STAT 521 Final Project

Health Insurance Plans and Individuals' Prescription drug Expenditures: How Expenditures vary among Plans along with underlying price inequality

Yan He

April 30th, 2021

Abstract

This study examines how insurance plans with different underlying prescription drug price affect individuals' prescription drug expenditures—total payments made by all sources, including out-of-pocket and insurance payments. Analysis of a national medical expenditure survey reveals that insured people with Medicare/Part D[1] and Private Health Insurance plans have significantly more Rx expenditure than individuals with Medicaid coverage only, largely driven by underlying price inequality.

When controlling for prescription drug utilization and other factors, on average, individuals enrolled in Medicare with or without Part D coverage spend around 48% more than Medicaid (only) enrollees in 2018, individuals with Private Health Insurance plans spent 35% more. The spending premium (%) for Medicare/Part D was smaller in 2017, around 30%. When further controlling for prescription drug type (in terms of Therapeutic class), on average, individuals with Medicare/Part D and PHI spent around 33% and 30% more than individuals with only Medicaid, respectively (in both 2017 and 2018). Since the out-of-pocket share is much lower for Medicaid than for other insurance plans, the underlying price inequality further increases the out-of-pocket burden for people with Medicare/Part D and PHI, especially for the elderly enrolled in Medicare and purchase prescription drug coverage themselves. We did not see any difference in prescription drug expenditures caused by the potential drug price difference between Medicaid only and Uninsured groups.

Overall, the findings suggest that the presumably higher drug price significantly increased the expenditures on prescription drugs for people with Medicare/Part D and Private Health Insurance plans. The effect is bigger for the elderly covered by Medicare with or without Part D.

[1] Medicare/Part D includes people who are on Medicare with/without Part D coverage

Contents

1. INTRODUCTION

1.1 Background

Prescription drug prices vary under different health insurance plans. The main insurance coverages in the U.S. include Medicaid, Medicare and Private Health Insurance (PHI), there are also some other health plans such as CHAMPUS (Civilian Health and Medical Programs of the Uniformed Services), TRICARE that provides comprehensive coverage for active-duty US military personnel, military retirees, their dependents and survivors. For Medicare, there is a Medicare Part D (starts 2006) that specifically covers prescription drugs. Medicare prescription drug coverage (Part D) is available in two ways. For those who have original/traditional Medicare (Part A and Part B), Medicare Part D is not automatically included, they can enroll in a stand-alone Medicare Prescription Drug Plan that works alongside the original Medicare benefits. For those who have Medicare Advantage Plan (Part C) or other Medicare health plan with drug coverage, they get all of their Part A, Part B and drug coverage (Part D). In both ways, Medicare Part D coverage and Medicare Advantage plan (or other Medicare health plan) are provided by private insurance companies that are contracted by Medicare, so costs and availability can differ between Medicare plans. Furthermore, Medicare Part B (Medical Insurance) covers a limited number of outpatient prescription drugs under limited conditions. Part B covers drugs that people would not usually give to themselves, like the drugs people get at a doctor’s office or hospital outpatient setting, certain oral cancer drugs, and drug used with some types of durable medical equipment.

Usually, Medicaid can negotiate the drug price while Medicare cannot[2]. Medicaid always pays the lowest price compared to Medicare and private insurance plans. From a consumer perspective, private insurance beneficiaries or uninsured people can use manufacturer coupons to lower their out-of-pocket spending while people with federally funded insurance (including Medicaid and Medicare) cannot. In STAT 520 final project, preliminary analyses of prescription drug expenditures were conducted to examine the underlying price difference among different types of insurance. The analyses showed that on average, compared to Medicaid (only) enrollees, Medicare (only) enrollees spent around 40% more in 2018 when controlling other covariates (including prescription utilization), and individuals with private insurance spent around 35% more. OLS analysis of multiple years’ data also indicates significant increase in Rx expenditures under Medicare and private insurance compared to under Medicaid, presumably caused by higher price. There were some limitations in the previous analyses and improvements can be made with more data.

1.2 Study objectives

This project will introduce more data to try additional models and/or improve the previous analyses. The main objective is still to study the impact of different health insurance coverage on the individual’s prescription drug spending, due to underlying price difference. But this project will focus on adults and some analyses will focus on certain types of drugs (in terms of therapeutic class). By controlling the therapeutic class of the drugs, we can study how the drug type/class can change the estimates of the health plan partial effects. It will also examine other factors that can influence the prescription drug spending, such as basic demographics, region, health conditions, prescription drug count. The hypothesis is that, compared to Medicaid, Medicare and private health

[2] For Medicare part D, the price is negotiated by pharmacy benefit manager (PBM) for the private health plans without government involvement.

insurance lead to higher prescription drug prices, thus higher expenditures. We will compare the results from the new models to the results of previous model.

2. METHOD

2.1 Data

Other than the 2017 and 2018 full year data from Medical Expenditure Panel Survey (MEPS, https://meps.ahrq.gov/mepsweb/), two more data sources are used in this study—MEPS Prescribed Medicines event-level data of each year and the 2017-2018 two-year longitudinal data. All MEPS data is produced by Agency for Healthcare Research and Quality (AHRQ), they provide data on health status, medical conditions, healthcare utilization, and healthcare expenditures for the U.S. civilian, non-institutionalized population. It is the most complete source of data on the cost and use of health care and health insurance coverage.

In the two-year longitudinal data, each person represents those who participated in the MEPS survey for two continuous years, here 2017-2018. Both MEPS full year data and longitudinal data have the same variables, including the demographic, socioeconomic information (such as employment status) of the individuals, data regarding the utilization of different health care services (here we focus on prescription drugs) as well as corresponding expenditures and source of payment—Medicaid, Medicare, private health insurance and out-of-pocket. In terms of insurance coverage, individuals were also asked whether they specifically have Medicare Advantage Plan (or Medicare managed care plan) coverage and Medicare Part D prescription drug coverage. For adults aged 18 and older as of certain interview date in MEPS, a variety of health status and health care quality and preventive health care measures are available in the individual-level data. This self-reported health information is collected from the Self-Administered Questionnaire (SAQ) that was administered only to adults.

In the prescription drug event-level data, each record represents one household-reported prescribed medicine with detailed information mainly collected from the pharmacies with permission, including drug NDC, therapeutic class, dosage, types of pharmacies and payment for each filled prescription. Specifically, in the main household component survey, the respondents were asked about the name of any prescribed medication they or their family members purchased or otherwise obtained during that round, and whether they send in claim forms for their prescriptions (self-filers) or if their pharmacy providers do this automatically for them at the point of purchase (non-self-filers). For non-self-filers, prescription drug data (including charge and payment) was collected based on the purchase records obtained in the pharmacy component survey with the permission from the respondents (unless the purchase was an insulin or diabetic supply/equipment event). Data for self-filers was collected directly from the household questionnaire (self-reported), as prescription drug data for self-filers' purchases would not be available from the pharmacy component. Uninsured persons were treated as non-self-filers. Persons who said they did not know if they sent in their own prescription claim forms were considered as self-filers. Also, the prescription drugs in MEPS only refer to outpatient prescription drugs, excluding over-the-counter medications and drugs that people may purchase from overseas.

Research shows that the number of drug fills and total expenditures for people with Medicare Part D coverage are reasonably accurate compared with claims data. It seems that household respondents tended to underreport the number of different drugs taken, but tended to overreport the number of fills of each drug. Behavioral analyses of the determinants of medication use and expenditures were largely unaffected because underreporting cut across most sociodemographic groups[3]. Therefore, the accuracy of prescription drug data in MEPS should be less concerning.

The prescribed medicines data can be merged with either annual or longitudinal data for analysis. To improve the analysis by controlling more information of the drugs in the model, we also extracted a cohort whose prescription drugs over one year all belong to the same therapeutic class. Eventually, we have three data used for final analyses in this study—the full year data of 2017 and 2018 that was used in the previous analyses and two new/main data. The first main data is the 2-year longitudinal data in which the individual adults can prescribe any prescription drugs and they appear in both 2017 and 2018 survey. The second is the cohort data where we restrict the sample to those who prescribe single (therapeutic) class of drugs within each year.

2.2 Measurements

The outcome variable in this study is still the annual total expenditures on prescription drugs—The prescription drug expenditures in MEPS are measured as the sum of payments from different sources, which incorporates discounts for point-of-purchase cost, but not manufacturer rebates[4].

The main variable of interest is the 6-level insurance plan type, which is derived according to the self-reported insurance coverages. We improved the categorization of this variable. Instead of grouping every individual into Medicaid Only, Medicare Only, Private Insurance Only, Medicaid+ (covered by more than 2 insurance plans, including Medicaid), Other plans, or Uninsured, in this study we replace "Medicare only" as "Medicare/Part D", and added another category of "Medicare+". Medicare/part D group include:

a) people who report being covered by Medicare only.
b) people who are covered by Medicare and one or more other insurances, have Part D coverage in the meantime, either not working or working but not covered by any group insurance (employer/union/other group).

Medicare+ group include people who have Medicare as well as one or more other insurances but not in Medicare/part D group. After this procedure, all people from the previous "Other" group are split into Medicare/Part D and Medicare+ groups, some of the people in the original Medicaid+ are grouped into those two groups as well. Eventually, every individual is categorized as having Medicaid only, Medicare/Part D, Medicare+, Private Health Insurance (PHI) only, Medicaid+ (covered by more than 2 insurance plans, including Medicaid but not Medicare), or Uninsured. Examples in ***Appendix 2*** further illustrate the definitions of different insurance types.

The rationale behind this reclassification of insurance type comes from multiple perspectives. On one hand, Medicare prescription drug coverage is part of Medicare (Part D). As mentioned in the Background session, both Medicare Part D and Medicare Advantage Plans are offered by Medicare-approved private companies that follow rules set by Medicare, most Medicare Advantage plans include prescription drug coverage (Part D), not necessary. Whichever way, individuals must have Medicare Part A and/or Medicare Part B (original Medicare) to join a separate Medicare drug plan or a Medicare Advantage plan. Therefore, it is possible for people who purchase Medicare Part D coverage separately from the original Medicare or through Medicare Average plan may report themselves covered by private insurance, because both Part D and Medicare Advantage plans can only be purchased from Private insurance companies (contract with Medicare). And even the person reports having Medicare coverage only (no coverage from other insurance such as PHI), he/she may just do not know the fact that prescription drug coverage

[3] https://pubmed.ncbi.nlm.nih.gov/22235548/
[4] Manufacture rebates are a separate transaction paid directly from manufactures to health plans.

(Part D) is a type of private health insurance so that he/she does not report PHI coverage (no matter whether he/she reported Part D coverage). In fact, we can assume people with Medicare only actually have Part D coverage but did not report due to lack of knowledge, MEPS data also supports this assumption—around 95% of people in Medicare/Part D category reported having Part D prescription drug coverage in both the pooled full year data of 2017 and 2018 and the 2-year longitudinal data.

On the other hand, it may be possible for an individual to have more than one insurance help pay for prescription drugs. In this case Medicare part D Drug Plans coordinate benefits with other drug coverage, where the primary insurance pays first and then the secondary insurance pays the remaining unpaid amount up to the plan's limit. Although usually, Medicare Part D pays first, for example, when the individuals are also covered by employer's or union's Group Health Plan but retired (not working), or the individuals are covered by TRICARE for life, or they are covered by State Pharmaceutical Assistance Program (SPAP). Medicare Part D is rarely the secondary payer, mainly when the individuals are still working and participate in the employer/group insurance from the employment. When the individuals have both Medicare and Medicaid, usually the drugs are covered by Part D, in rare cases, Medicaid pays for drugs that are NOT covered by Medicare Part D such as drugs for fertility, weight loss/gain. Therefore, it is reasonable to consider people who have Part D coverage are paying their prescription drugs with Part D plans most of the time. Because of this fact that Medicare Part D pays first in most of the cases, the Medicare+ category should have very small number of people, mainly including Medicare enrollees who either have prescription drug coverage from other insurance rather than Part D or have both Part D and the employer/union/other group insurance from working job. But the latter should be very few people because in that case the individual would pay double premiums for both Medicare Part D and group insurance. Insurance type distribution in ***Table 1*** and ***2*** also reflects this—only 4.3% people in the 2-year longitudinal data and 1.4% people in the cohort population (who use unique class of drugs over a year) are in Medicare+ group.

Another key variable is the count/fills of all prescribed medications purchased during the year, which directly affects the total expenditure on prescription drugs. Since we focus on the adults who are eligible for the self-administered questionnaire, measurements related to the individuals' health are also included in the analysis, such as whether the individual reported any limitation (yes or no, any IADL, ADL, functional or activity limitations), whether the individual ever been diagnosed with diabetes (yes or no/not mentioned). Other demographic and socioeconomic characteristics information we considered include education level (below Some College, Some College or above, Other degree or Unknown), race (White, Black or African American, Other), ethnicity (Hispanic or not), immigration status (native born or not), family size, poverty category (<200%, >=200% poverty line), region (Northeast, Midwest, South, West).

2.2 Empirical Framework

Our main strategy is to uses OLS, Pooled OLS and Random Effects models to examine the effects of different health insurance plans on individuals' prescription drug expenditures. The underlying assumption is that the prescription drug retail prices are higher for some insurance plans than others, which directly increase the total expenditures for people who are covered by insurance plans that pay higher drug prices. The OLS model is:

$$Y = \beta_0 + \beta_1 Instype + \sum_{i=2}^{K} \beta_i X_i + u \qquad (1)$$

When we use 2-year data, the Xs include the year dummy used to check the year effect, and the interaction of insurance plan and year to capture the potential difference in the health plan effect over time. When we use the data of cohort who only prescribed one type of therapeutic class drugs over one year, the Xs include the categorical variable of Therapeutic Class to control the effects of different prescription drug classes.

When working with the 2-year longitudinal data, the specified model is:

$$Y_{it} = \beta_0 + \beta_1 Instype_{it} + \beta_2 Year_{it} + \beta_3 Year_{it} * Instype_{it} + \sum_{i=4}^{K} \beta_i X_{it} + c_i + u_{it} \qquad (2)$$

The models fitted with 2-year longitudinal data are estimated with both POLS and Random Effects methods. Since insurance type may not change a lot over 2-year time periods, (individual) Fixed Effects model using the 2-year longitudinal data may not be appropriate, but we still try it and the results are presented in ***Appendix 3***.

In multi-year analysis model where we control for the year, the reference year is 2018, and the spending premium for non-Medicaid-only health plans in 2017 will be calculated by $\exp(\beta_1 + \beta_3) - 1$, and the corresponding standard error of the estimates is $\exp(\beta_1 + \beta_3) se(\beta_1 + \beta_3)$ (delta method), the p-value (about significance of hypothesis testing N_0: $\beta_1 + \beta_3 = 0$) of the spending premium in 2017 is then derived based on the t-statistic calculated by $(\beta_1 + \beta_3)/se(\beta_1 + \beta_3)$ and the degree of freedom of the model.

We restricted our sample to the population who have positive counts of prescription medications during the year. In the analyses, we used both log expenditures and log Rx count in the models to ameliorate the effect of the extreme values. We controlled for health information in all models. All the expenditures are indexed to constant dollar of 2018. Robust standard errors are calculated in all regressions to take the potential heteroscedasticity into account.

Before fitting the models discussed above, two sample T-test (for two-level category variable) and ANOVA (for category variables with three or more levels) were introduced to test the difference of log prescription drug expenditures between/among different groups. Scatter plot was employed to check potential relationship between log prescription drug expenditures and the log Rx count.

3. RESULTS

3.1 Descriptive results

Here in this study, we focused on adults who were eligible for MEPS self-administrated Questionnaire (SAQ, with health-related data) and those who prescribed at least 1 medication in the year. For the full year data, there are 14,714 and 14,881 such observations in 2017 and 2018, respectively. Among all the observations in the full year data, 6,123 sample respondents appear in both 2017 and 2018. Also, among the full year sample, 3,139 (2017) and 2,970 (2018) individuals prescribe only one class (therapeutic) of drugs during the year.

Table 1 and ***Table 2*** present with different data the mean and standard deviation of the prescription drug expenditures (log transformed) by different categorical variables. The p-values in the tables are from the T-test or ANOVA analysis. From the p-values, we can see that for each category variables, the mean of the log Rx expenditure is significantly different within different groups. For those who prescribe any kinds of prescription drugs and participated in the survey for two years, ***Table 1*** shows that the mean log Rx expenditure for Medicaid (only) enrollees is around 6, the mean log Rx expenditure for people with Medicare/Part D and Medicare+ is around 7, much higher than that of Medicaid (only) enrollees. People with other insurance types

(PHI, Medicaid+, Uninsured) have smaller mean of their log Rx expenditures than Medicaid (only) enrollees in both years. In general, the Rx expenditures seems slightly higher in 2018 than 2017 under all types of insurance plans, except for people with Medicaid+. For those who prescribe unique therapeutic class of drugs in one year, ***Table 2*** shows that the mean log Rx expenditure for Medicaid (only) enrollees is round 3.75 in 2017 and 2018, and the mean log Rx expenditure for other insured people is more or less, higher than that for Medicaid (only) enrollees. Uninsured people have similar mean log Rx expenditure, compared to Medicaid (only) enrollees. There seems to be no significant difference in the mean Rx expenditures between 2017 and 2018 for each type of insurance.

Figure 1 displays the scatter plots of log Rx expenditure and log Rx counts in each year, within different data. It indicates a positive correlation between Rx expenditure and the count of the prescription drugs.

In ***Table 3*** and ***Table 4***, we look at the mean (SD) of Rx count/fills and individual's out-of-pocket (OOP) share by health plans in 2017 and 2018, to better understand the data. With the 2-year longitudinal data, ***Table 3*** shows that the adult Medicaid (only) enrollees and people with Medicare and/or Part D in the survey have on average about 30 prescription drug fills in 2017 and 2018, with people under Medicare/Part D prescribing slightly more. People under Medicare+ have on average about 25 Rx fills, people under PHI have around 15 prescription fills on average. People under Medicaid+ and those who are uninsured have on average about 21 and 14 prescription fills, respectively. Overall, on average, people with different insurance plans all have about 3 more prescription fills in 2018 than in 2017. For the out-of-pocket (OOP) share of the Rx spending of the adults in both 2017 and 2018, people with Medicaid (only) have the lowest OOP share while uninsured population have the highest OOP share, around 12% and 73%, respectively. People with Medicare/Part D, with Medicare+ or Medicaid+ have about the same OOP share of around 25%. People with PHI have around 40% OOP share on average. With the data of cohort who have unique prescription drug class over a year, ***Table 4*** indicates that in 2017 and 2018, these people have on average around 3-6 prescription fills. Among those who prescribed the popular drugs (6 therapeutic classes), the OOP share differs across different plans and the OOP share is consistent in both years under each type of health plans. The OOP shares are about 20%, 44%, 54%, 35%, 80% for people with Medicaid (only), Medicare/Part D, PHI, Medicaid+ and Uninsured people, respectively. There is slightly bigger difference in the OOP share between 2017 and 2018 for people with Medicare+, 48% in 2017 and 61% in 2018.

Furthermore, ***Table 3*** and ***Table 4*** also present the mean and standard deviation of cost per fill[5] under different health insurance plans, in the year of interest. Results in both tables indicate that the mean cost per fill is below $160 dollar, among all different (therapeutic) class of drugs. Also, the box plots in ***Figure 2*** imply that, the median logged cost per fill is below 4 in all the data, which is about $50 per fill; the 75th quantile of the logged cost per fill is below 5 in all data, which is about $150 per fill. There is some variation in the quantiles of the Rx cost per fill under different health plans. Therefore, we can consider that most people in the sample are not using extremely expensive drugs (specialty). The fact that some people may tend to use specialty drugs can be less concerning in this study.

Table 1: Rx expenditure (log transformed) by different groups (2-year longitudinal data)

[5] Here in this study cost per fill is the same as cost per count, "fill" and "count" are interchangeable.

	Two-year Longitudinal data (adults)					
		2017			2018	
	Mean (SD)	**N=6,123**	**p-value**	**Mean (SD)**	**N=6,123**	**p-value**
Education Level			0.046*			0.034*
High School/GED or below	6.22(1.97)	2784(45.5)		6.47(1.98)	2784(45.5)	
Some College or above	6.09(1.94)	2700(44.1)		6.34(1.92)	2700(44.1)	
Other Degree or Unknown	6.20(2.01)	639(10.4)		6.47(2.05)	639(10.4)	
Race			0.003**			0.007**
White	6.21(1.95)	4722(77.1)		6.45(1.94)	4722(77.1)	
Black/African American	6.07(1.98)	933(15.2)		6.36(2.02)	933(15.2)	
Other	5.91(2.07)	468(7.6)		6.16(1.97)	468(7.6)	
Hispanic Origin			<0.001***			<0.001***
Non-Hispanic	6.23(1.94)	5219(85.2)		6.48(1.93)	5219(85.2)	
Hispanic	5.76(2.03)	904(14.8)		6.04(2.08)	904(14.8)	
Immigration Status			<0.001***			<0.001***
Native Born	6.23(1.96)	5225(85.3)		6.49(1.94)	5225(85.3)	
Immigrants/Unknown	5.76(1.97)	898(14.7)		5.96(2.01)	898(14.7)	
Poverty Level			<0.001***			<0.001***
<200% FPL	6.36(2.03)	2126(34.7)		6.61(2.06)	2063(33.7)	
>=200% FPL	6.06(1.92)	3997(65.3)		6.32(1.90)	4060(66.3)	
Region			<0.001***			<0.001***
Northeast	6.35(1.92)	1079(17.6)		6.57(1.96)	1069(17.5)	
Midwest	6.18(1.94)	1402(22.9)		6.46(1.94)	1401(22.9)	
South	6.22(1.97)	2335(38.1)		6.49(1.98)	2340(38.2)	
West	5.89(1.99)	1307(21.3)		6.11(1.93)	1313(21.4)	
Insurance Plan Type			<0.001***			<0.001***
Medicaid	5.98(2.15)	645(10.5)		6.13(2.22)	629(10.3)	
Medicare/Part D	6.81(1.79)	2173(35.5)		7.10(1.69)	2291(37.4)	
Medicare+	6.72(1.61)	262(4.3)		6.85(1.64)	262(4.3)	
PHI	5.72(1.93)	2632(43)		5.94(1.95)	2531(41.3)	
Medicaid+	6.29(1.95)	147(2.4)		6.04(2.00)	178(2.9)	
Uninsured	5.05(1.76)	264(4.3)		5.37(1.89)	232(3.8)	
Any Limitation			<0.001***			<0.001***
No	5.73(1.90)	3971(64.9)		5.99(1.91)	4093(66.8)	
Yes	7.00(1.81)	2104(34.4)		7.30(1.76)	1994(32.6)	
Unknown	5.31(1.94)	48(0.8)		6.29(1.85)	36(0.6)	
Ever told had diabetes			<0.001***			<0.001***
Not Mention/Unknown	5.85(1.89)	4881(79.7)		6.09(1.88)	4881(79.7)	
Mentioned	7.39(1.76)	1242(20.3)		7.67(1.75)	1242(20.3)	

Notes: 1. Results in the table refer to the mean (SD) of the log transformed Rx expenditures within different groups.
2. Table only presents the categorical variables where the mean log Rx counts is significantly different within levels.

Table 2: Rx expenditure (log transformed) by different groups (cohort with unique drug class in a year)

	Cohort with unique prescription drug class					
		2017			2018	
	Mean (SD)	**N=3,139**	**p-value**	**Mean (SD)**	**N=2,970**	**p-value**
Education Level			<0.001***			0.057*
High School/GED or below	3.73(1.74)	1352(43.1)		4.00(1.77)	1259(42.4)	
Some College or above	4.04(1.75)	1471(46.9)		4.14(1.74)	1433(48.2)	
Other Degree or Unknown	4.00(1.79)	316(10.1)		3.93(1.80)	278(9.4)	
Hispanic Origin			<0.001***			0.018*
Non-Hispanic	4.00(1.76)	2361(75.2)		4.10(1.76)	2393(80.6)	
Hispanic	3.60(1.70)	778(24.8)		3.91(1.74)	577(19.4)	
Immigration Status			<0.001***			<0.001***

	Cohort with unique prescription drug class					
		2017			2018	
	Mean (SD)	N=3,139	p-value	Mean (SD)	N=2,970	p-value
Native Born	3.98(1.76)	2428(77.3)		4.12(1.77)	2415(81.3)	
Immigrants/Unknown	3.62(1.70)	711(22.7)		3.82(1.69)	555(18.7)	
Poverty Level			0.043*			0.002**
<200% FPL	3.80(1.80)	932(29.7)		3.90(1.76)	860(29)	
>=200% FPL	3.94(1.74)	2207(70.3)		4.13(1.75)	2110(71)	
Region			0.009**			0.009**
Northeast	4.08(1.82)	505(16.1)		4.31(1.87)	467(15.7)	
Midwest	4.00(1.78)	642(20.5)		4.01(1.74)	611(20.6)	
South	3.85(1.73)	1128(35.9)		4.04(1.76)	1124(37.8)	
West	3.79(1.72)	864(27.5)		3.99(1.69)	768(25.9)	
Insurance Plan Type			<0.001***			<0.001***
Medicaid	3.70(1.82)	434(13.8)		3.80(1.89)	362(12.2)	
Medicare/Part D	4.18(1.84)	360(11.5)		4.51(1.83)	343(11.5)	
Medicare+	4.30(1.62)	41(1.3)		4.23(1.40)	43(1.4)	
PHI	3.92(1.73)	1945(62)		4.08(1.74)	1839(61.9)	
Medicaid+	4.27(2.09)	90(2.9)		4.15(1.74)	109(3.7)	
Uninsured	3.54(1.52)	269(8.6)		3.67(1.53)	274(9.2)	
Any Limitation			<0.001***			<0.001***
No	3.85(1.72)	2716(86.5)		3.99(1.73)	2622(88.3)	
Yes	4.27(1.93)	382(12.2)		4.64(1.90)	309(10.4)	
Unknown	4.07(2.04)	41(1.3)		4.21(1.51)	39(1.3)	
Therapeutic class			<0.001***			<0.001***
Anti-infectives	3.16(1.66)	561(17.9)		3.31(1.65)	594(20)	
Cardiovascular Agents	4.10(1.46)	578(18.4)		4.26(1.53)	538(18.1)	
Central Nervous System Agents	3.81(1.93)	900(28.7)		3.90(1.94)	821(27.6)	
Hormones/Hormone Modifiers	4.67(1.43)	496(15.8)		4.68(1.43)	534(18)	
Respiratory Agents	4.49(1.81)	186(5.9)		4.62(1.92)	198(6.7)	
Topical Agents	3.64(1.73)	418(13.3)		4.19(1.62)	285(9.6)	

Notes: 1. Results in the table refer to the mean (SD) of the log transformed Rx expenditures within different groups.
2. Table only presents the categorical variables where the mean log Rx counts is significantly different within levels.

Figure 1

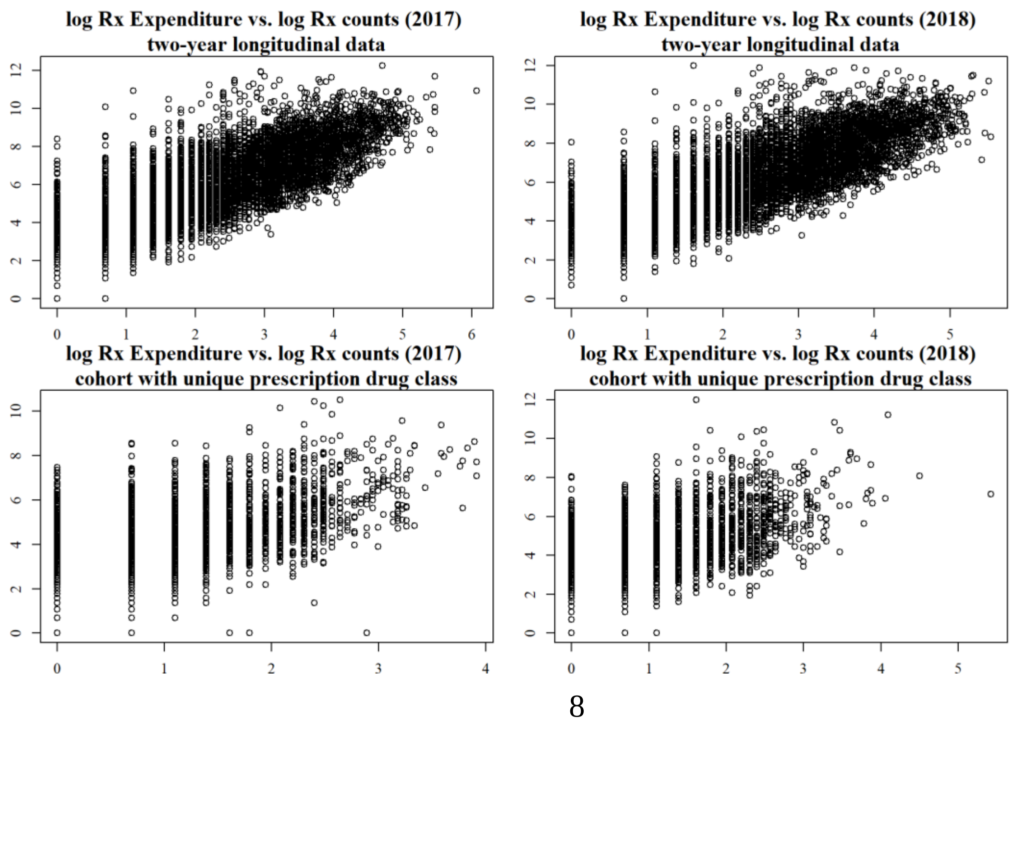


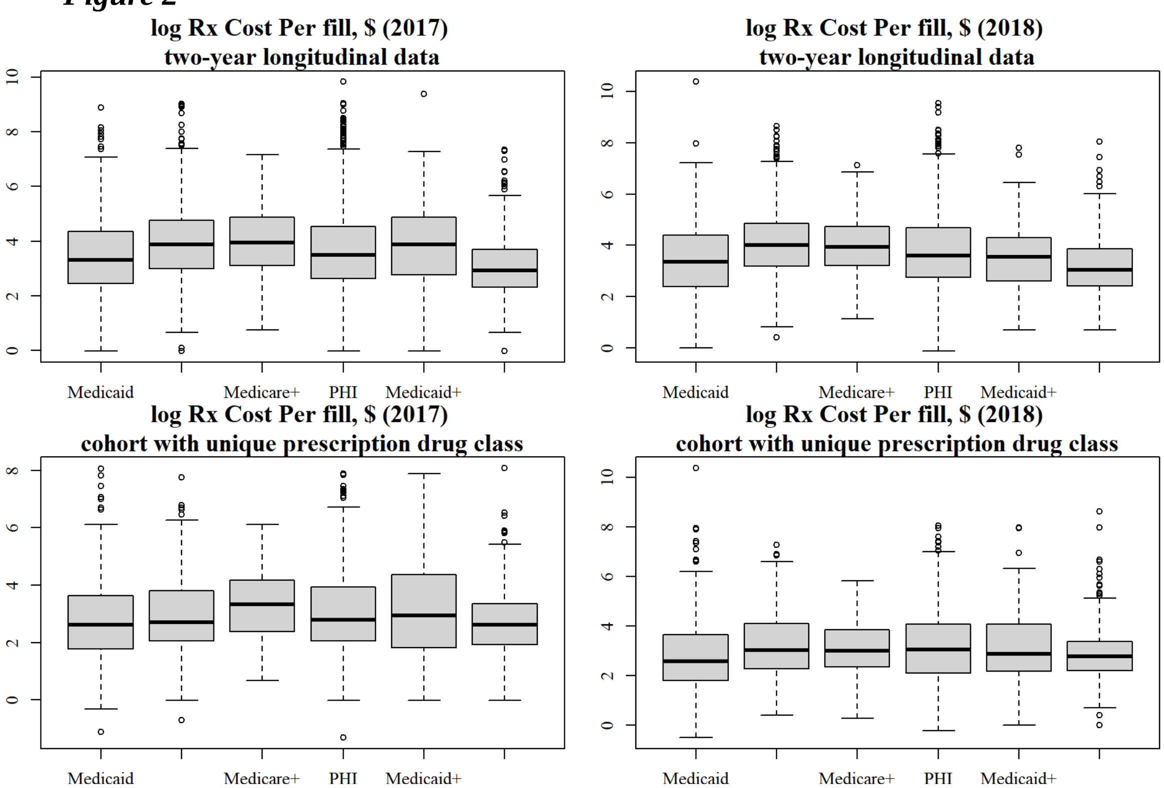
Table 3: Rx count, OOP share, Rx cost/fill by health insurance plan (2-year longitudinal data)

	Medicaid	Medicare/Part D	Medicare+	PHI	Medicaid+	Uninsured
Rx Count, mean (SD)						
2017 (N=6,123)	27.94(38.42)	28.99(28.95)	23.48(22.21)	13.93(17.06)	20.46(22.53)	12.52(15.57)
2018 (N=6,123)	31.18(37.59)	32.02(30.72)	26.24(24.79)	16.02(19.42)	22.76(26.83)	15.34(18.91)
OOP share, mean (SD)						
2017 (N=6,123)	0.1(0.22)	0.26(0.28)	0.26(0.29)	0.4(0.36)	0.24(0.32)	0.74(0.36)
2018 (N=6,123)	0.13(0.27)	0.25(0.27)	0.28(0.29)	0.39(0.35)	0.25(0.32)	0.72(0.38)
Rx cost per fill ($), mean (SD)						
2017 (N=6,123)	105.28 (399.3)	121.49 (405.18)	97.09 (132.91)	133.52 (576.41)	189.05 (985.87)	63.59 (168.91)
2018 (N=6,123)	134.11 (1304.11)	125.97 (295.79)	101.43 (152.93)	141.03 (556.78)	93.54 (246.99)	77.86 (254.21)

Table 4: Rx count, OOP share, Rx cost/fill by health insurance plan (cohort with unique drug class)

	Medicaid	Medicare/Part D	Medicare+	PHI	Medicaid+	Uninsured
Rx Count, mean (SD)						
2017 (N=3,139)	4.26(6.26)	5.27(6.53)	3.68(3.85)	3.73(4.44)	4.38(4.6)	3.34(4.56)
2018 (N=2,970)	5.23(13.6)	5.98(7.79)	4.6(5.18)	3.83(4.42)	4.66(6.06)	3.51(5.54)
OOP share, mean (SD)						
2017 (N=3,139)	0.18(0.34)	0.44(0.4)	0.48(0.39)	0.55(0.41)	0.36(0.43)	0.83(0.34)
2018 (N=2,970)	0.21(0.36)	0.44(0.4)	0.61(0.4)	0.52(0.41)	0.32(0.37)	0.79(0.37)
Rx cost per fill ($), mean (SD)						
2017 (N=3,139)	63.32 (232.87)	66.31 (169.9)	69.38 (105.43)	62.34 (163.19)	121.84 (333.08)	47.41 (209.02)
2018 (N=2,970)	159.6 (1721.09)	75.45 (160.12)	43.95 (66.39)	69.34 (173.36)	106.55 (401.55)	71.22 (393.36)

Figure 2



3.2 Model results

We have 5 regressions in this study using different data sets or different models. OLS1 is the general regression fitted on the pooled full-year data of 2017 and 2018. This model basically replicates the analyses from previous analyses in STAT 520 (except the classification of health insurance type is a little different), for comparison purpose. POLS (Pooled OLS) and RE (Random Effects) are regressions fitted on the 2-year longitudinal data. OLS2 is fitted on the 2018 data of cohort who have unique class of drugs. OLS3 is also fitted on the unique drug class cohort, but with 2017 and 2018 pooled data.

Table 5 displays the premium spending in terms of percentage difference under different Health plans. Compared to OLS1, the premium spending under Medicare/Part D and PHI estimated by POLS and RE models with the longitudinal data remain consistent. On average, holding other covariates fixed, including number of prescriptions but not drug classes/types, people with Medicare/Part D have around 48% and 30% (significantly) higher cost than Medicaid (only) enrollees in 2018 and 2017, respectively; people with Private health insurance have around 35% (significantly) higher cost in both 2017 and 2018. The spending premiums for people with Medicare+ (compared to Medicaid only enrollees) are 49%, 39% and 43% under OLS1, POLS and RE, respectively, all significant. For people with Medicaid+, OLS1 estimated the spending premium is 20% in 2018 and no significant difference in 2017; POLS and RE estimated no significant difference in the Rx expenditure between people with Medicaid+ and Medicaid only in 2018, when holding all other covariates fixed, but according to these two models, on average people with Medicaid+ have around 46% higher Rx expenditure compared to Medicaid (only) enrollees in 2017, again when holding other variables fixed.

When focusing on the cohort who prescribed unique class of drugs. We can see from the last 2 columns in ***Table 5*** that, if controlling for the prescription drug types, when holding all other variables fixed, on average, people with Medicare/Part D have around 33% percent higher cost on the prescription drugs than people with Medicaid only. The spending/cost premiums for people with PHI and Medicaid+ are 30% and 35%, respectively. We did not see significant spending/cost premium for people with Medicare+, noted in the cohort data, the group size for Medicare+ is very small (around 40). The results from fixed effects model using the 2-year longitudinal data are presented in ***Appendix 3***, and it indicates consistent partial effects of different health plans as the estimates presented in ***Table 5***. In general, according to OLS2 and OLS3, there is no very significant difference in any indicated spending premiums between 2017 and 2018.

In all 5 regressions, we did not see any significant difference in the Rx expenditure between Medicaid (only) and Uninsured people, holding other factors fixed (including number of prescriptions).

Figure 3 just visualized the spending premiums for different health plans that estimated by the 5 regressions in ***Table 5.*** It displays consistent information as ***Table 5***.

More comprehensive results of estimates for all explanatory variables (the original regression coefficients) are presented in ***Table 6***. We can see that the results from all regressions are consistent for almost all the other variables. For example, when holding other factors fixed, on average people with degree of some college or above are spending around 15% more compared to people with lower education degree, while immigrants are spending around 13% less compared to people that are native-born. On average 1% increase in the prescription drug count will lead to around 1.2% increase in the expenditure. People in Midwest, South and West all have lower Rx expenditures compared to people in East, when holding other covariates fixed, which indicates that the drug price may be higher in those regions.

In OLS2 and OLS3 where we controlled for the prescription drug class, we can see from the last two columns in ***Table 6*** that, when holding other explanatory variables fixed, compared to people only used Anti-infectives over one year, people using Cardiovascular drugs have significantly lower Rx expenditures (around 40% less), people using Central Nervous System Agents seem to have marginally smaller Rx expenditure, while people using drugs in other 3 therapeutic classes—Hormones/Hormone Modifiers, Respiratory Agents and Topical Agents, have significantly higher Rx expenditures. The spending premiums for those 3 classes are about 25%, 75% and 45%, respectively. These results imply the underly drug price difference among different Therapeutic drug classes.

Table 5: **Regression Results**

	Full-year data OLS1		**2-yr longitudinal POLS**		**2-yr longitudinal RE**		**Cohort 2018 OLS2**		**Cohort 17-18 OLS3**	
2018		*p-value*		*p-value*		*p-value*		*p-value*		*p-value*
Medicare/Part D	0.47 (0.06)	<0.01***	0.47 (0.09)	<0.01***	0.49 (0.09)	<0.01***	0.3 (0.14)	0.011**	0.36 (0.14)	<0.01***
Medicare+	0.49 (0.09)	<0.01***	0.39 (0.12)	<0.01***	0.43 (0.11)	<0.01***	0.11 (0.2)	0.638	0.17 (0.21)	0.463
PHI	0.36 (0.05)	<0.01***	0.36 (0.08)	<0.01***	0.34 (0.08)	<0.01***	0.3 (0.11)	<0.01***	0.29 (0.1)	<0.01***
Medicaid+	0.2 (0.08)	<0.01***	0.12 (0.12)	0.274	0.11 (0.1)	0.269	0.35 (0.2)	0.036**	0.33 (0.2)	0.047**
Uninsured	0.04 (0.06)	0.482	-0.01 (0.09)	0.907	0 (0.09)	0.991	0.16 (0.12)	0.167	0.16 (0.12)	0.152
2017										
Medicare/Part D	0.31 (0.05)	<0.01***	0.29 (0.08)	<0.01***	0.31 (0.08)	<0.01***				
Medicare+										
PHI										
Medicaid+			0.46 (0.19)	<0.01***	0.47 (0.16)	<0.01***				
Uninsured										
Observations	29,595		12,246		12,246		2,970		6,109	

Notes: 1. Results in the table refer to the partial effects in terms of percentage change, which are calculated by *exp(coef)-1* (since dependent variable is log Rx expenditures); the corresponding SE are calculated with Delta method, p-values are calculated based on the percentage change, SE and degree of freedom.
2. For 2017, only the partial effects that are indicated significantly different from that of 2018 are presented in the table.
3. The theta parameter for the Random Effects model in column 3 is 0.45.

Figure 3

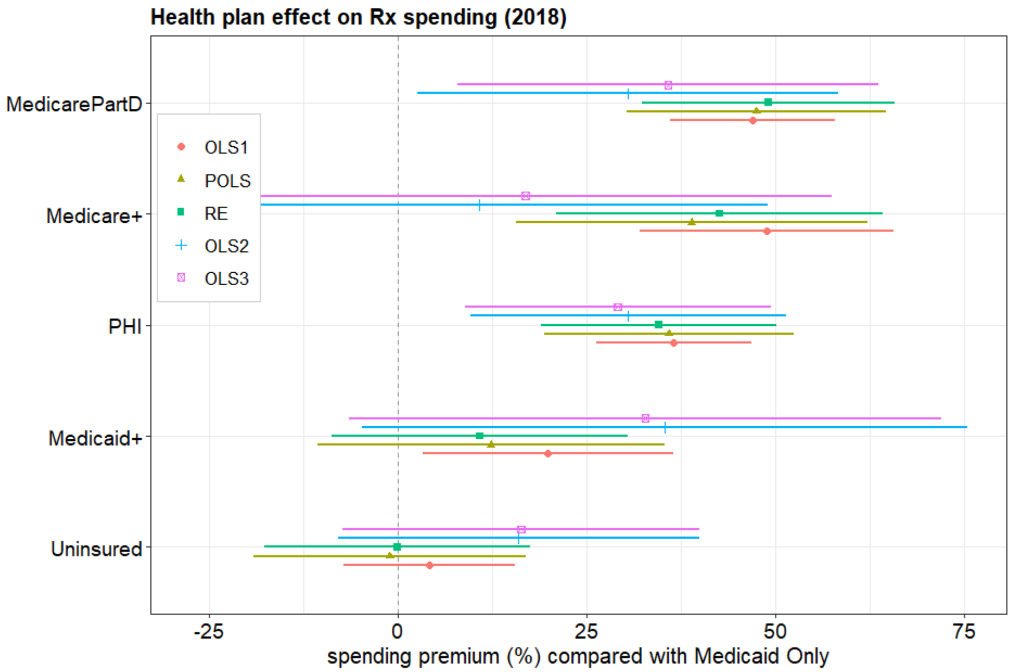


Table 6: Regression results

	Dependent variable: log Rx expenditure				
	OLS	*Panel linear*		*OLS*	
	OLS1	POLS	RE	OLS2	OLS3
Year2017	0.001	0.02	0.01		-0.01
	(0.04)	(0.05)	(0.05)		(0.10)
Medicare/Part D (ref: Medicaid Only)	0.39***	0.39***	0.40***	0.27**	0.31***
	(0.04)	(0.06)	(0.06)	(0.11)	(0.10)
Medicare+	0.40***	0.33***	0.35***	0.10	0.16
	(0.06)	(0.09)	(0.08)	(0.18)	(0.18)
PHI	0.31***	0.31***	0.30***	0.27***	0.26***
	(0.04)	(0.06)	(0.06)	(0.08)	(0.08)
Medicaid+	0.18**	0.12	0.10	0.30**	0.28*
	(0.07)	(0.10)	(0.09)	(0.15)	(0.15)
Uninsured	0.04	-0.01	-0.001	0.15	0.15
	(0.06)	(0.09)	(0.09)	(0.11)	(0.10)
Cardiovascular (ref: Anti-infectives)				-0.35***	-0.45***
				(0.09)	(0.07)
Central Nervous System Agents				-0.10	-0.11*
				(0.08)	(0.06)
Hormones/Hormone Modifiers				0.26***	0.25***
				(0.09)	(0.06)
Respiratory Agents				0.76***	0.75***
				(0.12)	(0.08)
Topical Agents				0.55***	0.34***
				(0.10)	(0.07)
Some College/above (ref: Below Some College)	0.14***	0.14***	0.14***	0.11**	0.15***
	(0.02)	(0.03)	(0.03)	(0.05)	(0.04)

	Dependent variable: log Rx expenditure				
	OLS	*Panel linear*		*OLS*	
	OLS1	POLS	RE	OLS2	OLS3
Other Degree/Unknown	0.04*	0.07	0.07	-0.04	0.07
	(0.03)	(0.05)	(0.05)	(0.09)	(0.06)
Black/African American (ref: White)	-0.14***	-0.18***	-0.18***		
	(0.02)	(0.04)	(0.04)		
Other Race	-0.01	-0.02	-0.02		
	(0.03)	(0.06)	(0.06)		
Hispanic (ref: Non-Hispanic)	-0.08***	-0.09*	-0.10**		
	(0.02)	(0.05)	(0.05)		
Immigrants/Unknown (ref: Native)	-0.14***	-0.12***	-0.13***	-0.11*	-0.13***
	(0.02)	(0.05)	(0.05)	(0.07)	(0.04)
Family size	-0.01**	-0.01	-0.02	-0.03*	-0.02*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
>=200% FPL (ref: <200% FPL)	0.07***	0.06**	0.04		
	(0.02)	(0.03)	(0.03)		
Log Rx Count	1.24***	1.21***	1.19***	1.20***	1.23***
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
Midwest (ref: East)	-0.18***	-0.16***	-0.16***	-0.23***	-0.13**
	(0.02)	(0.05)	(0.04)	(0.08)	(0.06)
South	-0.12***	-0.11***	-0.12***	-0.20**	-0.13**
	(0.02)	(0.04)	(0.04)	(0.08)	(0.05)
West	-0.17***	-0.19***	-0.19***	-0.23***	-0.17***
	(0.02)	(0.05)	(0.05)	(0.08)	(0.05)
Has Limitation (ref: No limitation)	0.12***	0.10***	0.07***	0.20**	0.14**
	(0.02)	(0.03)	(0.03)	(0.09)	(0.06)
Unknown	0.09	-0.03	0.02	0.30	0.28*
	(0.09)	(0.14)	(0.12)	(0.21)	(0.16)
Ever had Diabetes (ref: No diabetes)	0.42***	0.43***	0.44***		
	(0.02)	(0.04)	(0.04)		
Year2017: Medicare/Part D	-0.11**	-0.14**	-0.13**		-0.18
	(0.05)	(0.06)	(0.05)		(0.14)
Year2017: Medicare+	-0.05	-0.04	-0.02		0.35
	(0.08)	(0.09)	(0.08)		(0.28)
Year2017: PHI	-0.08	-0.07	-0.07		-0.11
	(0.05)	(0.06)	(0.06)		(0.11)
Year2017: Medicaid+	0.16	0.26*	0.29**		0.10
	(0.11)	(0.13)	(0.12)		(0.24)
Year2017: Uninsured	-0.05	-0.12	-0.07		-0.11
	(0.08)	(0.10)	(0.10)		(0.14)
Observations	29,595	12,246	12,246	2,970	6,109
R^2	0.65	0.60	0.55	0.45	0.45
Adjusted R^2	0.65	0.60	0.54	0.44	0.45

Notes: *the table presents the raw coefficients from the models. Standard errors in () are robust SE.*

*p<0.1**p<0.05***p<0.01

4. CONCLUSIONS

All models in the analysis imply that overall, holding other things equal, individuals with non-Medicaid-only health insurance plans have higher prescription drug expenditures, given the same prescription drug utilization, similar health situation and/or prescription drug type/class. Specifically, when including people who prescribe any drugs, the spending premiums (%) for Medicare/Part D and Private Health Insurance (PHI) estimated by models fitted with 2-year longitudinal data (POLS and RE) are almost identical to the estimates from STAT 520 (OLS1). When including only people who use single (Therapeutic) class of drugs over a year and controlling the drug type/class in the models (OLS2 and OLS3), the spending premiums (%) for Medicare/Part D and PHI are estimated to be slightly smaller than the estimates from models using the longitudinal data. These results indicate that people with Medicare/Part D and PHI health plans do pay higher price than Medicaid (only) enrollees, which presumably leads to the spending premiums (higher expenditures). The omitted drug type information inflates the spending premiums slightly in models that do not control Therapeutic class, which implies that the potential tendency or preference of utilization of certain therapeutic drug classes (that are more expensive) in Medicare/Part D and PHI can further cause higher spending premiums.

In the meantime, the spending premiums for Medicare/Part D enrollees are slightly higher than for people with Private Health Insurance (according to OLS3). We did not see any significant difference in prescription drug expenditures caused by the potential drug price difference between Medicaid only and Uninsured groups, this may be because the uninsured people can use manufactural coupons while the people covered by Medicaid cannot, the potential higher price for uninsured people is cancelled out by the coupon discount. Furthermore, when controlling drug utilization and drug type, also holding other variables fixed, the non-significant difference in spending premiums between 2017 and 2018 indicate that the price inequality between Medicaid (only) and Medicare/Part D and PHI may remain the same over the 2-year period. In terms of the price of different drugs, on average, Hormones/Hormone Modifiers, Respiratory Agents and Topical Agents are significantly more expensive than Anti-infectives, with Respiratory Agents most expensive, and Cardiovascular are significantly cheaper than Anti-infectives.

Table 3 and ***Table 4*** show that the average out-of-pocket payment of prescription drugs for Medicare/Part D and PHI only are higher than that of Medicaid (only), which means that people facing higher prices are paying more out of pocket. Therefore, the underlying price inequality further increases the financial burden for people with other than Medicaid health plans, especially for the elderly enrolled in Medicare and pay premiums for the prescription drug coverage, Part D.

5. DISCUSSIONS

5.1 Study strengths and limitations

In this study, most important factors that are available from MEPS data have been taken into consideration. Compared to previous analyses, we improve the models by controlling drug therapeutic class, which reduced the estimates bias. Even though we still did not capture the information about the preference/tendency to use Specialty drugs (Specialty drugs are much more expensive than non-specialty drugs, such as generic), the data shows only small number of people spend more than $1000 on average for one prescription fill, and table 1 of the Research Fining #44[6] from Agency for Healthcare Research and Quality showed that the distribution of outpatient prescriptions filled with different types of drugs (around 2%, 11%, 3%, 84% for Specialty, Single

[6] https://meps.ahrq.gov/data_files/publications/rf44/rf44.pdf

Source, Originator, Generic, respectively) is pretty similar across all different insurance plans (Medicaid, Medicare, Private, Tricare or VA, Uninsured), we may assume people with different health insurance plans have the same probabilities of using different types of drugs, and are less likely to use extremely expensive drugs. Also, some of the existing factors such as self-reported health limitation may potentially reflect the Specialty drug utilization to some degree (less healthy people may need to use more expensive drugs), which can reduce the bias caused by the missing information.

Overall, the analysis results from this study are consistent with the results of Retail drug prices by insurance status/plans in the table 2 of Research Fining #44, where they found the average retail unit price for most of the drugs are more expensive for Medicare and Private health insurance than for Medicaid, and interestingly, the average retail unit price are smaller or similar for uninsured group than that for Medicaid.

In terms of the insurance type definitions, the classification in this study better captures the price of prescription drugs used by the elderly with Medicare, even though the payment may be made by Part D plans as PHI, the price can directly impact the premiums that the elderly pays for Part D plans in general. One thing to be noted, as mentioned in the Method session and in the ***Table 3*** and ***4***, the majority people with Medicare have Part D coverage, the classification methods in this study lead to a very smaller number of people in Medicare+ group. So, in the analyses, we probably could have just excluded Medicare+ group in the analyses, or combined Medicare/Part D and Medicare+ groups into one category and conduct some sensitivity analysis to see how the results change.

Due to the feature of surveys, there may be some potential measurement error in some variables. For example, according to MEPS data document[7], in 2018 survey, the respondents were allowed to report both Medicaid and other public hospital/physician coverage. Previously, these types of coverage were mutually exclusive. Also, there can be some measurement error of the prescription medication counts, and the expenditure, since some of these data were self-reported if the individual self-filed the insurance claims or the individual is uninsured, which may be less accurate compared to the same data obtained from the pharmacy companies' records.

5.2 Possible implications and future research

As mentioned above, Specialty drug utilization information was not controlled in this analysis, further analysis can be done separately for Specialty drug users and Non-specialty drug users, to see the partial effects of health plans on specific drug type—how the underlying price difference may look like for Specialty drugs and Generic drugs. Or we can just exclude the individuals with extremely high cost per fill (count), and do the same analysis to see how results change.

In this study, we mainly focus only the retail price (upfront price) at purchase, which the health insurance plan can negotiate with the retail outlets. There is another transaction between manufactures and Health insurance plans, which is the manufacture rebates that usually happen behind the scenes and usually paid directly from manufactures to health plans (or health plan representatives) after the consumer purchasing transaction. The rebates can further offset some of what is paid at retail by health insurance plans. It is hard to get exact information about the manufacture rebates rate for different health plans because of confidential reasons and the flow through the system is mysterious, but previous research[8] did find that the rebates are much higher for Medicaid than for Medicare and private health insurances (51%, 22%, 12% for Medicaid,

[7] https://meps.ahrq.gov/data_stats/download_data/pufs/h209/h209doc.pdf
[8] https://altarum.org/sites/default/files/Altarum-Prescription-Drug-Rebate-Report_April-2018.pdf

Medicare, and PHI, respectively); the coupon discount/rebate rate for uninsured people is about 22% according to the same report, which matches the OOP share calculated based on MEPS data (72-83% in ***Table 3*** and ***4***). This means that the health insurance plans paying higher prices are getting less rebates from manufactures, so the health plan partial effects (spending premiums for non-Medicaid-only health plans) in this study could have be even higher. Overall, both health plans and consumers are paying more under Medicare/Part D and private health insurance plans compared to Medicaid. Further related research can be done to figure out more about the payment system if data available.

Also, the model results in this study imply that the underlying price seems not quite different between Medicaid only and Uninsured population, but of course uninsured people are paying much more out-of-pocket than Medicaid enrollees. Future research can be done to explore more about the prescription drug prices for uninsured people.

Appendix

Appendix 1: References

[1] Steven C Hill, Samuel H Zuvekas, Marc W Zodet. *Implications of the accuracy of MEPS prescription drug data for health services research*. Fall 2011. https://pubmed.ncbi.nlm.nih.gov/22235548/

[2] Miller, G.E., Hill, S.C., and Ding, Y. *Retail Drug Prices, Out-of-Pocket Costs, and Discounts and Markups Relative to List Prices: Trends and Differences by Drug Type and Insurance Status, 2011 to 2016*. Research Findings #44. October 2019. Agency for Healthcare Research and Quality, Rockville, MD. https://meps.ahrq.gov/data_files/publications/rf44/rf44.pdf

[3] Jessie X. Fan, Deanna L. Sharpe and Goog-Soog Hong: *Health care and prescription drug spending by seniors.* March 2003. https://www.bls.gov/opub/mlr/2003/03/art3full.pdf

[4] Charles Roehrig: *The impact of prescription drug rebates on Health plans and consumers*. April 2018. https://altarum.org/sites/default/files/Altarum-Prescription-Drug-Rebate-Report_April-2018.pdf

[5] Understanding Medicare Advantage Plans. CMS. https://www.medicare.gov/Pubs/pdf/12026-Understanding-Medicare-Advantage-Plans.pdf

[6] Part D / Prescription Drug Benefits. https://medicareadvocacy.org/medicare-info/medicare-part-d/#introduction

[7] How Does Medicare Part D Work with Other Insurance. https://www.ehealthmedicare.com/medicare-part-d-articles/how-does-medicare-part-d-work-with-other-insurance/

Appendix 2: Examples of insurance type category

Insurance Type	Example
Medicaid Only	Person A only has Medicaid coverage, no any other health insurance.
Medicare/Part D (Medicare w or w/o part D)	1. Person A has Medicare as well as Part D coverage (no other medical insurance), he reported in the survey that he only has Medicare coverage but did not report Part D coverage, since he did not know Part D is separate from Original Medicare 2. Person B has Medicare as well as Part D coverage (no other insurance), he reported both Medicare and Part D coverage in the survey, but he did not report that he has any Private health insurance, since he did not know Part D can only be purchased from Private Insurance company, Part D is a type of PHI. 3. Person C has Medicare as well as Part D (no other insurance), he reported both coverages in the survey, he also reported that he has Private Insurance coverage. 4. Person D has Medicare as well as Part D coverage (from any private insurance), he is **still working** but not getting any additional Health insurance from his employer/retirement group/other group insurance. 5. Person E has Medicare as well as Part D (from any private companies), he **retired**, but he still gets retiree benefits (including private health insurance) from his former employer. **NOTE**: This may be rare since the individual needs to pay double premiums.

Insurance Type	Example
	6. Person F has Medicare as well as Part D, also he has Medicaid and/or other public insurance such as TRICARE, CHAMPUS.
Medicare+	1. Person A has Medicare but no Part D coverage, he is working and getting Health insurance from his employer. 2. Person B has Medicare and Part D coverage, he is **working** and also getting additional Health insurance from his employer/retirement group/any group insurance. **NOTE:** This may be rare, since again he needs to pay premiums for both Part D and the group insurance. 3. Person C has Medicare but no Part D coverage, he has some other insurance such as Medicaid, TRICARE (for veterans), CHAMPUS.
Private Health Insurance (PHI)	Person A only has Private Insurance coverage, including TRICARE, CHAMPUS
Medicaid+	Person A has Medicaid but no Medicare coverage. He also has some other insurance in the meantime, can be either private or other public insurance

Appendix 3: Fixed effects estimates using 2017-2018 two-year longitudinal data.

```
t test of coefficients:

                              Estimate Std. Error t value  Pr(>|t|)    
year2017                    -0.0087296  0.0509894 -0.1712  0.864069    
instypeMedicarePartD         0.4476576  0.1559771  2.8700  0.004119 ** 
instypeMedicare+             0.4491865  0.1694079  2.6515  0.008034 ** 
instypePHI                   0.2230837  0.1288492  1.7314  0.083439 .  
instypeMedicaid+             0.0451745  0.1314296  0.3437  0.731071    
instypeUninsured             0.0855456  0.1510709  0.5663  0.571237    
famsize                     -0.0344617  0.0360626 -0.9556  0.339308    
povcat>=200%                 0.0161880  0.0376668  0.4298  0.667378    
logrxprmedsno                1.1333319  0.0219333 51.6717 < 2.2e-16 ***
regionmepsMidwest           -0.1423127  0.3565351 -0.3992  0.689793    
regionmepsSouth             -0.1203607  0.2775261 -0.4337  0.664528    
regionmepsWest               0.4977541  0.3586405  1.3879  0.165221    
anylmtYes                    0.0100550  0.0363041  0.2770  0.781816    
anylmtunk                    0.0469351  0.1491696  0.3146  0.753044    
year2017:instypeMedicarePartD -0.1187810  0.0553729 -2.1451  0.031983 *  
year2017:instypeMedicare+    0.0022128  0.0823261  0.0269  0.978557    
year2017:instypePHI         -0.0588587  0.0573211 -1.0268  0.304544    
year2017:instypeMedicaid+    0.3130752  0.1197336  2.6148  0.008951 ** 
year2017:instypeUninsured   -0.0385754  0.1001014 -0.3854  0.699981    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Notes: Some of the coefficients may be inappropriate and biased, since in Fixed effects model we should not include the time-constant variable, while in the model some of the explanatory variables may be time constant or change very little, such as family size, region, health limitation.

Appendix 4: Code

Data Processing

```
##############################################################################
############ Process the MEPS FYC data, identify whether the person has
Medicare part D coverage 2018 data
load("Data/fyc2018_for_analysis.rda")
fyc2018['mcrpho'] <- 0
fyc2018[fyc2018$mcrpho18==1 | fyc2018$mcrpho31==1 |
fyc2018$mcrpho42==1,'mcrpho'] <- 1

fyc2018['mcrpb'] <- 0
fyc2018[fyc2018$mcrpb18==1 | fyc2018$mcrpb31==1 | fyc2018$mcrpb42==1,'mcrpb']
<- 1

fyc2018['mcrpd'] <- 0
fyc2018[fyc2018$mcrpd18x==1 | fyc2018$mcrpd31x==1 |
fyc2018$mcrpd42x==1,'mcrpd'] <- 1

fyc2018['employed'] <- 0 # employment status
fyc2018[fyc2018$empst31h==1 | fyc2018$empst42h==1 |
fyc2018$empst53h==1,'mcrpd'] <- 1

fyc2018['mepsid'] <- fyc2018$dupersid
fyc2018['year'] <- 2018
fyc2018 <- fyc2018[,c('year', 'mepsid', 'mcrpho', 'mcrpb', 'mcrpd',
'employed')]

# 2017 data
load("Data/fyc2017_for_analysis.rda")
fyc2017['mcrpho'] <- 0
fyc2017[fyc2017$mcrpho17==1 | fyc2017$mcrpho31==1 |
fyc2017$mcrpho42==1,'mcrpho'] <- 1

fyc2017['mcrpb'] <- 0
fyc2017[fyc2017$mcrpb17==1 | fyc2017$mcrpb31==1 | fyc2017$mcrpb42==1,'mcrpb']
<- 1

fyc2017['mcrpd'] <- 0
fyc2017[fyc2017$mcrpd17x==1 | fyc2017$mcrpd31x==1 |
fyc2017$mcrpd42x==1,'mcrpd'] <- 1

fyc2017['employed'] <- 0 # employment status
fyc2017[fyc2017$empst31h==1 | fyc2017$empst42h==1 |
fyc2017$empst53h==1,'mcrpd'] <- 1

fyc2017['mepsid'] <- fyc2017$panel*100000000 + as.numeric(fyc2017$dupersid)
fyc2017['mepsid'] <- as.character(fyc2017$mepsid)
fyc2017['year'] <- 2017

fyc2017 <- fyc2017[,c('year', 'mepsid', 'mcrpho', 'mcrpb', 'mcrpd',
'employed')]

# append fyc 2017 and 2018 data
fyc = rbind(fyc2017, fyc2018)
```

```r
# Load the IPUMS MEPS data
load("Data/MEPS_for_analysis.rda")
# CPI
# https://meps.ipums.org/meps/userNotes_cpi2009.shtml

# fill the missing ages with the agelast variable
meps[meps$age==99, 'age'] <- meps[meps$age==99, 'agelast']

# fill out people who had positive prescription fills
df <- meps %>%
  filter(rxprmedsno>0 & year>=2017 & regionmeps>0) %>%
  inner_join(fyc, by=c('year', 'mepsid'))

df['agegrp'] <- 0
df$agegrp[df$age>30 & df$age<=45] <-1
df$agegrp[df$age>45 & df$age<=60] <-2
df$agegrp[df$age>60 & df$age<=75] <-3
df$agegrp[df$age>75] <-4
table(df$agegrp)

df$agegrp <- factor(df$agegrp, labels=c("<31", "31-45", "46-60", "61-75",
">75"))

df$sex <- df$sex-1
table(df$sex)
df$sex <- factor(df$sex, labels=c("Male", "Female"))

# need to recode marital status var
df["marriage"] <- 0 #"NotMarried"
df$marriage[df$marstat == 10] <-1 #"Married"
table(df$marriage)

df$marriage <- factor(df$marriage, labels=c("Unmarried", "Married"))

# need to recode race--White, Black, Alaskan Native or American Indian,
# Asian or Pacific Islander, Other
df["race"] = 0 #"White"
df$race[df$racea==200] <- 1 #"Black/African American"
df$race[df$racea>300] <- 2 #"Other"
table(df$race)

df$race <- factor(df$race,
                  labels=c("White", "Black/African American", "Other"))

# maybe use educ to create a new education level
df["edulvl"] <-0 #"NoDegree"
df$edulvl[df$educ>400] <- 1
df$edulvl[df$educ>=604] <- 2
df$edulvl <- factor(df$edulvl,
                    labels=c("HighSchool/GED or below", "SomeCollege or
above",
                             "OtherDegree or Unknown"))
table(df$edulvl)

df$povcat <- ifelse(df$povcat>=4, 1, 0)
table(df$povcat)
df$povcat <- factor(df$povcat,
```

```
                     labels=c("<200%", ">=200%"))

df[df$regionmeps==0, 'regionmeps'] <- 5
df$regionmeps <- df$regionmeps-1
table(df$regionmeps)
df$regionmeps <- factor(df$regionmeps,
                        labels=c("Northeast", "Midwest", "South", "West"))

df["immigr"] <- 0 #"NativeBorn"
df$immigr[df$usborn != 20] <- 1 #"ImmigrantOrUnknown"
table(df$immigr)
df$immigr <- factor(df$immigr, labels=c("NativeBorn", "Immigrants/unk"))

df$hispyn <- df$hispyn-1
table(df$hispyn)
df$hispyn <- factor(df$hispyn, labels=c("NonHispanic", "Hispanic"))

df$usualpl[!df$usualpl %in% c(1,2)] <- 3 # unknown
df$usualpl <- df$usualpl-1
table(df$usualpl)
df$usualpl <- factor(df$usualpl, labels=c("No", "Yes", "unk"))

# create dummy indicating whether the person is covered by
employer/union/other group insurance.
df['pegcov'] <- ifelse(df$pegja==2 | df$pegfe==2 | df$pegma==2 | df$pegap==2
|
                          df$pegmy==2 | df$pegju==2 | df$pegjl==2 |
df$pegau==2 |
                          df$pegse==2 | df$pegoc==2 | df$pegno==2 |
df$pegde==2, 1, 0)
df['pogcov'] <- ifelse(df$pogja==2 | df$pogfe==2 | df$pogma==2 | df$pogap==2
|
                          df$pogmy==2 | df$pogju==2 | df$pogjl==2 |
df$pogau==2 |
                          df$pogse==2 | df$pogoc==2 | df$pogno==2 |
df$pogde==2, 1, 0)
# any employer/group/union coverage
df['pgrcov'] <- ifelse(df$pegcov==1 | df$pogcov==1, 1, 0)

# for the health insurance, there are many combinations
df$hiprivate <- df$hiprivate-1
df$himcare <- df$himcare-1
df$himachip <- df$himachip-1
df$hichampany <- df$hichampany-1
df["instypetot"] <- df$hiprivate+df$himcare+df$himachip+df$hichampany
table(df$instypetot)

df["instype"] <- 5 # uninsured
df$instype[df$instypetot==1 & df$himachip==1] <- 0 # Medicaid only
# Medicare only or Covered by Medicare Part D
df$instype[(df$instypetot==1 & df$himcare==1) |
             (df$instypetot>1 & df$himcare==1 & df$mcrpd==1 & (df$employed==0
| df$pgrcov ==0))] <- 1
# Covered by Medicare but no Part D coverage (Medicare or Medicare
combination)
df$instype[df$instypetot>1 & df$himcare==1 & df$instype>1] <- 2 # &
df$rxexpmc>0
```

```
# df$instype[df$instypetot>1 & df$himcare==1 & df$himachip==0 & df$instype>1]
<- 2 # & df$rxexpmc>0
df$instype[(df$instypetot==1 & (df$hiprivate==1 | df$hichampany==1)) |
             (df$instypetot==2 & df$hiprivate==1 & df$hichampany==1)] <- 3 #
Private only
df$instype[df$instypetot>1 & df$himachip==1 & df$himcare==0] <- 4 # Medicaid
plus other insurances
# df$instype[df$instype>4 & df$instypetot>1 & df$himachip==0 & df$himcare==0]
<- 5 # No such case
table(df$instype)

df$instype <- factor(df$instype,
                     labels=c("Medicaid", "MedicarePartD", "Medicare+",
"PHI",
                              "Medicaid+", "Uninsured"))

# create some other variables
df['oopshare'] <- df$rxexpself/df$rxexptot
df["totvists"] <- df$obtotvis + df$optotvis
#table(df$totvists)

# adjust the dollar into 2018 dollar amount (CPI is 4-digit number, so need
to divide by 1000)
df['rxexptot_adj'] <- df$rxexptot*df$cpi2009*0.854/1000
df['logRxExp'] <- log(df$rxexptot_adj+1)

# df['loginc'] <- log(df$inctot)
df['exptot_adj'] <- df$exptot*df$cpi2009*0.854/1000
df['loghealthexp'] <- log(df$exptot_adj+1)
df['rxprmedsno_stnd'] <- (df$rxprmedsno-
mean(df$rxprmedsno))/sd(df$rxprmedsno)
df['rxprmedsno_sqr'] <- df$rxprmedsno^(1/2)
df['rxprmedsno_qdr'] <- df$rxprmedsno^(1/3)
df['logrxprmedsno'] <- log(df$rxprmedsno)
df['logrxprmedsno_sqr'] <- df$logrxprmedsno^2

df['agesqr'] <- df$age^2
df['totvisitsqr'] <- df$totvists^2
# create a indicator of whether the person is above 18 or not
# df['above18'] = ifelse(df$age>=18, 1, 0)
# df$above18 <- factor(df$above18, labels=c("No", "Yes"))

#######################################################
# restrict to people who are eligible for the SAQ part #
#######################################################

table(df$saqelig)
df_adt <- df %>%
  filter(saqelig>1)
# inctot>=0

df_adt$health[df_adt$health %in% c(0,7,8,9)] <-6 # health status unknown
df_adt$health <- df_adt$health -1
# 0-Excellent, 1-very good, 2-good, 3-fair, 4-poor, 5-unknown
df_adt$health.status <- 0
df_adt[df_adt$health %in% c(0,1), 'health.status'] <- 1
df_adt[df_adt$health %in% c(2,3,4), 'health.status'] <- 2
```

```
df_adt[df_adt$health == 5, 'health.status'] <- 3
df_adt$health.status <- factor(df_adt$health.status,
                        labels=c("VeryGood/above", "Good/below", "Unknown"))

# df_adt$health <- factor(df_adt$health,
#                         labels=c("Excellent", "VeryGood", "Good",
#                                  "Fair", "Poor", "unk"))
table(df_adt$health.status)
df_adt$anylmt[df_adt$anylmt>2] <- 3
df_adt$anylmt <- df_adt$anylmt-1
df_adt$anylmt <- factor(df_adt$anylmt,
                        labels=c("No", "Yes", "unk"))
table(df_adt$anylmt)
# 0-No, 1-Yes, 2-unknown
df_adt$diabeticev[df_adt$diabeticev>2 | df_adt$diabeticev==0] <- 1
df_adt$diabeticev <- df_adt$diabeticev-1
# recode do not know to 0, as not mentioned or do not know
df_adt$diabeticev <- factor(df_adt$diabeticev,
                            labels=c("NotMention/unk", "Mentioned"))
table(df_adt$diabeticev)

df_adt$addaya[df_adt$addaya>2] <- 3 # unknown

df_adt$addaya <- factor(df_adt$addaya,
                        labels=c("NotLimited", "AlittleLimited",
                                 "LimitedAlot", "unk"))
# Health now limits moderate activities
# 0-Not limited, 1-Limited a little, 2-Limited a lot, 3-Unknown
table(df_adt$addaya)

# 3 or above-- major depressive disorder is likely.
df_adt["likelydepressed"]<- 0 # no
df_adt$likelydepressed[df_adt$phq2>=3 & df_adt$phq2<=6] <- 1 # yes
df_adt$likelydepressed[df_adt$phq2>=96] <- 2 # unknown
table(df_adt$likelydepressed)

df_adt$likelydepressed <- factor(df_adt$likelydepressed,
                                 labels=c("No", "Yes", "unk"))
# keep only needed variables
keep_var = c("year", "age", 'duid', "mepsid", "sex", "agegrp", "marriage",
"race", "hispyn", "edulvl", "immigr",
             "regionmeps", "famsize", "inctot", "cpi2009", "povcat",
"povlev", "saqelig",
             "health", "usualpl", "hinotcov", "hiprivate", "hichampany",
"himachip","himcare",
             "hiothgova", "hiothgovb", "covertype", "anylmt", "cancerev",
"diabeticev",
             "chgtot", "exptot", "rxprmedsno", "rxexptot", "rxexpself",
"rxexpmc", "rxexpma",
             "rxexppr", "rxexpva", "rxexptr", "rxexpof", "rxexpol",
"rxexpwc", "rxexpopr",
             "rxexpopu", "rxexpos", "rxexpptr", "rxexpoth", "addaya", "phq2",
"perweight",
             "totvists", "logRxExp", "loghealthexp", "rxprmedsno_stnd",
             "logrxprmedsno", "logrxprmedsno_sqr", "agesqr", "totvisitsqr",
"mcrpb", "mcrpd", "mcrpho",
```

```
              "instype", "rxexptot_adj", "rxprmedsno_sqr", 'rxprmedsno_qdr',
'oopshare', 'pegcov', 'pogcov', 'pgrcov')

meps_all <- df[, keep_var]
meps_adt <- df_adt[, c(keep_var, "likelydepressed", 'health.status')]
save(meps_all, file = "Data/meps_all_processed.rda")
save(meps_adt, file='Data/meps_adult_processed.rda')
######################################################################
####################Process the Rx data###############################
load("Data/rx18_for_analysis.rda")
load("Data/rx17_for_analysis.rda")

rx17['mepsid'] <- (as.numeric(rx17$panel)+20)*100000000 +
as.numeric(rx17$dupersid)
rx17['mepsid'] <- as.character(rx17$mepsid)

# change the factor to numeric for the TC1 variable
rx17['tc1'] <- as.character(rx17$tc1)

rx17['tc1'] <- gsub("[a-zA-Z]", "", rx17$tc1)
rx17['tc1'] <- gsub("[/]", "", rx17$tc1)

rx17['tc1'] <- gsub("[-]$", "", rx17$tc1)
rx17['tc1'] <- as.numeric(rx17$tc1)

# keep the people who prescribed only one type of Drug
rx17 <- rx17 %>%
  group_by(mepsid) %>%
  mutate(tc.max = max(tc1),
         tc.min = min(tc1),
         daysup.max = max(rxdaysup),
         daysup.min = min(rxdaysup)) %>%
  filter(tc.min==tc.max & tc1>0) %>%
  group_by(mepsid, tc1) %>%
  summarise(rxfills = n(),
            rxexp = sum(rxxp17x)) %>%
  mutate(year=2017)

rx18['mepsid'] <- rx18$dupersid
rx18 <- rx18 %>%
  group_by(mepsid) %>%
  mutate(tc.max = max(tc1),
         tc.min = min(tc1),
         daysup.max = max(rxdaysup),
         daysup.min = min(rxdaysup)) %>%
  # specialty = ifelse(rxxp18x>200, 1, 0),
  # use.specialty = max(specialty)
  filter(tc.min==tc.max & tc1>0) %>%
  group_by(mepsid, tc1) %>%
  summarise(rxfills = n(),
            rxexp = sum(rxxp18x))%>%
  mutate(year=2018)
rx <- rbind(rx17, rx18)
rx['tc1'] <- as.factor(rx$tc1)
save(rx, file='Data/rx_1718.rda')

######################################################################
```

```
#####Merge MEPS data with Rx TC data to get people who prescribed######
####unique type of drugs############################################
df <- meps_all[meps_all$year>=2017,]
table(df$instype)

uniqueRx_all <- df %>%
  inner_join(rx, by=c('year', 'mepsid'))

save(uniqueRx_all, file='Data/uniqueRx_all.rda')

# adt
df <- meps_adt[meps_adt$year>=2017,]
table(df$instype)

uniqueRx_adt <- df %>%
  inner_join(rx, by=c('year', 'mepsid'))

save(uniqueRx_adt, file='Data/uniqueRx_adt.rda')
##################################################################
#################Process the longitudinal data ###################
load('Data/meps_all_processed.rda')
load('Data/meps_adult_processed.rda')

# all people
meps_all2yr <- meps_all %>%
  filter(year %in% c(2017,2018)) %>%
  group_by(mepsid) %>%
  mutate(idcnt = n()) %>%
  filter(idcnt==2)
table(meps_all2yr$idcnt)

# We can actually get the longitudinal data directly from IPUMS MEPS
save(meps_all2yr, file='Data/meps_all2yr.rda')

# adults
meps_adt2yr <- meps_adt %>%
  filter(year %in% c(2017,2018)) %>%
  group_by(mepsid) %>%
  mutate(idcnt = n()) %>%
  filter(idcnt==2)
table(meps_adt2yr$idcnt)

save(meps_adt2yr, file='Data/meps_adt2yr.rda')
```

Analysis

```
# function used to create the descriptive tables (of t-tests and anova)
test <- function(dt, byvar) {
  # by_var <- enquo(byvar)
  by_var <- rlang::sym(byvar)
  temp <- dt %>%
    group_by(!!by_var) %>%
    summarise(mean=mean(logRxExp),
              sd=sd(logRxExp),
              N = n())
  varname <- colnames(temp)[1] # mean of logRxExp by groups
```

```
  # anova test/t-test
  if (length(table(dt[,byvar])>2)) {
    anv=anova(lm(dt[['logRxExp']] ~ dt[[byvar]]))
    pval <- anv[['Pr(>F)']][1] # extract the p value
  } else {
    ttest=t.test(dt[['logRxExp']] ~ dt[[byvar]])
    pval <- ttest[[3]]
  }

  colnames(temp) <- c('var', 'mean', 'sd', 'N')
  temp['pct'] <- round(temp$N/sum(temp$N)*100, 1)
  temp['N'] <- paste0(temp$N, '(', temp$pct, ')')
  temp <- temp %>%
    select(-pct) %>%
    add_row(var = varname, mean = NA, sd=NA, N=NA, .before = 1) %>%
    mutate(p_val = ifelse(var==varname, pval, NA))
  return (temp)
}

# pick a df and run testing for all categorical variables
test_table <- function(df, uniquerx=F) {
  test_tbl <- data.frame()
  varlist <- c('sex', 'marriage', 'edulvl', 'race','hispyn', 'immigr',
                'povcat', 'usualpl', 'regionmeps', 'instype','health.status',
'anylmt', 'diabeticev')
  if (uniquerx==T){
    varlist <- c(varlist, 'tc1')
  }
  for (var in varlist) {
    temp <- test(df, var)
    test_tbl <- rbind(test_tbl, temp)
  }

  # format the table cells
  test_tbl$mean.sd <- paste0(format(round(test_tbl$mean, 2), nsmall = 2),
                   '(', format(round(test_tbl$sd, 2), nsmall = 2),')')
  test_tbl$mean.sd[test_tbl$mean.sd=="  NA(  NA)"] <- NA
  test_tbl$p_value <- format(round(test_tbl$p_val, 3), nsmall = 3)
  test_tbl$p_value[is.na(test_tbl$p_val)]=NA
  test_tbl$p_value[test_tbl$p_val<=0.001 & !is.na(test_tbl$p_val)] <-
'<0.001***'
  test_tbl$p_value[test_tbl$p_val<=0.01 & test_tbl$p_val>0.001
& !is.na(test_tbl$p_val)] <-
    paste0(test_tbl$p_value[test_tbl$p_val<=0.01 & test_tbl$p_val>0.001
& !is.na(test_tbl$p_val)], "**")
  test_tbl$p_value[test_tbl$p_val<=0.1 & test_tbl$p_val>0.01
& !is.na(test_tbl$p_val)] <-
    paste0(test_tbl$p_value[test_tbl$p_val<=0.1 & test_tbl$p_val>0.01
& !is.na(test_tbl$p_val)], "*")
  return (test_tbl)
}
```

Descriptive tables comparing logRxExp by groups

```
tbl.2yr17 <- test_table(meps_adt2yr[meps_adt2yr$year==2017,])
tbl.2yr17['id'] <- seq(1, nrow(tbl.2yr17),1)
```

```
tbl.2yr18 <- test_table(meps_adt2yr[meps_adt2yr$year==2018,])
tbl.2yr18['id'] <- seq(1, nrow(tbl.2yr18),1)
tbl.rx17 <- test_table(uniquerx[uniquerx$year==2017,], T)
tbl.rx17['id'] <- seq(1, nrow(tbl.rx17),1)
tbl.rx18 <- test_table(uniquerx[uniquerx$year==2018,], T)
tbl.rx18['id'] <- seq(1, nrow(tbl.rx18),1)

# merge four outputs together
keep_var <- c('var', 'mean.sd', 'N', 'p_value', 'id')
tbl.2yr <-merge(tbl.2yr17[,keep_var], tbl.2yr18[,keep_var],
             by=c('id', 'var'), all=T,
             suffixes = c(".2yr17",".2yr18"), sort=F)
tbl.rx <-merge(tbl.rx17[,keep_var], tbl.rx18[,keep_var],
             by=c('id', 'var'), all=T,
             suffixes = c(".rx17",".rx18"), sort=F)
table_combined <- merge(tbl.2yr, tbl.rx,
             by=c('id', 'var'), all=T,sort=F)

write.xlsx(table_combined,
           "descriptive_table.xlsx", col.names=T, row.names=F, overwrite =
TRUE)
```

Summary statistics of continuous explanatory variables

```
# rx count and oop share by insurance type
time.series <- function(input) {
  agg_ins <- input %>%
    group_by(year,instype) %>%
    summarise(mean_costperfill= mean(rxexptot_adj/rxprmedsno),
            sd_costperfill= sd(rxexptot_adj/rxprmedsno),
            med_costperfill= median(rxexptot_adj/rxprmedsno),
            q25 = quantile(rxexptot_adj/rxprmedsno, 0.25),
            q75 = quantile(rxexptot_adj/rxprmedsno, 0.75),
            mean_rxn= mean(rxprmedsno),
            sd_rxn= sd(rxprmedsno),
            mean_oop= mean(oopshare, na.rm =T),
            sd_oop= sd(oopshare, na.rm =T))
  agg_ins['Rx Cost/fill(mean)'] <- paste0(
    round(agg_ins$mean_costperfill,2), "(", round(agg_ins$sd_costperfill,2),
")")
  agg_ins['Rx Cost/fill(median)'] <- paste0(
    round(agg_ins$med_costperfill,2), "[", round(agg_ins$q25,2),", ",
round(agg_ins$q75,2), "]")
  agg_ins['Rx Count'] <- paste0(
    round(agg_ins$mean_rxn,2), "(", round(agg_ins$sd_rxn,2), ")")
  agg_ins['OOP share'] <- paste0(
    round(agg_ins$mean_oop,2), "(", round(agg_ins$sd_oop,2), ")")
  return(agg_ins)
}
byins.2yr <- time.series(meps_adt2yr)
byins.rx <- time.series(uniquerx)

wb <- createWorkbook()
addWorksheet(wb, 'summary.stats')
addWorksheet(wb, 'byins.2yr')
addWorksheet(wb, 'byins.rx')
```

```
# output our model result into the worksheet
writeData(wb, 1, table_combined, rowNames = F, colNames = T)
writeData(wb, 2, byins.2yr, rowNames = F, colNames = T)
writeData(wb, 3, byins.rx, rowNames = F, colNames = T)

# save worksheet
saveWorkbook(wb, "descriptive_table.xlsx", overwrite = TRUE)
par(mfrow=c(2,2), tcl=-0.5, family="serif", mai=c(0.3,0.3,0.5,0.3))
temp <- meps_adt2yr[meps_adt2yr$year==2017, ]
plot(temp$instype, log(temp$RxCostPerfill), main="log Rx Cost Per fill,
$ (2017) \ntwo-year longitudinal data",
     xlab='', ylab='',cex.main=1.5,cex.axis=1.1)
temp <- meps_adt2yr[meps_adt2yr$year==2018, ]
plot(temp$instype, log(temp$RxCostPerfill), main='log Rx Cost Per fill,
$ (2018) \ntwo-year longitudinal data',
     xlab='', ylab='',cex.main=1.5,cex.axis=1.1)
temp <- uniquerx[uniquerx$year==2017, ]
plot(temp$instype, log(temp$RxCostPerfill), main="log Rx Cost Per fill,
$ (2017) \ncohort with unique prescription drug class",
     xlab='', ylab='',cex.main=1.5,cex.axis=1.1)
temp <- uniquerx[uniquerx$year==2018, ]
plot(temp$instype, log(temp$RxCostPerfill), main="log Rx Cost Per fill,
$ (2018) \ncohort with unique prescription drug class",
     xlab='', ylab='',cex.main=1.5,cex.axis=1.1)
par(mfrow=c(1,1), mai=c(1.02,0.82,1.02,0.82))
```

Scatter plot of logRxExp vs. continuous variables

```
par(mfrow=c(2,2), tcl=-0.5, family="serif", mai=c(0.3,0.3,0.5,0.3))
temp <- meps_adt2yr[meps_adt2yr$year==2017, ]
plot(temp$logrxprmedsno, temp$logRxExp, main="log Rx Expenditure vs. log Rx
counts (2017) \ntwo-year longitudinal data",
     cex.main=1.5,cex.axis=1.1)
temp <- meps_adt2yr[meps_adt2yr$year==2018, ]
plot(temp$logrxprmedsno, temp$logRxExp, main='log Rx Expenditure vs. log Rx
counts (2018) \ntwo-year longitudinal data',
     cex.main=1.5,cex.axis=1.1)
temp <- uniquerx[uniquerx$year==2017, ]
plot(temp$logrxprmedsno, temp$logRxExp,main="log Rx Expenditure vs. log Rx
counts (2017) \ncohort with unique prescription drug class",
     cex.main=1.5,cex.axis=1.1)
temp <- uniquerx[uniquerx$year==2018, ]
plot(temp$logrxprmedsno, temp$logRxExp,main="log Rx Expenditure vs. log Rx
counts (2018) \ncohort with unique prescription drug class",
     cex.main=1.5,cex.axis=1.1)
par(mfrow=c(1,1), mai=c(1.02,0.82,1.02,0.82))
```

Running the regression

```
############################################################
#####Use data of all prescriptions to run the regression#####
gols <-
lm(logRxExp~year+instype+year*instype+edulvl+race+hispyn+immigr+famsize+
               povcat+logrxprmedsno+regionmeps+anylmt+diabeticev,
data=meps_adt)
# coeftest(gols, vcov = vcovHC(gols, type = "HC1"))
robust.gols <- sqrt(diag(vcovHC(gols, type="HC1")))
```

```
pols <-
plm(logRxExp~year+instype+year*instype+edulvl+race+hispyn+immigr+famsize+
               povcat+logrxprmedsno+regionmeps+anylmt+diabeticev,
            data=meps_adt2yr,model="pool")
#coeftest(pols, vcov = vcovHC(pols, type = "HC1"))
robust.pols <- sqrt(diag(vcovHC(pols, type="HC1")))

re <-
plm(logRxExp~year+instype+year*instype+edulvl+race+hispyn+immigr+famsize+
               povcat+logrxprmedsno+regionmeps+anylmt+diabeticev,
data=meps_adt2yr,model="random")
summary(re)
# coeftest(re, vcov = vcovHC(re, type = "HC1"))
robust.re <- sqrt(diag(vcovHC(re, type="HC1")))

# models using the Unique Rx data
rxols18 <-
lm(logRxExp~instype+tc1+edulvl+immigr+famsize+logrxprmedsno+regionmeps+anylmt
, data=uniquerx[uniquerx$year==2018,])
robust.rxols18 <- sqrt(diag(vcovHC(rxols18, type="HC1")))
# summary(rxols18)
rxols1718 <-
lm(logRxExp~year+instype+year*instype+tc1+edulvl+immigr+famsize+logrxprmedsno
+regionmeps+anylmt, data=uniquerx)
robust.rxols1718 <- sqrt(diag(vcovHC(rxols1718, type="HC1")))

stargazer(gols, pols, re, rxols18, rxols1718, type="html",
          se = list(robust.gols,robust.pols, robust.re, robust.rxols18,
robust.rxols1718),
          dep.var.labels = "model for all prescription data",
          covariate.labels = c("Year2017", "Medicare PartD (ref: Medicaid
Only)", "Medicare+", "PHI","Medicaid+","Uninsured", "Cardiovascular (ref:
Anti-infectives)", "Central Nervous System Agents", "Hormones/Hormone
Modifiers", "Respiratory Agents", "Topical Agents", "SomeCollege/above (ref:
Below some college)","OtherDegree/Unknown","Black/African American (ref:
White)", "Other","Hispanic (ref: Non-Hispanic)","Immigrants/Unknown (ref:
Native)","Family size",">=200% FPL (ref: <200% FPL)", "Log Rx Count","Midwest
(ref: East)","South","West", "Has Limitation(ref:No
limitation)","Unknown","Ever had Diabetes (ref: No diabetes)",
"Year2017:MedicarePartD",
"Year2017:Medicare+","Year2017:PHI","Year2017:Medicaid+",
"Year2017:Uninsured", "Constant"),
          column.labels = c("General OLS", "POLS", "Random Effects",
"OLS_Rx18", "OLS_Rx1718"),
          title = "Table 3: Regression Results",
          digits = 2,
          model.numbers = F,
          font.size = "small",
          align = TRUE,
          no.space = TRUE,
          single.row = FALSE, out="model.htm")
```

Visualization of the coefficients about health plan spending premium

```
# diagnosis
dig <- function(model) {
```

```
  vif <- vif(model)
  summary_model <- summary(model)
  par(mfrow=c(1,2), mar=c(2,2,2,1),tck=-0.04)
  plot(model, which=2)
  plot(model, which=1)
  par(mfrow=c(1,1), mai=c(1.02,0.82,1.02,0.82))
  hats=hatvalues(model)
  summary_hats <- summary(hats)
  return_obj <- list(summary_hats, vif, summary_model)
  names(return_obj)<-c('summary.hats', 'VIF','summary.model')

  return(return_obj)
}

create_output <- function(model, hc_se=T) {
  s <- summary(model)$coefficient
  # try to get hetero SE using sandwich
  se_hc <- data.frame(sqrt(diag(vcovHC(model, type="HC1"))))
  s <- merge(s, se_hc,by=0, all=T, sort=F)
  colnames(s) <- c('var', "coef", "se", "tval", "pval", "se_hc")
  if (hc_se==T) {
    colnames(s) <- c('var', "coef", "se_raw", "tval", "pval", "se")
  }
  s$change <- exp(s$coef)-1
  s$changese <- exp(s$coef)*s$se
  # for the coefficient of logRxcount and Rxcount, RxCount^2, keep original
coefficient
  if ("logrxprmedsno" %in% rownames(s)){
      s$change <- s$coef
      s$changese <- s$se}
  s$changelcl <- s$change-1.96*s$changese
  s$changeucl <- s$change+1.96*s$changese
  s$est_ci <- paste0(format(round(s$change, 2), nsmall = 2),
                     '[', format(round(s$changelcl, 2), nsmall = 2),
                     ' ',format(round(s$changeucl, 2), nsmall = 2), ']')
  s$est_se <- paste0(format(round(s$change, 2), nsmall = 2),
                     '(', format(round(s$changese, 2), nsmall = 2),')')
  s$p_value <- format(round(s$pval, 3), nsmall = 3)
  s$p_value[s$pval<=0.01] <- '<0.01***'
  s$p_value[s$pval<=0.05 & s$pval>0.01] <-
    paste0(s$p_value[s$pval<=0.05 & s$pval>0.01], "**")
  s$p_value[s$pval<=0.1 & s$pval>0.05] <-
    paste0(s$p_value[s$pval<=0.1 & s$pval>0.05], "*")

  return (s)
}

create_outputyr <- function(model, hc_se=T) {
  s <- data.frame(summary(model)$coefficient)
  # try to get hetero SE using sandwich
  se_hc <- sqrt(diag(vcovHC(model, type="HC1")))
  s <- merge(s, se_hc,by=0, all=T, sort=F)
  colnames(s) <- c('var', "coef", "se", "tval", "pval", "se_hc")
  if (hc_se==T) {
    colnames(s) <- c('var', "coef", "se_raw", "tval", "pval", "se")
  }
  s['pval_raw'] <- s$pval
```

```
  # get the covariance matrix to calculate the se(beta_1+beta_2)
  V <- vcov(model)
  if (hc_se==T) {
    V <- vcovHC(model, type="HC1")
  }

  # update the coefficient for the year and instype interaction, to
  # represent the spending premium (from Medicaid) in each year
  for (var in c("instypeMedicarePartD", "instypeMedicare+", "instypePHI",
                "instypeMedicaid+", "instypeUninsured")){
    for (yr in c(2017)) {
      var_to_update <- paste0("year",yr,":", var)
      # update the coefficient
      s[s$var==var_to_update, 'coef'] <-
        s[s$var==var_to_update, 'coef']+s[s$var==var, 'coef']
      # !!IMPORTANT: update the se for the interaction
      related_var <- c(var, var_to_update)
      cov_matrix <- V[related_var, related_var]
      s[s$var==var_to_update, 'se'] <- sqrt(t(c(1,1))%*%cov_matrix%*%c(1,1))

      # !! update the p-value
      # https://www.cyclismo.org/tutorial/R/pValues.html
      s[s$var==var_to_update, 'pval'] <- 2*pt(
        -abs(s[s$var==var_to_update, 'coef']/s[s$var==var_to_update, 'se']),
        df.residual(model)) # the degree of freedom
    }
  }

  s$change <- exp(s$coef)-1
  s$changese <- exp(s$coef)*s$se
  # for the coefficient of logRxcount and Rxcount, RxCount^2, keep original
coefficient
  if ('logrxprmedsno' %in% rownames(s)){
      s$change <- s$coef
      s$changese <- s$se
    }
  s$changelcl <- s$change-1.96*s$changese
  s$changeucl <- s$change+1.96*s$changese
  s$est_ci <- paste0(format(round(s$change, 2), nsmall = 2),
                     '[', format(round(s$changelcl, 2), nsmall = 2),
                     ' ',format(round(s$changeucl, 2), nsmall = 2), ']')
  s$est_se <- paste0(format(round(s$change, 2), nsmall = 2),
                     '(', format(round(s$changese, 2), nsmall = 2),')')
  s$p_value <- format(round(s$pval, 3), nsmall = 3)
  s$p_value[s$pval<=0.01] <- '<0.01***'
  s$p_value[s$pval<=0.05 & s$pval>0.01] <-
    paste0(s$p_value[s$pval<=0.05 & s$pval>0.01], "**")
  s$p_value[s$pval<=0.1 & s$pval>0.05] <-
    paste0(s$p_value[s$pval<=0.1 & s$pval>0.05], "*")

  return (s)
}

wb <- createWorkbook()
addWorksheet(wb, 'OLS1')
addWorksheet(wb, 'POLS')
addWorksheet(wb, 'RE')
```

```
addWorksheet(wb, 'OLS2')
addWorksheet(wb, 'OLS3')

# output our model result into the worksheet
writeData(wb, 1, create_outputyr(gols), rowNames = F, colNames = T)
writeData(wb, 2, create_outputyr(pols), rowNames = F, colNames = T)
writeData(wb, 3, create_outputyr(re), rowNames = F, colNames = T)
writeData(wb, 4, create_output(rxols18), rowNames = F, colNames = T)
writeData(wb, 5, create_outputyr(rxols1718), rowNames = F, colNames = T)

# save worksheet
saveWorkbook(wb, "regression_summary.xlsx", overwrite = TRUE)
# make the plot of the coefficient
# https://cran.r-project.org/web/packages/dotwhisker/vignettes/dotwhisker-
vignette.html
# function to create the table for plot
dwplot_tbl <- function(modelname){
  model <- eval(sym(modelname))
  keep_row <- c("instypeMedicarePartD", "instypeMedicare+", "instypePHI",
             "instypeMedicaid+", "instypeUninsured")
  keep_column <- c("var","change", "changese")
  temp=create_output(model) %>%
    filter(var %in% keep_row) %>%
    select(keep_column)
  colnames(temp) <- c("term","estimate", "std.error")
  temp$term <- sub("^instype", "", temp$term)
  temp$model<- modelname
  return (temp)
}

OLS1 <- gols
POLS <- pols
RE <- re
OLS2 <- rxols18
OLS3 <- rxols1718

tbl <- data.frame()
for (m in c("OLS1", "POLS", "RE", "OLS2", "OLS3")) {
  temp <- dwplot_tbl(m)
  temp$estimate <- temp$estimate*100
  temp$std.error <- temp$std.error*100
  tbl <- rbind(tbl,temp)
}

#### make the plots for model 1-8
dwplot(tbl,
       vline = geom_vline(xintercept = 0, colour = "grey60", linetype = 2), #
plot line at zero _behind_ coefs
       dot_args = list(aes(shape = model), size=2),
       whisker_args = list(size=0.8)
       ) +
    theme_bw() + xlab("spending premium (%) compared with Medicaid Only") +
ylab("") +
    ggtitle("Health plan effect on Rx spending (2018)") +
    theme(plot.title = element_text(size=15, face="bold"),
          legend.position = c(0.007, 0.54),
          legend.text=element_text(size=13),
```

```
        legend.key.size = unit(0.8, "cm"),
        legend.justification = c(0, 0),
        legend.background = element_rect(colour="grey80"),
        legend.title=element_blank(),
        axis.text = element_text(color="black", size=14),
        axis.title = element_text(color="black", size=15))
```