Luke Pulaski

Prof. Aslim

ECON 4850 – Health Economics

4 December 2024

Testing for Adverse Selection in the Market for Medicare Advantage and Medicare Part D

Introduction

Ever since the introduction of Medicare Advantage (MA) and Medicare Part D (Part D) with the Medicare Modernization Act of 2003 (CMS, 2024), takeup of these supplemental health insurance coverage options for senior citizens in the U.S. has steadily increased up to the present day. Today, more than half of Medicare beneficiaries are enrolled in MA plans, or Medicare-subsidized plans sold by private companies that offers an alternative to traditional Medicare (Freed, et. al., 2024). Additionally, almost 80% of Medicare beneficiaries are enrolled in Part D, Medicare's outpatient prescription drug insurance program (Cubanski, 2024). While takeup rates this high are certainly worth celebrating, it is also worth wondering whether these Medicare supplements have attracted disproportionate numbers of beneficiaries with higher healthcare costs. Any insurer that offers plans with voluntary enrollment (such as MA and Part D) must be wary of this phenomenon, also known as adverse selection. While it has been shown to be a concern among private insurers, it is an even more pressing threat to markets for public health insurance (like Medicare), where the federal government, and ultimately, taxpayers, shoulder the burden of paying for care as opposed to private companies. Given the severity of the adverse selection if it exists public health insurance, in this study, I look for any indication of adverse selection in the markets for both MA and Part D. Using a five-year pooled subsample of individuals surveyed in the National Health Interview Survey (NHIS), I investigate if unhealthier, and therefore, riskier, individuals are more likely to enroll in MA or Part D as opposed to healthier individuals.

Literature Review

Several studies have documented evidence of adverse selection within the MA and Part D programs. For example, Lustig (2010) finds that, in the market for Medicare HMO plans (a subset of MA coverage), the more costly beneficiaries tend to have the highest preferences for insurance. Levy and Weir (2009), in an analysis of the predictors of Part D take-up among the elderly, note that most uninsured individuals in their sample made a rational economic choice in rejecting drug coverage; those who chose not to enroll in Part D likely elected not to because they already had low prescription drug costs. Using enrollment and claims data from the Medicare Current Beneficiary Survey, as well individual health risk scores based on the Center for Medicare and Medicaid Services Hierarchical Condition Category (CMS-HCC) risk adjustment model, Riley et. al. (2009) find that enrollees in both MA prescription drug plans and Part D fee-for-services plans tend to have higher risk scores than uninsured individuals or individuals who otherwise receive drug coverage from another provider. In a similar vein, Heiss et. al. (2009) provide evidence that while individuals with higher risk scores might not necessarily adversely select into Part D, they do tend to favor more generous plans within the program. Polyakova (2016) similarly finds that Part D beneficiaries with higher health risk scores tend to prefer plans where they are offered more comprehensive coverage.

The body of past literature indicates that adverse selection may be plaguing the markets for MA and Part D plans. However, an issue not addressed by some of the aforementioned studies is the endogeneity of the relationship between health risk and enrollment decision. While how costly an individual might be to insure has an effect on whether they enroll in MA or Part D, it is also true that their decision to enroll in either program can have a tangible effect on their health or healthcare utilization and thereby have an effect on how costly they are to insure. To combat this endogeneity problem, Chiappori and Salanie (2000), in experiments looking for adverse selection in any insurance market, recommend sampling only individuals who have never been insured before. This way, it is impossible for prior insurance status to have affected the "riskiness" of an individual. In the auto insurance market, for example, they suggest studying only "beginning drivers", who have no established pattern of driving behavior under any kind of insurance coverage. Following this logic, in the case of the market for MA and Part D, it would be wise to limit sampling to new enrollees, or individuals who are exactly of the age where they are first eligible to enroll.

Data

I utilize pooled cross-sectional data from the NHIS, an annual population health survey of the civilian noninstitutionalized population of the United States, over a period from 2019 to 2023. The NHIS redesigned its questionnaire ahead of 2019, so, to ensure the consistency of the content and questions of the survey over time, I choose 2019 as my first year of observation. I apply the NHIS-provided survey weights (which help correct for nonresponse), adjusting for the fact that my sample contains five years of data as opposed to one.

Following Chiappori and Salanie (2000), I test for the presence of adverse selection into MA and Part D using a sample of individuals who are of the age where they are first eligible to enroll in Medicare. Seniors are first eligible to enroll in Medicare three months before the month of their 65th birthday, and the initial enrollment period (IEP) ends three months after their birthday month. Since the NHIS does not include the birth dates of respondents, I limit my sample to individuals aged 65 and 66 in an attempt to ensure that all individuals have at least had the opportunity to enroll in both MA and Part D if they desired. Even with this approach, I still find it necessary to pool all available data-years to ensure that the resulting subsample is sufficiently as representative of all IEP-aged individuals in the U.S. as possible.

The NHIS inquires about a variety of health-related topics, but I am particularly interested in respondents' demographic characteristics, chronic health conditions, as well as their enrollment status in MA or Part D. Unfortunately, not all IEP-aged respondents indicated their enrollment status in either MA, Part D, or both, so those observations were dropped. Another handful of observations were dropped because information on one or more of a respondent's demographic characteristics was missing. In total, 5,480 individuals of IEP age remain in the sample to test for adverse selection into MA, with 5,532 in the sample for Part D.

To quantify an individual's health risk, I employ a similar strategy to that of Riley, et. al. (2009), which entails the usage of the CMS-HCC model for the MA sample and the related RxHCC model for the Part D sample. However, since I do not have access to individuals' risk score data as assigned by Medicare, I must calculate these scores on my own, for both enrollees and non-enrollees, using the information I have. The CMS-HCC model assigns a risk adjustment factor (RAF) to acute and chronic health conditions based on how costly an individual with that condition would be to insure relative to someone without it. RAFs are also assigned to demographic characteristics; for example, factors are added based on an individual's sex and age. For every health condition an individual has, the corresponding RAF is added to their risk score. Additionally, CMS uses an HCC model with different RAF scores, the RxHCC model, to determine how costly individuals with certain conditions will be to the Medicare Part D program, since some conditions rely more on prescription drugs to treat. So, I also calculate the health risk score of every individual in the Part D sample using the RxHCC model. The NHIS offers data on whether the respondents have previously been diagnosed with a number of medical conditions; I outline the conditions as well as my criteria to count a respondent as affected by that condition in Table 1. It should be mentioned that the HCC-based model does not add factors if an individual has hypertension or high cholesterol, whereas they are accounted for in the RxHCC model. For both the CMS-HCC and

CMS-RxHCC models, CMS normalizes risk scores so that 1 is the average score among the entire population covered under Medicare in a given year. This is done by simply dividing the risk scores of all Medicare enrollees by the mean risk score. To stay as faithful to the CMS-HCC model as possible, and to make interpretation of risk scores easier, I choose to normalize my risk scores in the same manner. The distributions of HCC- and RxHCC-based risk scores are displayed in Figure 1 and Figure 2, respectively. As expected, there is a pronounced right skew in the distribution of scores for either sample, meaning much of the IEP-aged sample is of below-average health risk. While it seems that individuals with lower health risk scores (especially those with scores between 0 and 0.5, for either sample) are less likely to be enrolled in MA or Part D, higher-risk individuals do not appear to be significantly more likely to be enrolled.

To control for observable factors that might have an effect on either health risk or enrollment decision, I gather demographic data on selected respondents from the NHIS. These factors are sex, race, whether a respondent lives in an urban or rural area, educational attainment, income (which I present as the ratio of a respondent's income to the federal poverty line, or "income/FPL ratio"), and "other care" (whether a respondent reported receiving regular medical care from the Veterans Administration (VA), Tricare, the Civilian Health and Medical Program of the Department of Veterans Affairs (CHAMPVA), or the Indian Health Service (IHS). Descriptive statistics for the MA and Part D samples, grouped by enrollees and non-enrollees, are shown in Table 2. To test if there are significant differences in the distributions of each control variable between enrollees and non-enrollees, I test whether the proportions of individuals in each subcategory of each variable are significantly different between enrollees and non-enrollees. In Table 2a, I display these percentage point differences. I additionally report the statistical significance of the difference between the two proportions for each subcategory. While it seems that most differences between the enrollee and non-enrollee groups, for both the MA and Part D samples, are negligible in terms of factors like sex, race, and educational attainment, the differences in terms of income and "other care" are much starker. For either sample, it seems that lower-income individuals are represented far more in the enrollee subset, whereas higher-income individuals seem less likely to be enrolled. Additionally, it seems that those receiving "other care" are far less likely to be enrolled in either MA or Part D, though these individuals constitute a relatively small portion of either sample. Lastly, it seems that individuals living in rural areas are less likely to enroll in MA, and, to a lesser degree, in Part D. For these reasons, I believe that my empirical model will predict that several mediating factors will have substantial effects on enrollment decision.

Model Specification

To model the effect that health risk has on enrollment decision for IEP-aged individuals, I employ two separate logistic models: one for the MA sample, and one for the Part D sample, both of which are presented below.

$$(1)\ logit(MA_i) = \beta\, Risk_i + \gamma\, X_i + \delta Z_i$$

$$(2)\ logit(D_i) = \beta\, Risk_i + \gamma\, X_i + \delta Z_i$$

Model 1 seeks to identify the effect of HCC-based risk score on the decision to enroll in Medicare Advantage. Model 2 seeks to identify the effect of RxHCC-based risk score on the decision to enroll in Medicare Part D. In both models, X represents a set of controls that include sex, race, educational attainment, income, and Z represents a set of year dummy variables. Subscript i indexes survey respondents.

Results

Results from Model 1 are displayed in Table 3, and results from Model 2 are displayed in Table 4. In both tables, Column 1 displays the coefficients of all variables from the logistic regression, while Column 2 displays the average marginal effect (AME) corresponding with each variable. I will focus on the results from Column 2 for ease of interpretation.

I find that a 1-unit increase in HCC-based health risk score, on average, increases the likelihood that an IEP-aged individual is enrolled in MA by 1.9%. To visualize this effect, assume there is an IEP-aged individual with a risk score of 1 (i.e. the HCC-based model predicts that they are of average health risk). My model predicts that, if that same individual's score were to suddenly rise to 2 (they become twice as risky, in this scenario), they are only 1.9% more likely to be enrolled in MA. Given that most sampled individuals' risk scores fall between 0 and 4, this effect, while somewhat significant, is not all that impactful. It is worth noting, however, that a number of the control variables in this model have an even larger, more significant effects on enrollment decision than health risk scores. For one, as I expected, living in a rural area decreases an individual's likelihood of enrollment, on average, by almost 7%. Other notable AMEs include the effects of having a bachelor's degree or higher (+3.3%), receiving "other care" (-23.9%), as well as being surveyed in 2019 (-9.7%) or 2020 (-6%). Finally, a one-unit increase in income/FPL ratio decreases the probability of enrollment by, on average, 2.6%.

The estimates for Model 2, on the other hand, show that RxHCC-based health risk score might have a greater effect on enrollment in Part D. I find that, on average, a 1-unit increase in health risk score increases the probability an individual is enrolled in Part D by 5.2%. While this effect is significant and larger than the effect of health risk score on enrollment in MA, I likewise cannot say that health risk has a noteworthy impact on Part D enrollment. As was the case with the analysis of the MA model results, a 5.2% increase in the likelihood of enrollment resulting from a 1-unit increase in risk score is simply not substantial when most RxHCC-based health risk scores in the sample fall between 0 and 3. That said, many of the same covariates that had noteworthy AMEs in Model 1 are also worth pointing out for their effect on enrollment in Part D. Having a bachelor's degree (+7%), receiving "other care" (-38.2%), and being surveyed in the year 2020 (-5.9%) all had relatively large, statistically significant AMEs, and while the AME of a one-unit increase in income/FPL ratio was relatively small (-2.2%), it, too, was significant.

Discussion

Despite the statistical significance of the impact of health risk score on enrollment in both MA and Part D, neither effect appears to be that meaningful in terms of magnitude. This could be the case for a number of reasons. First of all, it could simply be the case that adverse selection is not actually a large concern in the market for MA or Part D. As mentioned before, takeup rates of plans in both programs are already quite high, so it is very plausible that high- and low-risk beneficiaries are enrolling at similar rates. It is also very likely, however, that my results are marred by measurement error, mainly in the construction of my health risk score variables. My calculations of each individual's HCC- and RxHCC-based risk scores do not exactly reflect how they would be calculated by CMS, as NHIS does not survey individuals about every single medical condition that is fed into the HCC- and RxHCC-based models. Furthermore, my criteria to determine if an individual is currently affected by cancer and diabetes are likely flawed. If I had relaxed these criteria and allowed any individual who had been diagnosed with either condition at any point in their lives to be considered "affected", marked changes may have occurred in the distribution of health risk scores, which in turn may have modified the effect of health risk scores on enrollment status. Aside from measurement error, the NHIS data also did not allow me to determine if respondents had other health insurance coverage (either privately purchased or employer sponsored) that might be a substitute for MA or Part D outside of the "other care" categories I described. It could very well be the case that unhealthier individuals in either sample not enrolled in either program elected not do so because they had another source of coverage, muddying my results.

It seems as though several mitigating factors had far greater impacts on enrollment decisions (into both MA and Part D), some of which should be acted upon with policy interventions. For example, individuals with higher levels of educational attainment appear to be more likely to be enrolled in MA and Part D, which could be indicative of health literacy and financial literacy barriers preventing certain individuals from enrolling. There seem to be barriers to enrollment for individuals living in rural areas, which could require action to improve access to care outside urban areas through improvements in transportation and internet connection. Lastly, I find it interesting that the dummy variable for the year 2020, which coincides with the peak of the COVID-19 pandemic, was negative and significant in both models. This could mean that enrollment in MA and Part D was down in 2020 because new enrollees did not have access to resources to help them through the enrollment process, like in-person government offices, highlighting the need for further health insurance and technological literacy among seniors. While this analysis is not directly related to my research question, I nonetheless believe that the effects of these control variables merit discussion.

Conclusion

Using NHIS data to form a pooled subsample of IEP-aged individuals from 2019 to 2023, I find that, all else held equal, individuals with higher health risk are somewhat more likely to be enrolled in Medicare Advantage and Medicare Part D, although the effect of health risk on enrollment decision seems to be more than twice as strong in the market for Part D than for MA. It is worth mentioning that demographic and socioeconomic factors such as income, education, and receiving care from the VA, Tricare, CHAMPVA, or the IHS had sizeable effects, suggesting that

IEP-aged individuals' decisions to enroll in MA or Part D might be motivated by variables outside of their health status.

This study necessitates further research into the nature of adverse selection in the market for public health insurance. The next logical step would be to run a similar experiment with a dataset in which respondents are surveyed on a broader variety of health conditions and are asked more specific questions about the nature of their insurance coverage for hospital visits, ER visits, PCP visits, and prescription drugs. If such a dataset exists, it would be especially beneficial if it offered a larger sample size than the one presented in this study. Another avenue of research could be to investigate the degree to which adverse selection affects plan choice within MA and Part D over a similar time period, as not all MA and Part D plans are the same and instead offer varying levels of coverage at different costs. With either route, I hope that some of the flaws of this study can potentially be mitigated.

References

Center for Medicare and Medicaid Services (CMS). (2024) *History*. https://www.cms.gov/about-cms/who-we-are/history

Chiappori, P.A., & Salanie, B. (2000). Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy*, 108(1), 56-78.

Cubanski, J. (2024). A Current Snapshot of the Medicare Part D Prescription Drug Benefit. Kaiser Family Foundation. https://www.kff.org/medicare/issue-brief/a-current-snapshot-of-the-medicare-part-d-prescription-drug-benefit/

Freed, M., Biniek, J.F., Damico, A., & Neuman, T. (2024). *Medicare Advantage in 2024: Enrollment Updates and Key Trends*. Kaiser Family Foundation. https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2024-enrollment-update-and-key-trends/

Heiss, F., McFadden, D., & Winter, J. (2009). *The workings of a private health insurance market: enrollment decisions, plan choice, and adverse selection in Medicare Part D*. University of Munich working paper.

Levy, H., & Weir, D.R. (2009). Take-up of Medicare Part D: Results from the Health and Retirement Study. *The Journals of Gerontology: Social Sciences*, 65B(4), 492-501.

Lustig, J. (2010). "The Welfare Effects of Adverse Selection in Privatized Medicare." Yale University working paper.

Polyakova, M. (2016). Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D. *American Economic Journal: Applied Economics*, 8(3), 165-195.

Riley, G.F., Levy, J.M., & Montgomery, M.A. (2009). Adverse Selection in the Medicare Prescription Drug Program. *Health Affairs*, 28(6), 1826-1837.

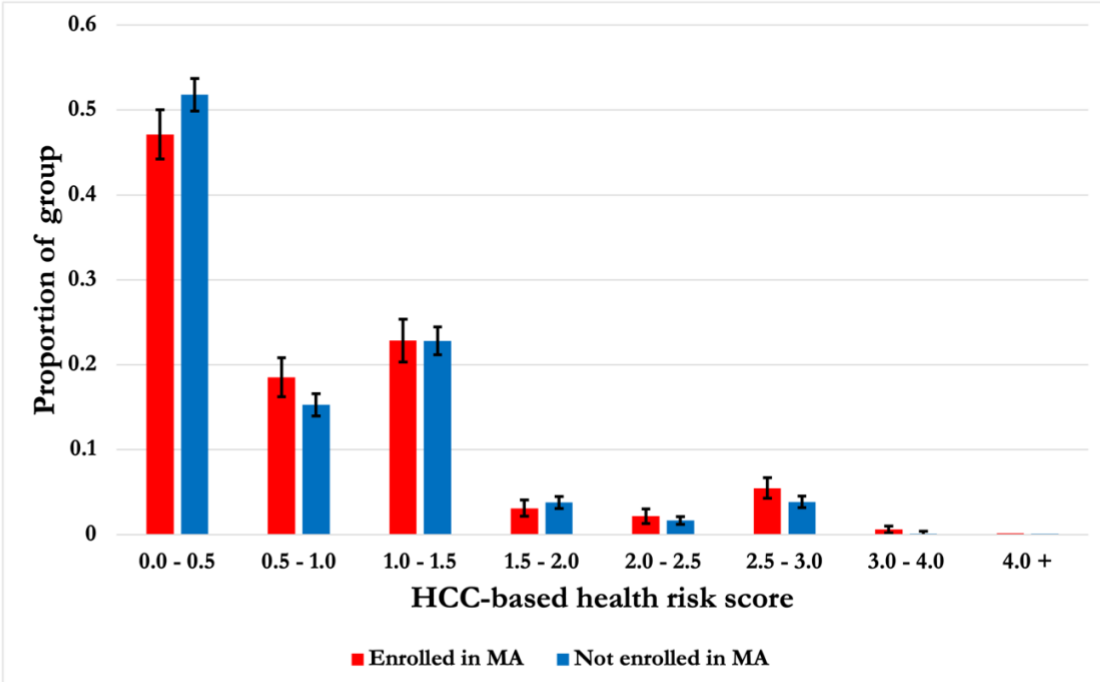
Appendix

Table 1: *Criteria to be counted as affected by a given condition*

Condition	Time Frame of Diagnosis
Hypertension	Within 12 months prior to the time of the survey
High cholesterol	Within 12 months prior to the time of the survey
Coronary artery disease (CAD)	Ever
Chronic obstructive pulmonary disorder (COPD)	Ever
Asthma	Within 12 months prior to the time of the survey
Diabetes[1]	Within 5 years prior to the time of the survey
Arthritis	Ever
Cancer[1]	Within 5 years prior to the time of the survey

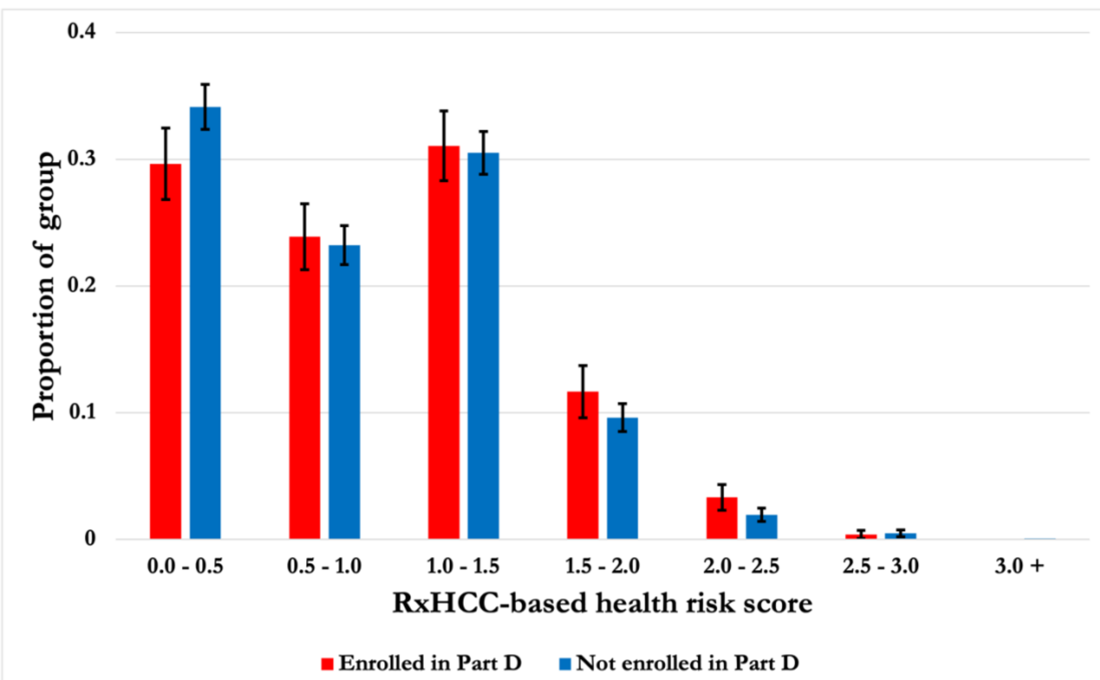
[1] The criteria to be considered "affected" by cancer and diabetes was determined personally; NHIS asks respondents if they have ever been diagnosed with cancer or diabetes, yet there is no way to determine if all these individuals still have either condition, given that they aren't necessarily permanent and incurable, unlike CAD, COPD, or arthritis. Therefore, I created these thresholds to ensure that individuals considered to be "affected" by cancer or diabetes to currently have or recently have had the condition.

Figure 1: *Distribution of HCC-based risk scores among MA enrollee and non-enrollee groups*



95% confidence intervals shown

Figure 2: *Distribution of RxHCC-based risk scores among Part D enrollee and non-enrollee groups*



95% confidence intervals shown

Table 2: *Descriptive statistics (all covariates)*

	MA Sample (N = 5,480)		Part D Sample (N = 5,532)	
	Enrolled (N = 1,592)	Not enrolled (N = 3,888)	Enrolled (N = 2,216)	Not enrolled (N = 3,316)
Sex				
Male	48.4	48.0	46.4	48.7
Female	51.6	52.0	53.6	51.3
Race				
White	72.3	74.1	75.0	72.7
Black	12.4	9.3	10.8	9.7
Hispanic	9.5	10.5	9.2	10.8
Asian	4.3	4.7	3.5	5.2
AI/AN[1]	0.5	0.4	0.5	0.5
AI/AN & other	0.4	0.6	0.7	0.5
Other race or multiple races	0.6	0.5	0.3	0.7
Area of residence				
Urban	85.5	82.4	82.6	83.8
Rural	14.5	17.6	17.4	16.2
Education				
Bachelor's degree or higher	28.9	31.0	30.9	29.8
Less than bachelor's degree	71.1	69.0	69.1	70.2
Income/FPL ratio				
Less than 1	10.2	7.7	11.6	06.6
1 – 2.49	32.5	21.7	28.4	22.8
2.5 – 4.99	30.7	33.8	30.2	34.4
5 – 7.49	15.4	17.2	15.2	17.5
7.5 – 9.99	6.5	10.2	7.7	10.1
10 or greater	4.8	9.5	7.0	8.7
Other care				
Yes	2.3	6.7	1.7	7.6
No	97.7	93.3	98.3	92.4

[1] American Indian or Alaska Native

Table 2a: *Percentage-point differences in proportions of enrollees and non-enrollees in each subcategory*

	MA Sample	**Part D Sample**
Sex		
Male	0.4	-2.3
Female	-0.4	2.3
Race		
White	1.8	2.3*
Black	-3.1***	1.1
Hispanic	-1.0	-1.6*
Asian	-0.4	-1.7***
AI/AN	0.1	0.0
AI/AN & other	-0.2	0.2
Other race or multiple races	0.1	-0.4**
Area of residence		
Urban	3.1***	-1.2
Rural	-3.1***	1.2
Education		
Bachelor's degree or higher	-2.1	1.1
Less than bachelor's degree	2.1	-1.1
Income/FPL ratio		
Less than 1	2.5***	5.0***
1 – 2.49	10.8***	5.6***
2.5 – 4.99	-3.1**	-4.2***
5 – 7.49	-1.8*	-2.3**
7.5 – 9.99	-3.7***	-2.4***
10 or greater	-4.7***	-1.7**
Other care		
Yes	-4.4***	-5.9***
No	4.4***	5.9***

Significance of two-proportion Z-test shown
**** p<0.01, ** p<0.05, * p<0.1*

Table 3: *Logistic model results, MA sample*

Dependent variable: *Enrollment status in MA*

	(1) Logit model results	(2) AME
HCC-based risk score	0.09374* (0.05236)	0.01852* (0.01031)
Female	0.07712 (0.07445)	-0.01523 (0.0147)
White	-0.07712 (0.07445)	-0.04602 (0.10125)
Black	-0.23294 (0.51244)	-0.02064 (0.10481)
Hispanic	-0.10446 (0.53052)	-0.10837 (0.10599)
Asian	-0.4244 (0.55104)	-0.08383 (0.10888)
AI/AN	0.31664 (0.74742)	0.06255 (0.14759)
AI/AN & other	-0.48441 (0.68999)	-0.0957 (0.13638)
Rural	-0.35346*** (0.1065)	-0.06982*** (0.02098)
Bachelor's degree or higher	0.16761** (0.07905)	0.03311** (0.01558)
Income/FPL ratio	-0.13064*** (0.01445)	-0.02581*** (0.00274)
Other care	-1.21004*** (0.19788)	-0.23903*** (0.03896)
Year 2019	-0.4934*** (0.10881)	-0.09747*** (0.02135)
Year 2020	-0.30328*** (0.11403)	-0.05991*** (0.02247)
Year 2021	-0.16162 (0.10812)	-0.03193 (0.02133)
Year 2022	-0.11379 (0.11026)	-0.02248 (0.02177)
Constant	0.12247 (0.51558)	
Observations	5,480	5,480

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

Table 4: *Logistic model results, Part D sample*

Dependent variable: *Enrollment status in Part D*

	(1) Logit model results	(2) AME
RxHCC-based risk score	0.22937*** (0.06268)	0.05223*** (0.01421)
Female	0.008215 (0.06299)	0.00187 (0.01435)
White	0.82784 (0.52125)	0.18852 (0.11861)
Black	0.79327 (0.52925)	0.18065 (0.12044)
Hispanic	0.50009 (0.5412)	0.11388 (0.12324)
Asian	0.31275 (0.55455)	0.07122 (0.12629)
AI/AN	1.35743* (0.7486)	0.30912* (0.17046)
AI/AN & other	1.59536** (0.64529)	0.3633** (0.1466)
Rural	-0.029728 (0.0879)	-0.00677 (0.02002)
Bachelor's degree or higher	0.30619*** (0.07936)	0.06973*** (0.01793)
Income/FPL ratio	-0.09592*** (0.0134)	-0.02184*** (0.003)
Other care	-1.67539*** (0.2112)	-0.38153*** (0.04688)
Year 2019	-0.14173 (0.09804)	-0.03227 (0.02231)
Year 2020	-0.25895** (0.10836)	-0.05897** (0.02461)
Year 2021	-0.08187 (0.10097)	-0.01864 (0.02299)
Year 2022	0.01319 (0.10748)	0.003 (0.02448)
Constant	-0.9593** (0.5271)	
Observations	5,532	5,532

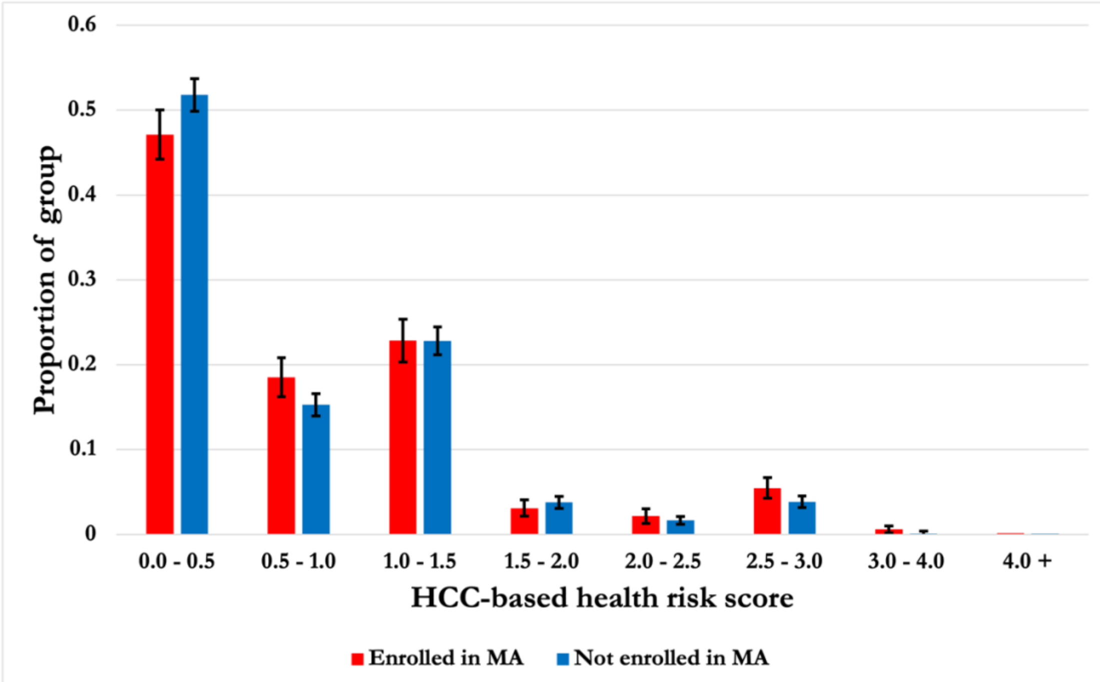
Standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table 1: *Criteria to be counted as affected by a given condition*

Condition	**Time Frame of Diagnosis**
Hypertension	Within 12 months prior to the time of the survey
High cholesterol	Within 12 months prior to the time of the survey
Coronary artery disease (CAD)	Ever
Chronic obstructive pulmonary disorder (COPD)	Ever
Asthma	Within 12 months prior to the time of the survey
Diabetes[1]	Within 5 years prior to the time of the survey
Arthritis	Ever
Cancer[1]	Within 5 years prior to the time of the survey

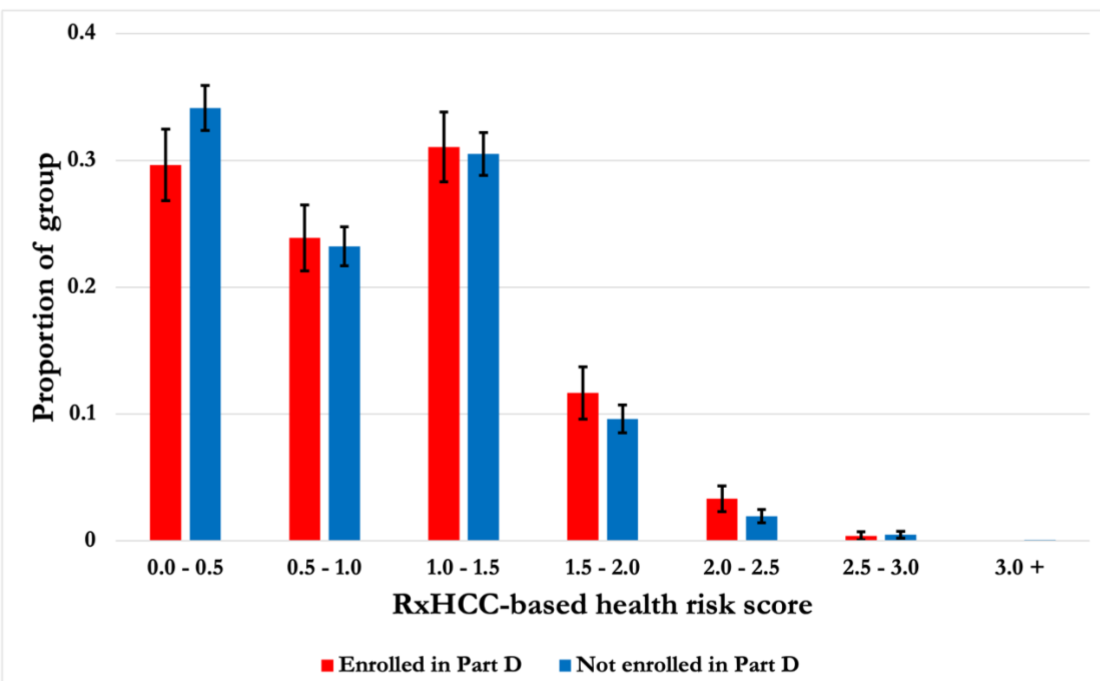
[1] The criteria to be considered "affected" by cancer and diabetes was determined personally; NHIS asks respondents if they have ever been diagnosed with cancer or diabetes, yet there is no way to determine if all these individuals still have either condition, given that they aren't necessarily permanent and incurable, unlike CAD, COPD, or arthritis. Therefore, I created these thresholds to ensure that individuals considered to be "affected" by cancer or diabetes to currently have or recently have had the condition.

Figure 1: *Distribution of HCC-based risk scores among MA enrollee and non-enrollee groups*



95% confidence intervals shown

Figure 2: *Distribution of RxHCC-based risk scores among Part D enrollee and non-enrollee groups*



95% confidence intervals shown

Table 2: *Descriptive statistics (all covariates)*

	MA Sample (N = 5,480)		Part D Sample (N = 5,532)	
	Enrolled (N = 1,592)	Not enrolled (N = 3,888)	Enrolled (N = 2,216)	Not enrolled (N = 3,316)
Sex				
Male	48.4	48.0	46.4	48.7
Female	51.6	52.0	53.6	51.3
Race				
White	72.3	74.1	75.0	72.7
Black	12.4	9.3	10.8	9.7
Hispanic	9.5	10.5	9.2	10.8
Asian	4.3	4.7	3.5	5.2
AI/AN[1]	0.5	0.4	0.5	0.5
AI/AN & other	0.4	0.6	0.7	0.5
Other race or multiple races	0.6	0.5	0.3	0.7
Area of residence				
Urban	85.5	82.4	82.6	83.8
Rural	14.5	17.6	17.4	16.2
Education				
Bachelor's degree or higher	28.9	31.0	30.9	29.8
Less than bachelor's degree	71.1	69.0	69.1	70.2
Income/FPL ratio				
Less than 1	10.2	7.7	11.6	06.6
1 – 2.49	32.5	21.7	28.4	22.8
2.5 – 4.99	30.7	33.8	30.2	34.4
5 – 7.49	15.4	17.2	15.2	17.5
7.5 – 9.99	6.5	10.2	7.7	10.1
10 or greater	4.8	9.5	7.0	8.7
Other care				
Yes	2.3	6.7	1.7	7.6
No	97.7	93.3	98.3	92.4

[1] American Indian or Alaska Native

Table 2a: *Percentage-point differences in proportions of enrollees and non-enrollees in each subcategory*

	MA Sample	Part D Sample
Sex		
Male	0.4	-2.3
Female	-0.4	2.3
Race		
White	1.8	2.3*
Black	-3.1***	1.1
Hispanic	-1.0	-1.6*
Asian	-0.4	-1.7***
AI/AN	0.1	0.0
AI/AN & other	-0.2	0.2
Other race or multiple races	0.1	-0.4**
Area of residence		
Urban	3.1***	-1.2
Rural	-3.1***	1.2
Education		
Bachelor's degree or higher	-2.1	1.1
Less than bachelor's degree	2.1	-1.1
Income/FPL ratio		
Less than 1	2.5***	5.0***
1 – 2.49	10.8***	5.6***
2.5 – 4.99	-3.1**	-4.2***
5 – 7.49	-1.8*	-2.3**
7.5 – 9.99	-3.7***	-2.4***
10 or greater	-4.7***	-1.7**
Other care		
Yes	-4.4***	-5.9***
No	4.4***	5.9***

Significance of two-proportion Z-test shown
**** $p<0.01$, ** $p<0.05$, * $p<0.1$*

Table 3: *Logistic model results, MA sample*

Dependent variable: *Enrollment status in MA*

	(1) Logit model results	(2) AME
HCC-based risk score	0.09374* (0.05236)	0.01852* (0.01031)
Female	0.07712 (0.07445)	-0.01523 (0.0147)
White	-0.07712 (0.07445)	-0.04602 (0.10125)
Black	-0.23294 (0.51244)	-0.02064 (0.10481)
Hispanic	-0.10446 (0.53052)	-0.10837 (0.10599)
Asian	-0.4244 (0.55104)	-0.08383 (0.10888)
AI/AN	0.31664 (0.74742)	0.06255 (0.14759)
AI/AN & other	-0.48441 (0.68999)	-0.0957 (0.13638)
Rural	-0.35346*** (0.1065)	-0.06982*** (0.02098)
Bachelor's degree or higher	0.16761** (0.07905)	0.03311** (0.01558)
Income/FPL ratio	-0.13064*** (0.01445)	-0.02581*** (0.00274)
Other care	-1.21004*** (0.19788)	-0.23903*** (0.03896)
Year 2019	-0.4934*** (0.10881)	-0.09747*** (0.02135)
Year 2020	-0.30328*** (0.11403)	-0.05991*** (0.02247)
Year 2021	-0.16162 (0.10812)	-0.03193 (0.02133)
Year 2022	-0.11379 (0.11026)	-0.02248 (0.02177)
Constant	0.12247 (0.51558)	
Observations	5,480	5,480

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

Table 4: *Logistic model results, Part D sample*

Dependent variable: *Enrollment status in Part D*

	(1) Logit model results	(2) AME
RxHCC-based risk score	0.22937*** (0.06268)	0.05223*** (0.01421)
Female	0.008215 (0.06299)	0.00187 (0.01435)
White	0.82784 (0.52125)	0.18852 (0.11861)
Black	0.79327 (0.52925)	0.18065 (0.12044)
Hispanic	0.50009 (0.5412)	0.11388 (0.12324)
Asian	0.31275 (0.55455)	0.07122 (0.12629)
AI/AN	1.35743* (0.7486)	0.30912* (0.17046)
AI/AN & other	1.59536** (0.64529)	0.3633** (0.1466)
Rural	-0.029728 (0.0879)	-0.00677 (0.02002)
Bachelor's degree or higher	0.30619*** (0.07936)	0.06973*** (0.01793)
Income/FPL ratio	-0.09592*** (0.0134)	-0.02184*** (0.003)
Other care	-1.67539*** (0.2112)	-0.38153*** (0.04688)
Year 2019	-0.14173 (0.09804)	-0.03227 (0.02231)
Year 2020	-0.25895** (0.10836)	-0.05897** (0.02461)
Year 2021	-0.08187 (0.10097)	-0.01864 (0.02299)
Year 2022	0.01319 (0.10748)	0.003 (0.02448)
Constant	-0.9593** (0.5271)	
Observations	5,532	5,532

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