Medicare Advantage Market: Implications for Premiums, Benefits, and Efficiency.* Working Paper. National Bureau of Economic Research.

Town, Robert, & Liu, Su. 2003. The welfare impact of Medicare HMOs. *RAND Journal of Economics*, 719–736.

Wedig, G.J., & Tai-Seale, M. 2002. The effect of report cards on consumer choice in the health insurance market. *Journal of Health Economics*, **21**(6), 1031–1048.

8 Tables and Figures

Table 1: **Summary Statistics**

	2007	2009	2011	2013
Plan/County Data				
Enrollment	283	246	359	420
	(1,382)	(1,266)	(1,494)	(1,643)
MA Market Share	8.3%	6.1%	8.7%	8.3%
	(0.144)	(0.097)	(0.125)	(0.124)
Premium	44.04	42.10	42.31	54.00
	(42.52)	(43.06)	(46.64)	(53.69)
Drug Coverage	67.7%	65.7%	75.6%	79.6%
HMO	13.9%	19.4%	29.7%	35.7%
PPO	13.77%	15.4%	40.9%	45.0%
Observations[a]	75,677	113,227	62,031	60,639
Contract/County Data				
Star Rating 1.5 to 2.5		48.9%	15.2%	12.3%
Star Rating 3 to 3.5		25.9%	53.8%	59.4%
Star Rating 4 to 5		3.4%	12.5%	24.5%
Number of Plans	3.24	3.18	2.25	2.19
	(3.29)	(2.80)	(1.81)	(1.85)
Plan Mix	1.15	1.20	0.86	0.89
	(0.74)	(0.69)	(0.62)	(0.61)
New Contract	27.5%	4.0%	14.4%	1.7%
Observations	23,366	35,615	27,615	27,701
County Data				
MA Penetration	0.165	0.153	0.165	0.194
	(0.114)	(0.109)	(0.119)	(0.127)
Number of Plans	30.32	36.07	19.98	19.55
	(22.30)	(29.03)	(20.19)	(21.93)
Population (1,000s)	157.34	96.06	98.70	100.35
	(394.00)	(309.52)	(312.71)	(318.39)
Percent >65	0.137	0.153	0.158	0.164
	(0.034)	(0.043)	(0.042)	(0.043)
Employed Full Time	0.375	0.378	0.378	0.373
	(0.053)	(0.061)	(0.062)	(0.063)
White	0.827	0.839	0.841	0.840
	(0.145)	(0.165)	(0.163)	(0.163)
Black	0.093	0.089	0.091	0.091
	(0.126)	(0.144)	(0.146)	(0.146)
College Graduate	0.132	0.124	0.127	0.130
	(0.056)	(0.053)	(0.053)	(0.054)
Observations	1,817	3,138	3,104	3,102

[a]Due to missing enrollment data, enrollments and market shares are available for 18,826 observations in 2007, 27,253 observations in 2009, 22,122 observations in 2011, and 22,714 observations in 2013.

Table 2: **Fixed Effects IV Regression Results for MA Shares**[a]

	Overall Effects	By Star Rating
Premium	-0.005***	-0.002***
	(0.001)	(0.001)
$ln\left(S_{c(j)m\|g}\right)$	0.809***	0.854***
	(0.122)	(0.093)
Contract Age < 2 Years	-0.199***	
	(0.017)	
Post 2009	0.098***	
	(0.008)	
Undisclosed × Post	0.320***	0.219***
	(0.024)	(0.027)
Underlying Rating		
3 to 3.5-star		-0.013
		(0.031)
4-star or more		-0.320***
		(0.043)
Disclosed Rating ×		
3 or 3.5-star		0.092***
		(0.028)
4-star or more		0.312***
		(0.039)
Observations	118,793	101,192
First-stage IV Results		
Global F-statistic:		
Premium	1,507	921
	(0.000)	(0.000)
$ln\left(S_{c(j)m}\right)$	99.20	123.42
	(0.000)	(0.000)
Hansen's J-statistic	1.763	1.582
	(0.623)	(0.664)

[a]Results based on linear fixed effects instrumental variable regressions, with standard errors in parentheses clustered at the county level. Premium and within-group shares were instrumented with number of hospitals in the county, the hospital HHI in the county, and the minimum, maximum, and mean premium in the contract across all other counties in the state. Additional independent variables not in the table include county demographics (measures total population, age, race, income, education, and employment), contract age (indicator variables for contracts 4-8 years old or more than 8 years old), number of other counties in which the contract operates, number of other plans offered by the contract in the same county, and year fixed effects. * p<0.1. ** p<0.05. *** p<0.01.

Table 3: **Fixed Effects Regression Results for Supply-side Outcomes**[a]

	Premiums		Plan Count		Plan Mix	
Contract Age < 2 Years	-1.958*	-3.724**	0.154**	0.162	0.016	-0.044
	(1.079)	(1.611)	(0.074)	(0.118)	(0.028)	(0.041)
Anticipated Disclosure	2.058***	4.429***	0.148***	0.115	0.044***	0.120***
	(0.444)	(1.305)	(0.038)	(0.124)	(0.012)	(0.038)
% Low Quality	4.743***	2.349	-0.047	0.322	-0.032	0.202***
	(1.360)	(2.328)	(0.117)	(0.210)	(0.037)	(0.066)
Underlying Rating						
3 to 3.5-star	-8.952***	-5.864***	-0.205**	-0.422***	-0.044	-0.104**
	(1.186)	(1.616)	(0.090)	(0.126)	(0.034)	(0.046)
4-star or more	-12.674***	-6.995**	-0.112	-0.254	-0.007	-0.099
	(1.948)	(2.977)	(0.126)	(0.215)	(0.048)	(0.073)
Disclosed Rating ×						
3 to 3.5-star	0.619	-2.489	0.133	0.353***	0.000	0.062
	(1.172)	(1.615)	(0.088)	(0.124)	(0.034)	(0.046)
4-star or more	7.767***	2.039	-0.008	0.138	-0.022	0.072
	(1.933)	(2.982)	(0.124)	(0.215)	(0.046)	(0.072)
% Low Quality	-0.982	1.485	0.231**	-0.141	0.055	-0.181***
	(1.325)	(2.299)	(0.113)	(0.209)	(0.036)	(0.066)
Anticipated Disclosure ×						
3 to 3.5-star		-4.755***		0.433***		0.121**
		(1.435)		(0.131)		(0.047)
4-star or more		-9.255***		0.302		0.164**
		(3.018)		(0.200)		(0.070)
% Low Quality		4.292*		-0.646***		-0.357***
		(2.532)		(0.213)		(0.067)
Observations	101,062		62,313			

[a]Results based on linear fixed effects regressions, with standard errors in parentheses clustered at the county level. Additional independent variables not in the table include county demographics (measures total population, age, race, income, education, and employment), contract age (indicator variables for contracts 4-8 years old or more than 8 years old), measures of the hospital market (number of hospitals, number of hospital beds, and the hospital HHI at the county level), the number of total Medicare Advantage enrollees in the county, the Medicare Advantage benchmark rate, and year fixed effects. * p<0.1. ** p<0.05. *** p<0.01.

Table 4: **Fixed Effects Regression Results for Supply-side Outcomes (without predicted underlying quality ratings)**[a]

	Premiums	Plan Count	Plan Mix
Contract Age < 2 Years	-7.748***	-0.032	-0.033**
	(0.568)	(0.043)	(0.014)
Anticipated Disclosure	2.709***	0.149***	0.042***
	(0.436)	(0.037)	(0.012)
Disclosed Rating ×			
3 to 3.5-star	-7.917***	-0.064**	-0.042***
	(0.442)	(0.025)	(0.008)
4-star or more	-4.109***	-0.114***	-0.029**
	(0.835)	(0.039)	(0.013)
% Low Quality	3.803***	0.193***	0.026**
	(0.724)	(0.037)	(0.012)
Observations	101,062	62,313	

[a]Results based on linear fixed effects regressions, with standard errors in parentheses clustered at the county level. Additional independent variables not in the table include county demographics (measures total population, age, race, income, education, and employment), contract age (indicator variables for contracts 4-8 years old or more than 8 years old), measures of the hospital market (number of hospitals, number of hospital beds, and the hospital HHI at the county level), the number of total Medicare Advantage enrollees in the county, the Medicare Advantage benchmark rate, and year fixed effects. * p<0.1. ** p<0.05. *** p<0.01.

Table 5: **Fixed Effects Regression Results for Supply-side Outcomes (among contracts with stable quality ratings)**[a]

	Premiums		Plan Count		Plan Mix	
Contract Age < 2 Years	1.144	4.185**	0.417***	0.428**	0.136**	0.111*
	(1.757)	(2.127)	(0.155)	(0.214)	(0.056)	(0.066)
Anticipated Disclosure	2.011***	-1.874	0.233***	0.221	0.066***	0.100**
	(0.559)	(1.165)	(0.050)	(0.174)	(0.016)	(0.044)
% Low Quality	11.565***	1.614	-0.714***	-0.338	-0.266***	-0.012
	(2.076)	(2.505)	(0.164)	(0.230)	(0.052)	(0.077)
Underlying Rating						
3 to 3.5-star	-19.607***	-17.493***	-0.182	-0.405*	-0.053	-0.160**
	(2.153)	(2.214)	(0.164)	(0.219)	(0.063)	(0.073)
4-star or more	-27.059***	-24.630***	0.393	0.172	-0.035	-0.242*
	(3.486)	(3.740)	(0.261)	(0.349)	(0.105)	(0.133)
Disclosed Rating ×						
3 to 3.5-star	8.783***	6.641***	0.258	0.485**	0.030	0.140*
	(2.151)	(2.215)	(0.160)	(0.214)	(0.063)	(0.073)
4-star or more	19.943***	17.350***	0.084	0.324	0.149	0.366***
	(3.422)	(3.710)	(0.241)	(0.336)	(0.100)	(0.131)
% Low Quality	-7.553***	2.474	0.911***	0.525**	0.286***	0.026
	(2.407)	(2.733)	(0.165)	(0.235)	(0.054)	(0.080)
Anticipated Disclosure ×						
3 to 3.5-star		-4.692***		0.426**		0.203***
		(1.496)		(0.183)		(0.057)
4-star or more		-5.248		0.410*		0.354***
		(3.312)		(0.237)		(0.084)
% Low Quality		15.574***		-0.686***		-0.420***
		(2.770)		(0.229)		(0.076)
Observations	41,530		25,790			

[a]Results based on linear fixed effects regressions, with standard errors in parentheses clustered at the county level. "Stable" quality rating is defined as a contract whose quality does not change over a two-year period. Additional independent variables not in the table include county demographics (measures total population, age, race, income, education, and employment), contract age (indicator variables for contracts 4-8 years old or more than 8 years old), measures of the hospital market (number of hospitals, number of hospital beds, and the hospital HHI at the county level), the number of total Medicare Advantage enrollees in the county, the Medicare Advantage benchmark rate, and year fixed effects. * p<0.1. ** p<0.05. *** p<0.01.

Figure 1: Share Change following Quality Disclosure

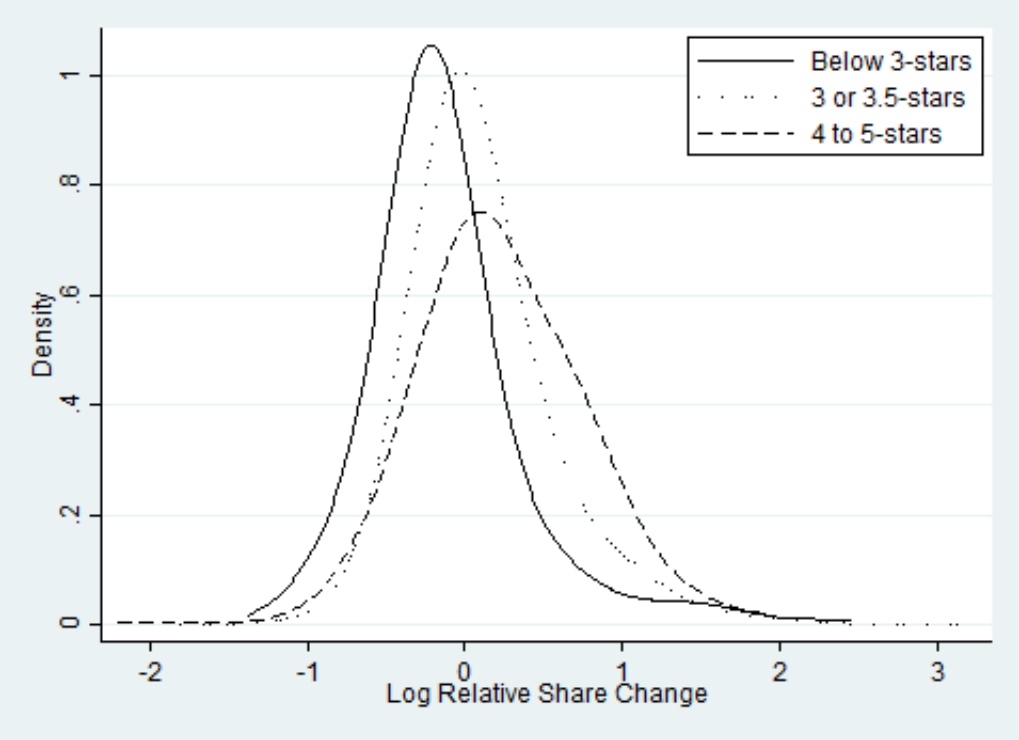


Figure 2: Premiums by Age of Contract and Underlying Rating

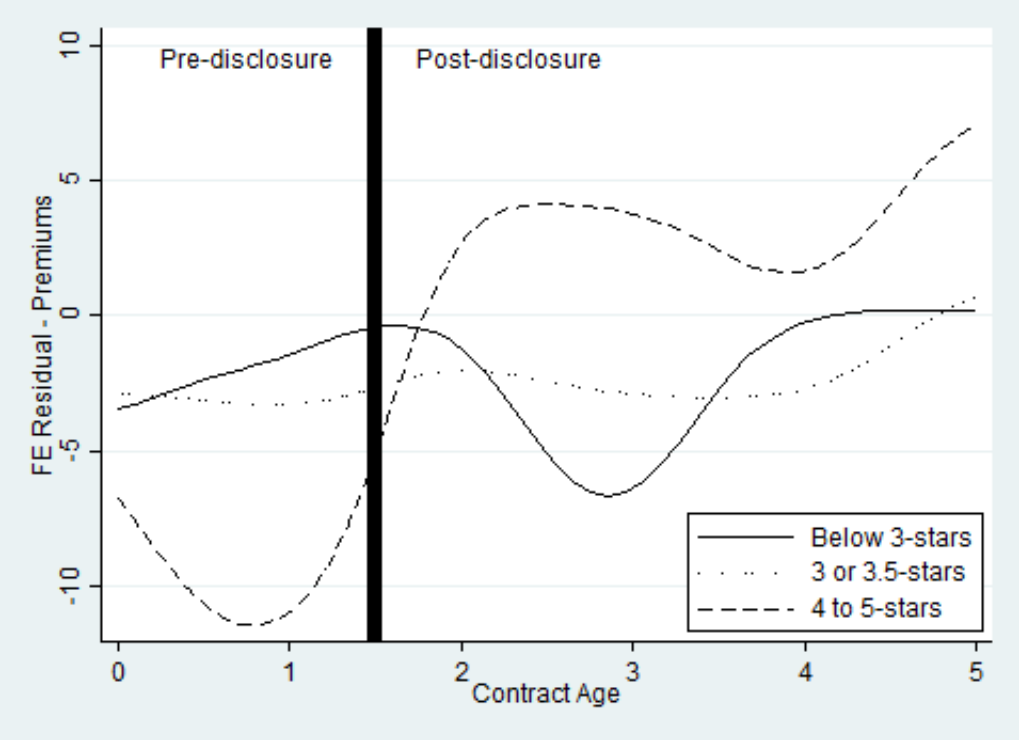


Figure 3: Plan Count by Age of Contract and Underlying Rating

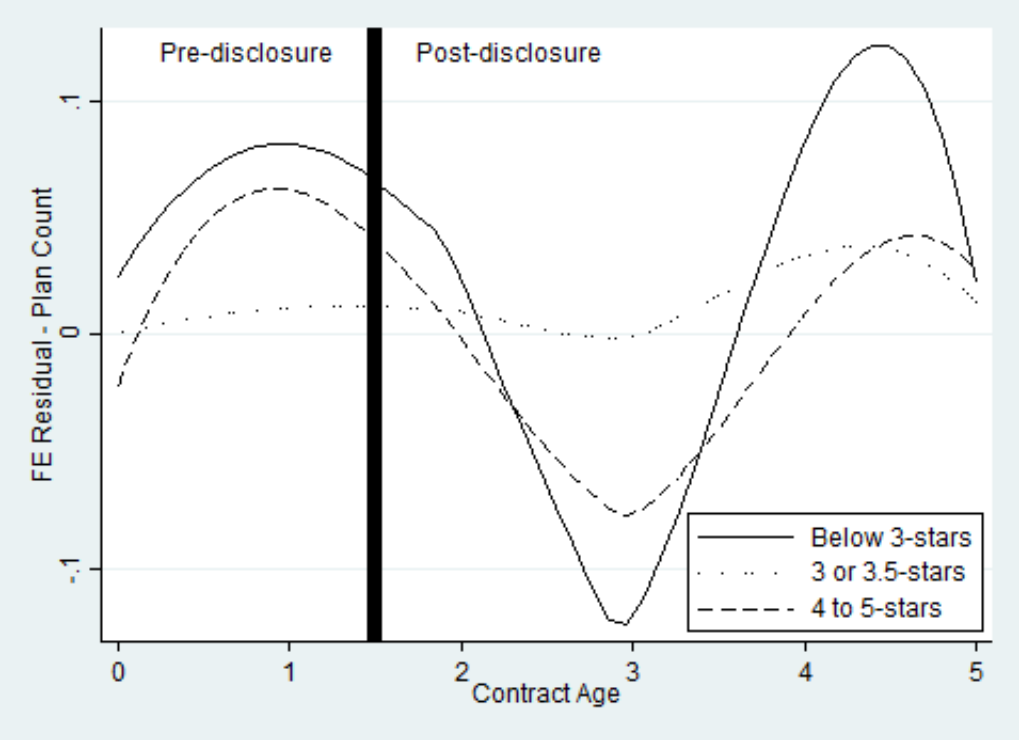
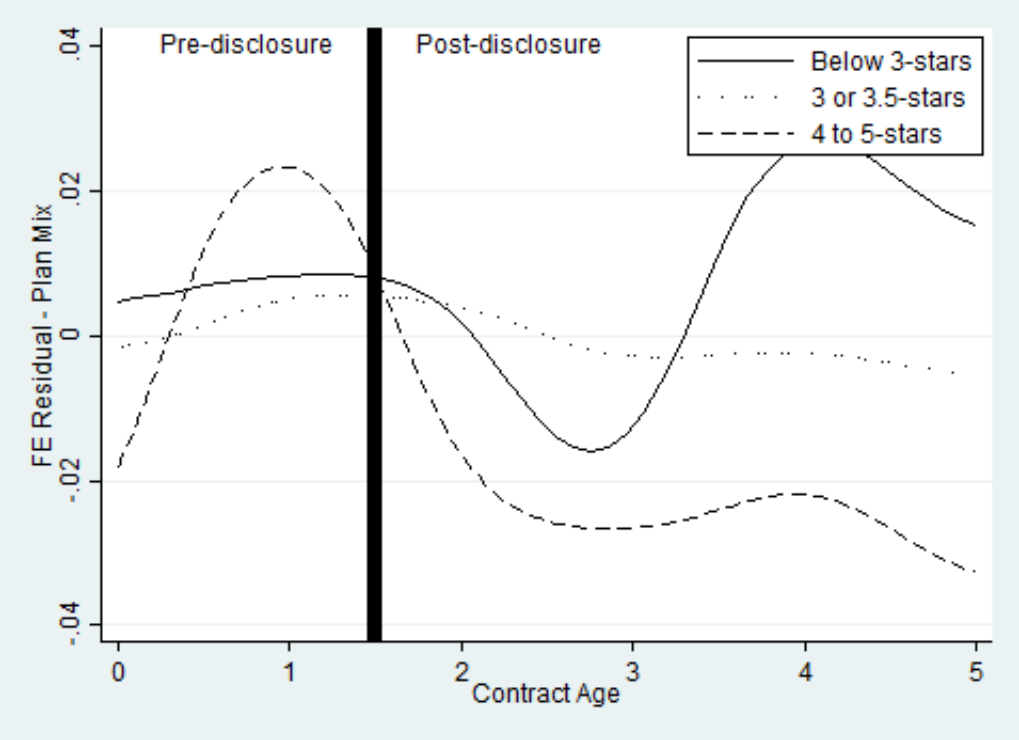


Figure 4: Plan Mix by Age of Contract and Underlying Rating



A Appendix Tables

Table A.1: **Fixed Effects Regression Results for MA Shares**[a]

	Overall Effects	By Star Rating
Premium	-0.003***	-0.003***
	(0.000)	(0.000)
$ln\left(S_{c(j)m\|g}\right)$	0.679***	0.729***
	(0.007)	(0.007)
Contract Age < 2 Years	-0.215***	
	(0.014)	
Post 2009	0.079***	
	(0.008)	
Undisclosed × Post	0.323***	0.215***
	(0.016)	(0.024)
Underlying Rating		
3 to 3.5-star		-0.024
		(0.027)
4-star or more		-0.346***
		(0.037)
Disclosed Rating ×		
3 or 3.5-star		0.105***
		(0.026)
4-star or more		0.337***
		(0.035)
Observations	156,774	115,149

[a]Results based on linear fixed effects, with standard errors in parentheses clustered at the county level. * p<0.1. ** p<0.05. *** p<0.01.

Table A.2: **Fixed Effects Results for MA Shares with Share Persistence**[a]

	Standard Fixed Effects		
	with IV	without IV	Arellano-Bond
$ln\left(S_{c(j)m,t-1}\right)$	0.263***	0.271***	0.236***
	(0.051)	(0.007)	(0.014)
Premium	-0.004***	-0.004***	-0.004***
	(0.001)	(0.000)	(0.000)
$ln\left(S_{c(j)m\|g}\right)$	0.603***	0.578***	0.678***
	(0.157)	(0.009)	(0.012)
Undisclosed × Post	0.196***	0.193***	0.257***
	(0.025)	(0.020)	(0.034)
Underlying Rating			
3 to 3.5-star	0.004	0.004	-0.125***
	(0.023)	(0.021)	(0.036)
4-star or more	-0.177***	-0.172***	-0.126***
	(0.044)	(0.041)	(0.042)
Disclosed Rating ×			
3 or 3.5-star	0.057***	0.058***	0.144***
	(0.021)	(0.020)	(0.035)
4-star or more	0.203***	0.198***	0.124***
	(0.041)	(0.039)	(0.041)
Observations	74,743		33,963

[a]Results based on fixed effects IV, linear fixed effects without IV, and Arellano-Bond estimator for dynamic panels. Robust standard errors in parentheses, clustered at the county level for the fixed effects analysis. In the fixed effects IV estimates, premium and within-group shares were instrumented with number of hospitals in the county, the hospital HHI in the county, and the minimum, maximum, and mean premium in the contract across all other counties in the state. Additional independent variables not in the table include county demographics (measures total population, age, race, income, education, and employment), contract age (indicator variables for contracts 4-8 years old or more than 8 years old), number of other counties in which the contract operates, number of other plans offered by the contract in the same county, and year fixed effects. * p<0.1. ** p<0.05. *** p<0.01.