Paper 126-2010

Investigating the Impact of Medicare, Part D on the Diabetes Medications Using Enterprise Miner and Survival Analysis

Xiao Wang, Department of Mathematics, University of Louisville, KY

ABSTRACT

The purpose of this paper is to examine the influence of Medicare, Part D on diabetes medications since its introduction in 2006, with the prescribed medicines data from the Medical Expenditure Panel Survey collected by the Agency for Healthcare Research and Quality.

In this paper, Summary Statistics, Kernel Density Estimation and Survival Analysis in Base SAS® and SAS/STAT® were used. The association analysis and the text mining analysis in Enterprise Miner™ were also utilized. Before that, SAS SQL, some SAS data step procedures or functions such as the transpose procedure were used to preprocess the data. Summary Statistics show that Medicare, part D increases the average Medicare payment for diabetes medications from $20 in 2005 to $130 in 2006. The kernel density estimation demonstrates that in 2005, the male beneficiaries pay more for the drugs, metformin and glyburide, while the female beneficiaries spend more on insulin and supplies. In 2006, the costs of the diabetes supplies and metformin remain higher than the other costs. The association analysis provides such information that the Medicare enrollees have fewer choices of their drugs than those with private insurance. Survival analysis indicates that insulin and metformin are more stable in terms of usage than the other drugs in general, while glyburide itself is very unstable in 2005. Moreover, the diabetes medications in 2005 are more stable than those in 2006.

INTRODUCTION

MEDICARE AND MEPS

As part of the Medicare reform, legislators have agreed to support legislation that will help close the coverage gap in the Medicare prescription drug program (Part D). We estimated the prescribed diabetes drugs for the beneficiaries in the U.S. prior to and after the implementation of Medicare Prescription Drug Coverage, or Medicare part D.

Medicare is the health insurance for people of age 65 or older, under age 65 with certain disabilities, and any age with permanent kidney failure (called “End-Stage Renal Disease”). It mainly consists of four parts, Part A (Hospital Insurance), Part B (Medical Insurance), Part C (Medicare Advantage Plans) and Part D (The optional prescription drug program).Part D uses competing private plans to provide beneficiaries access to appropriate drug therapies. As of January 2008, almost 90 percent of Medicare beneficiaries were enrolled in the Part D plan, or had other sources of creditable drug coverage.

The Medical Expenditure Panel Survey (MEPS), collected by the Agency for Healthcare Research and Quality, is a set of large-scale surveys of families and individuals, their medical providers and employers across the United States. Every year, MEPS covers such information as the usage of medical services, sources of payments, and health insurance cost. However, such data have some disadvantages; for instance, time information is incomplete.

DIABETES MEDICATIONS

Diabetes medications refer to both the oral medicines and the supplies that are related. However, we mainly focus on the former in this paper. The sulfonylureas such as glyburide, glipizide, and glimepiride are all generic medications, which make them inexpensive drugs for the management of Type II diabetes mellitus. Another class of medication that has a similar mechanism of action to the sulfonylurea medications is the meglintinides, including repaglinide (prandin) and neteglinide (starlix). Another drug class is alpha glucosidase inhibitors, such as precose. Metformin, the only biguanide is recommended as a mainstay in therapy in patients with type II diabetes. When the above medications become ineffective, insulin should be used alone or combined with the other drugs.

METHODS

In statistics, kernel density estimation is a non-parametric way of estimating the probability density function of a random variable. Here, we use the univariate case. The kernel estimator for the univariate can be defined as follows:

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \qquad (1)$$

where $X_1, X_2, \ldots X_n$ are independent and identically distributed random variables, f is the density function, K stands for some known density function and h is a smoothing parameter called the bandwidth. As an illustration, given some data about a sample of a population, kernel density estimation makes it possible to extrapolate the data to the entire population. To better display the distribution, it is important to make sure that the smoothness of the graph is reasonable, which is decided by the bandwidth. One simple method to choose the optimal bandwidth is to find the minimum of the asymptotic mean integrated squared error (AMISE, defined below) with respect to h,

$$\text{AMISE} = \frac{R(K)}{nh} + \frac{1}{4}\sigma_k^4 h^4 R(f''(x)) \tag{2}$$

where $R(K)= \int K^2(x)dx$.

Cluster analysis in the text mining analysis (first used by Tryon, 1939) encompasses a number of different algorithms and methods for grouping objects of a similar kind into respective categories. In this paper, we chose the default algorithm in Text Miner, expectation maximization, a very general iterative algorithm for parameter estimation by maximum likelihood when some of the random variables involved are not observed. The algorithm consists of two steps, E-step (estimation) and M-step (maximization).

E-step: to compute the conditional expectation:

$$Q(\theta;\theta^{(t)}) = E_{\theta^{(t)}}(\log L(\theta, y)\,|\, y_{obs}) \tag{3}$$

M-step: to find the θ , which maximizes the expectation:

$$\hat{\theta}^{(t+1)} = \arg\max_{\theta} Q(\theta\,|\,\hat{\theta}^{(t)}) \tag{4}$$

These two steps are repeated until

$$\left|L(\theta^{(t+1)}) - L(\theta^{(t)})\right| < \varepsilon, \tag{5}$$

where $L(\theta, y)$ is a likelihood function, $Q(\theta,\theta^{(t)})$ is a conditional expectation, θ is an unknown parameter and $\hat{\theta}^{(t)}$ is the estimate of the unknown parameter at iteration t >0, and ε is a very small positive number. In most cases, cluster analysis is only a useful starting point for other purposes. We used it for kernel density estimation of the Medicare payments.

All of the standard approaches to survival analysis are probabilistic. One way of describing it is the survival function, denoted by $S(t)$, defined as the probability that an individual survives longer than t:

$$S(t) = P(T > t) \tag{6}$$

From the definition of the cumulative distribution function $F(t)$ of T,

$$S(t) = 1 - F(t) \tag{7}$$

If the event of interest is a medicine switch, then the survivor function gives the probability that the drug remains to be used beyond time t. In this paper, due to a large amount of data, we preferred the life table method, in which the survival estimate is obtained by calculating the conditional probabilities of surviving beyond time t, defined by

$$S(t_i) = \prod_{i=1}^{n-1}(1-q_i) = \prod_{i=1}^{n-1} p_i \tag{8}$$

For each time interval i, t_i is the start time, q_i is the conditional probability of failure and p_i is the conditional probability of surviving to t_i or beyond that time.

METHOD

In this paper, we used four different data sets for the prescribed medications information and the demographic information for the year 2005 and the year 2006. Additionally, we utilized the office-based visit and outpatient visit files for the year 2005. All of them were taken from the Medical Expenditure Panel Survey collected by AHRQ (Agency for Healthcare Research and Quality).The following variables were used.

DUPERSID	PERSON ID
RXNAME	DRUG NAME
RXMR05(06)X	AMOUNT PAID, MEDICARE (IMPUTED) IN 2005(2006)
RXXP05(06)X	SUM OF PAYMENTS RXSF06X-RXOU06X(IMPUTED)
RXBEGD	DAY PERSON STARTED TAKING MEDICINE
RXBEGMM	MONTH PERSON STARTED TAKING MEDICINE
RXBEGYRX	YEAR PERSON STARTED TAKING MEDICINE
ICD9	INTERNATIONAL CLASSFICATION OF DISEASE, 9TH EDITION

In the study, we mainly demonstrated using the data for the year 2006. First, we filtered out the records for the diabetes medications and then we consolidated the drug names. Next, we used Summary Statistics to get the average Medicare payment and the average total payment. For comparison, the average Medicare payment in 2005 is expressed in 2006 dollars with 2005 data inflated based on the CPI-U (Consumer Price Index for all Urban Consumers) for prescription drugs. That is to say, once we got the mean values for the year 2005, we multiplied them by the index 1.043.

Year	Variable	Mean	N
2005*	SUM OF PAYMENTS MEDICARE (IMPUTED)	501.66 20.58	1759 1759
2006	SUM OF PAYMENTS MEDICARE (IMPUTED)	558.11 129.02	1994 1994

Table 1. Average overall payment and Medicare payment in 2005 & 2006

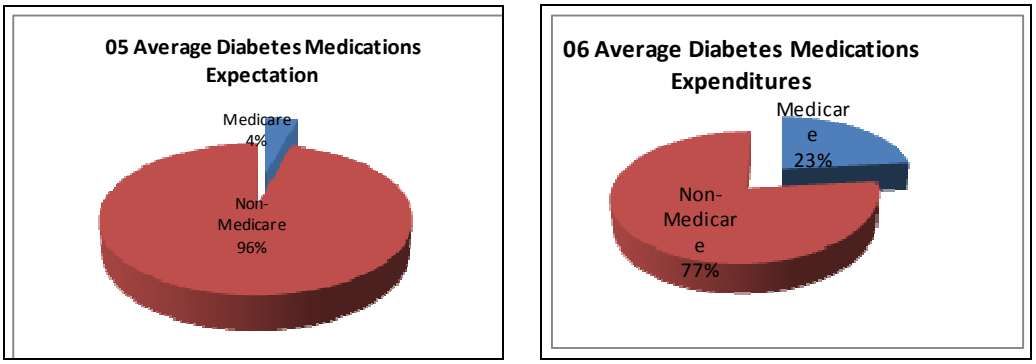


Figure1. Pie charts of payments in 2005 &2006

Table 1 and Figure 1 show that the average of the total payments for the prescription increases of approximately 12 percent from the year 2005 to the year 2006, while the average Medicare payment in 2006 is 6 times as much as that in 2005. The ratio of the average Medicare payment to the total expenditures also increases from 4 per cent to 23 per cent. Results demonstrate that the plan, Part D, indeed increases the Medicare drug expenditures.

Next, we want to see how the Medicare payments distribute among different groups of drugs using Text Miner in SAS Enterprise Miner and kernel density estimation. We compared the Medicare payments in two cases, the general case and the Medicare case in which the beneficiaries join Medicare. We also needed to preprocess the data sets. First, we converted the variable, NRXNAME, into observations by the transpose procedure, the trim function, the translate function and the concatenation operator. The SAS code is shown below.

```
PROC SORT DATA=SASUSER.SMED06 OUT=SASUSER.SORTMED06;
BY DUPERSID NRXNAME; RUN; OPTIONS OBS=MAX;
DATA SASUSER.SORTMR06; SET SASUSER.SORTMED06;
NRXNAME=TRANSLATE(LEFT(TRIM(NRXNAME)),'_',' '); RUN;
PROC TRANSPOSE DATA=SASUSER.SORTMR06 OUT=SASUSER.TRANMR06
PREFIX=MED_; VAR NRXNAME; BY DUPERSID; RUN;
DATA SASUSER.CONMR06 (KEEP= DUPERSID SSNRXNAME);
LENGTH SSNRXNAME $ 32767; SET SASUSER.TRANMR06;
ARRAY CONCAT MED_:; SSNRXNAME =LEFT(TRIM(MED_1));
DO I=2 TO DIM(CONCAT); SSNRXNAME=LEFT(TRIM(SSNRXNAME)) ||''||
LEFT (TRIM(CONCAT[I])); END; RUN;
```

GENERAL CASE

First, we input the new data into Text Miner, changing the default setting of Different Parts of Speech and Noun Groups from Yes to No, setting the number of clusters to 10 and using the default cluster algorithm. The algorithm returned 5 clusters in 2005 shown in figure 2 and 6 clusters in 2006 displayed in figure 3. We merged the two new data sets by clusters to do kernel density estimation on the Medicare payments by cluster. We also converted the payments in 2005 into 2006 dollars by CPI-U. We used the following SAS code.

```
/* Sort and Merge the data*/
DATA SASUSER.DOCUMENT06(DROP=_DOCUMENT_ _SVD_1-_SVD_10 PROB1-PROB7);
SET EMWS1.TEXT2_DOCUMENTS;RUN;
DATA SASUSER.CLUSTER06(DROP=_SVD_1-_SVD_4);
SET EMWS1.TEXT2_CLUSTER;RUN;
PROC SORT DATA=SASUSER.DOCUMENT06;BY _CLUSTER_;
PROC SORT DATA=SASUSER.CLUSTER06;BY _CLUSTER_;
DATA SASUSER.MERGEDOCCLU06;
MERGE SASUSER.DOCUMENT06 SASUSER.CLUSTER06;BY _CLUSTER_;RUN;
/*Kernel density estimation*/
PROC SORT DATA=SASUSER.MERGEDOCCLU06
OUT=SASUSER.SORT06;BY SEX  _CLUSTER_ ;RUN;
PROC KDE DATA=SASUSER.SORT06; UNIVAR RXMR06X_SUM /GRIDL=0 GRIDU=500 METHOD=SNR
OUT=SASUSER.KDE06; BY SEX _CLUSTER_ ; RUN;
```

Clusters

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	glyburide, rosiglitazone, precose, metformin, glyburide-metformin	271	0.154328018...	0.0711406...
2	glyburide-metformin, + supply, starlix, precose, prandin	493	0.280751708...	0.1198462...
3	insulin, pioglitazone, + supply, starlix, prandin	237	0.134965831...	0.0374267...
4	metformin, glimepiride, tolazamide, prandin, rosiglitazone	494	0.281321184...	0.0838507...
5	glipizide, rosiglitazone, precose, starlix, metformin	261	0.148633257...	0.0693840...

Figure 2. Clusters of drugs in 2005

Clusters

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	glyburide-metformin, glimepiride, starlix, insulin, prandin	214	0.108961303...	0.2865909...
2	+ supply, glimepiride	358	0.182281059...	0.0148972...
3	glyburide, rosiglitazone, precose, glyburide-metformin, metformin	272	0.138492871...	0.0554535...
4	insulin, + supply, starlix, precose, glyburide-metformin	228	0.116089613...	0.0484771...
5	tolazamide, metformin, glimepiride, prandin, precose	611	0.311099796...	0.0897675...
6	glipizide, rosiglitazone, pioglitazone, precose, metformin	281	0.143075356...	0.0510506...

Figure 3. Clusters of drugs in 2006

Figure 4 for the year 2005 demonstrates that most Medicare payments for diabetes medications are fewer than 200 dollars. For the males, most costs of the drugs are lower than 50 dollars. The only exception is cluster 1, which indicates that Medicare pays more for metformin, glyburide and their combination. For the female patients, the densities of clusters #2 and #4 are higher than those of the other clusters under 50 dollars; after that, the density of cluster 3 is the highest. Hence, the female patients spend more on insulin and supplies. In 2006, the ordering of the males' expenditures is cluster# 2 > #5> #6>#1>#3 under 120 dollars; after the threshold point, the densities for clusters #1 and #3 become higher than the others. Women spend much more on clusters #2 and # 5 of the drugs under 140 dollars, indicating that metformin and supplies cost the females more than the others do. Hence, most Medicare expenditures are on supplies, metformin and glyburide.

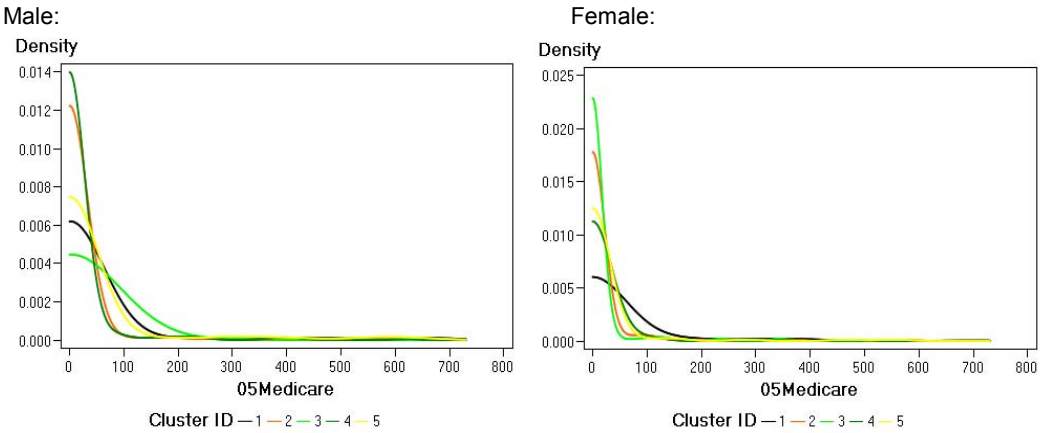


Figure 4. Kernel density estimation for Medicare in 2005

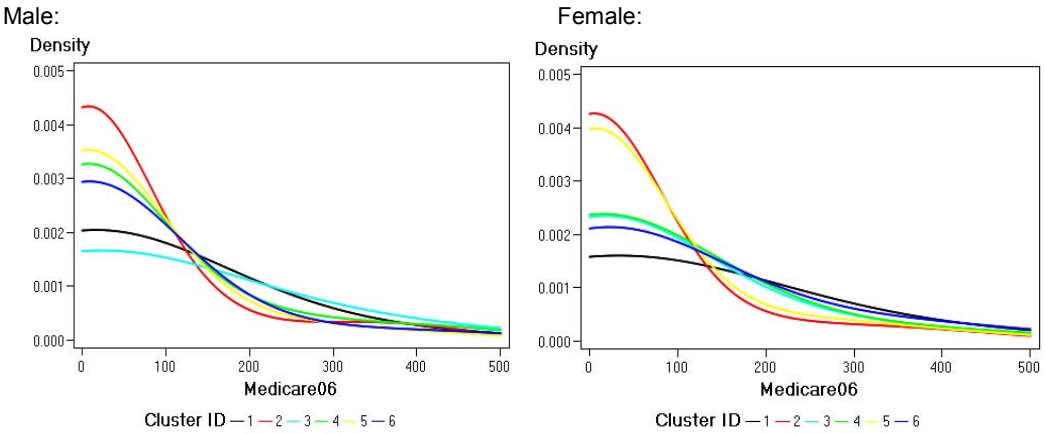


Figure 5. Kernel density estimation for Medicare in 2006

MEDICARE CASE

In this case, we filtered out the beneficiaries whose Medicare payments are greater than 0. Then, we repeated the same steps to get kernel density functions for Medicare payments.

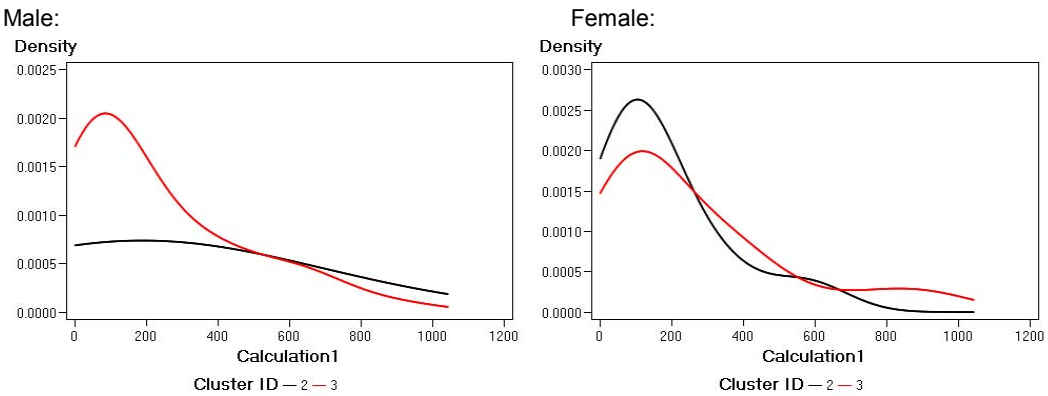


Figure 6. Kernel density estimation of 2005 Medicare

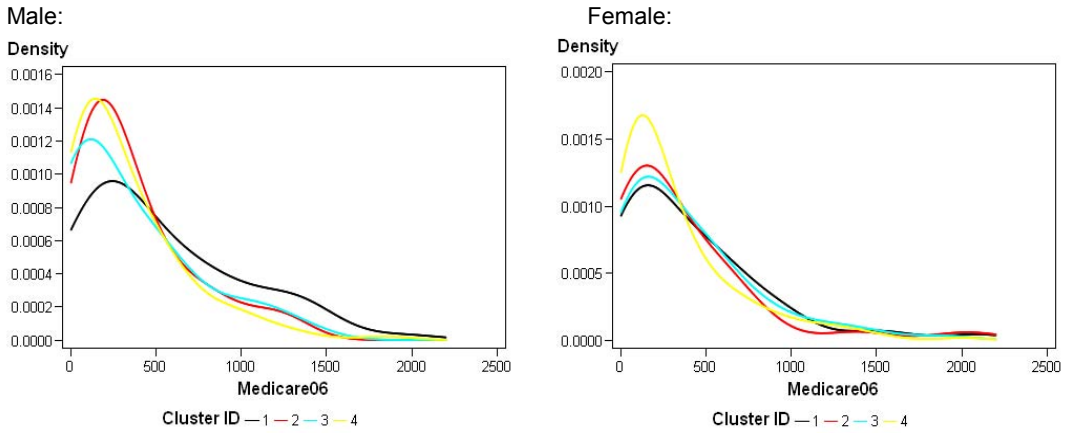
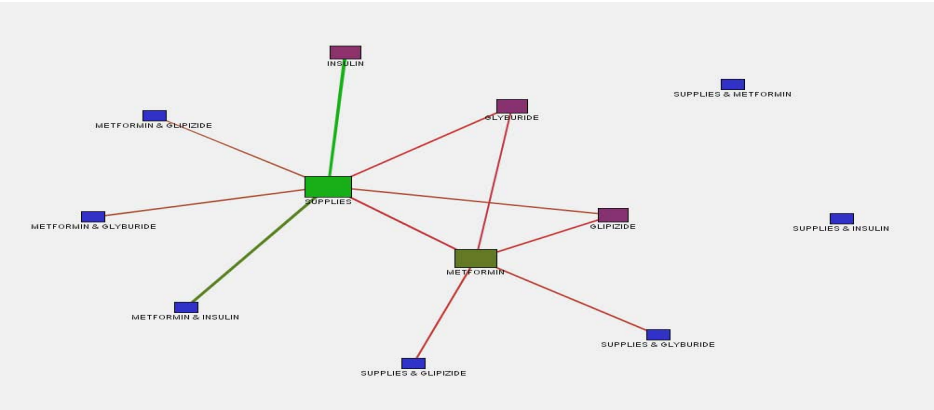
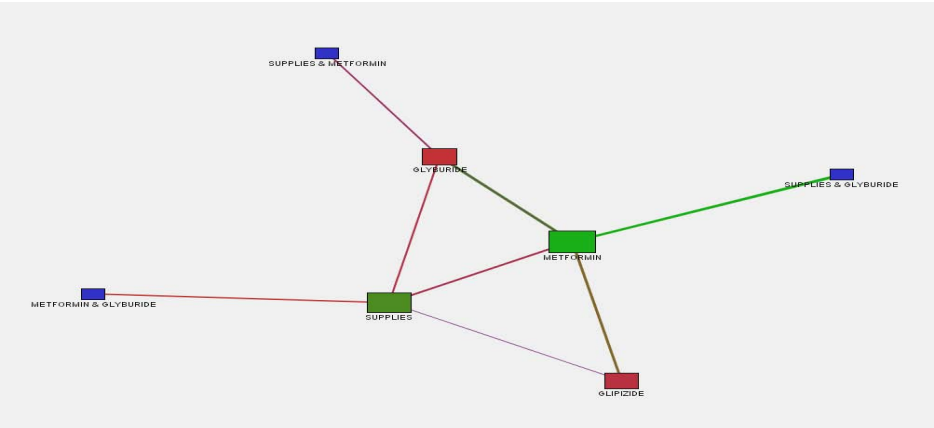


Figure 7. Kernel density estimation of 2006 Medicare

Figures 6 and 7 show that in 2006, with Part D in Medicare, the expenditures on drugs are greatly increased. The costs of the diabetes supplies, metformin and insulin remain higher than the other costs. Next, we performed market basket analysis.

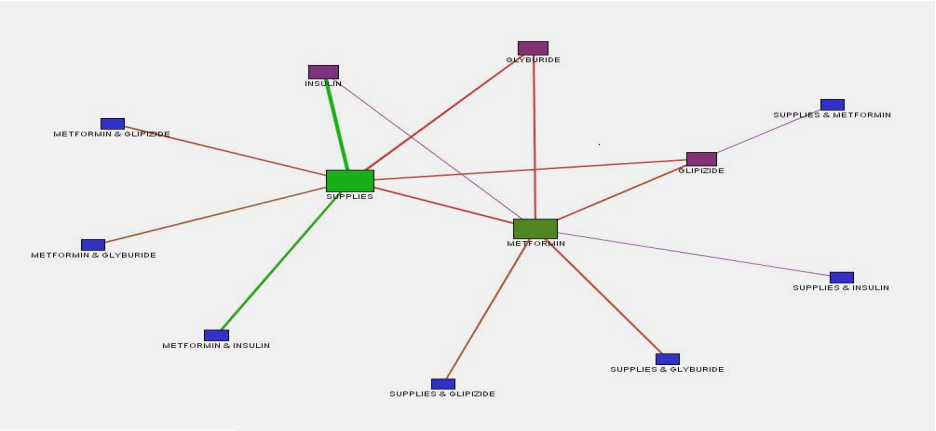


Output 1. General case

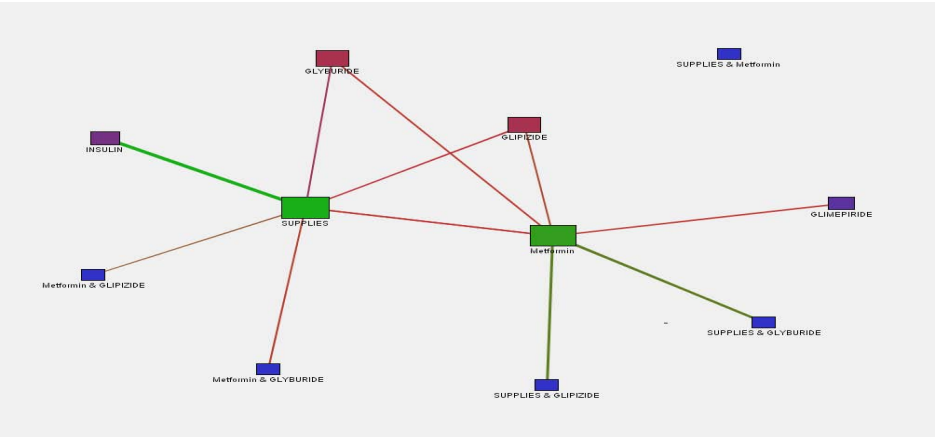


Output 2. Medicare case

Figure 8.Link graphs for the drugs in 2005



Output 1.General case



Output 2. Medicare case

Figure 9.Link graphs for the drugs in 2006

We used the association node in EM 5.2 to get the link graphs shown in Figures 8 and 9. We also discussed two cases, the general case and the Medicare case. In 2005 in general, the diabetes supplies and metformin are two centers of the graphs. Insulin has a strong relationship to the supplies, although it is not related to the other factors. Glyburide, and glipizide are also important. When we studied the beneficiaries who got their drugs paid by Medicare, metformin becomes more important and it is strongly related to glyburide and glipizide. In 2006, the general case remains almost the same except that the supplies are connected to the combination of insulin and metformin. In the Medicare case, there are fewer connections between different drugs. Output 2 in Figure 8 indicates that if insulin is prescribed, then the supplies are very likely to be prescribed, too. Also, once the combinations of the supplies with glyburide are used, then metformin will probably be utilized.

Finally, we performed survival analysis by the life test procedure. For a better comparison, we also needed physician visit information in 2005. For the year 2006, we processed the missing time information using the following SAS code and then sorted out the records for the year 2006 to get the data set shown in Figure 10. We set all the unknown values for the day to the 1st day of each month, incomplete month information to missing and all the missing years to 2006.

```
PROC SQL;
CREATE TABLE SASUSER.SMRS06 AS SELECT SRSMED06.DUPERSID,
SRSMED06.RNRXNAME,(( CASE WHEN -1 = SRSMED06.RXBEGDD THEN 1
WHEN -8 = SRSMED06.RXBEGDD THEN 1 WHEN -9 = SRSMED06.RXBEGDD THEN 1 ELSE
SRSMED06.RXBEGDD END)) AS RXDD,
((CASE WHEN -1 = SRSMED06.RXBEGMM THEN. WHEN -8 = SRSMED06.RXBEGMM THEN. WHEN -9 =
SRSMED06.RXBEGMM THEN. ELSE SRSMED06.RXBEGMM END)) AS RXMM,
```

```
((CASE WHEN -1 = SRSMED06.RXBEGYRX THEN 2006 WHEN -14 = SRSMED06.RXBEGYRX THEN 2006
WHEN -7 = SRSMED06.RXBEGYRX THEN 2006 WHEN -8 = SRSMED06.RXBEGYRX THEN 2006 WHEN -9
= SRSMED06.RXBEGYRX THEN 2006 ELSE SRSMED06.RXBEGYRX END)) AS RXYY
FROM SASUSER.SRSMED06 AS SRSMED06 WHERE CALCULATED RXYY = 2006; QUIT;
```

Next, we suppressed the data by removing the repeated information and the SAS code is as follows:

```
PROC SORT DATA=SASUSER.SMRS06 OUT=SASUSER.UNISMRS06 NODUPKEY;
BY DUPERSID SSNRXNAME; RUN;
```

	DUPERSID	SSNRXNAME	RXDD	RXMM	RXYY
1	30078019	METFORMIN	1	.	2006
2	30121012	GLIPIZIDE	1	.	2006
3	30149010	METFORMIN	1	.	2006
4	30177026	GLIPIZIDE	27	12	2006
5	30177026	METFORMIN	27	12	2006
6	30180024	GLYBURIDE	1	.	2006
7	30192012	METFORMIN	20	1	2006
8	30206025	GLIMEPIRIDE	1	.	2006
9	30217015	GLYBURIDE	1	2	2006
10	30300013	METFORMIN	30	8	2006

Figure 10. Diabetes medication in 2006

Next, we converted the date into a SAS date by using the MDY function, transposed the data by NRXNAME and DATE and finally merged the two new data sets to get the data displayed in Figure 11.The SAS code is shown below.

```
/*Change the date into SAS date*/
  DATA SASUSER.DATE06;SET SASUSER.SURMED06; IF RXMM=. THEN DATE=.;
  ELSE DATE=MDY(RXMM,RXDD,RXYY);RUN;
PROC SORT DATA=SASUSER.DATE06; BY DUPERSID; RUN;
/*Change the records to variables*/
PROC TRANSPOSE DATA=SASUSER.DATE06 OUT=SASUSER.NDATE06
PREFIX=MED_ NAME=VAR; BY DUPERSID; VAR SSNRXNAME; RUN;
PROC TRANSPOSE DATA=SASUSER.DATE06 OUT=SASUSER.N1DATE06
REFIX=DATE_ NAME=VAR; BY DUPERSID; VAR DATE; RUN;
/*Merge the new data*/
DATA SASUSER.MERGEDATA06;
MERGE SASUSER.NDATE06 SASUSER.N1DATE06; BY DUPERSID; RUN;
```

	DUPERSID	VAR	MED_1	MED_2	MED_3	DATE_1	DATE_2	DATE_3
1	30078019	DATE	METFORMIN			.	.	.
2	30121012	DATE	GLIPIZIDE			.	.	.
3	30149010	DATE	METFORMIN			.	.	.
4	30177026	DATE	GLIPIZIDE	METFORMIN		17162	17162	.
5	30180024	DATE	GLYBURIDE			.	.	.
6	30192012	DATE	METFORMIN			16821	.	.
7	30206025	DATE	GLIMEPIRIDE			.	.	.
8	30217015	DATE	GLYBURIDE			16833	.	.
9	30300013	DATE	METFORMIN			17043	.	.
10	30363015	DATE	GLIPIZIDE			.	.	.

Figure 11. Analysis data in 2006

Next, we searched for the first switching of the drugs and defined the variable, STATUS. During the analysis, we made some assumptions:

#1. When we use an array statement, we assume the missing date to be the end of the year 2006 and also we converted it into a SAS date.
#2. If the drug is continued during the survival time, then it is censored and the value of STATUS is 0; otherwise, the value of STATUS is 1.
#3.If CHMED is equal to the drug, it means the drug is switched to another drug; in other words, it is not censored.
#4.Due to a lack of information, we set the start date equal to the beginning of the year 2006 and the end

date to the end of the year 2006 if such information is unknown.
#5. We also define the value of STATUS as 0 when the survival time DAYS is equal to 364.
#6.We supposes the frequency of prescription for the year 2005 is at most 12.

The SAS code is shown below.

```
 DATA SASUSER.T06; SET SASUSER.MERGEDATA06;
 ARRAY MEDS(3) MED_1 - MED_3; ARRAY DATES(3) DATE_1 - DATE_3;
 DO J=1 TO 3; IF DATES(J)=. THEN DATE='31DEC2006'D; END;
   DO I=1 TO 3;
   IF I=1 THEN TEMP=MEDS(I);
   IF MEDS(I) NE TEMP THEN DO;
  MED_NUM=I; DATE_NUM=DATES(I);  CHMED=MEDS(I);
  STATUS=1; I=3;
  END;
END;
/*Define 0-1 indicators and status*/
IF CHMED=' ' THEN STATUS=1;
IF CHMED='GLYBURIDE' THEN GLYBURIDE=0 AND STATUS=1;ELSE GLYBURIDE=1;
IF CHMED='METFORMIN' THEN METFORMIN=0 AND STATUS=1;ELSE METFORMIN=1;
IF CHMED='STARLIX' THEN STARLIX=0 AND STATUS=1; ELSE STARLIX=1;
IF CHMED='PRECOSE' THEN PRECOSE=0 AND STATUS=1; ELSE PRECOSE=1;
IF CHMED='INSULIN' THEN INSULIN=0 AND STATUS=1; ELSE INSULIN=1;
/*Define the variables days*/
IF DATE_1^=. THEN SDATE=DATE_1; ELSE SDATE='01JAN2006'D;
IF DATE_2^=. THEN EDATE=DATE_2; ELSE EDATE='31DEC2006'D;
FORMAT SDATE EDATE DATE9; DAYS=DATDIF (SDATE,EDATE,'ACT/ACT');
IF DAYS=364 THEN STATUS=0; RUN;
```

Finally, we sorted the new data by CHMED to get the data shown in Figure 12.

	STATUS	GLYBURIDE	METFORMIN	STARLIX	PRECOSE	INSULIN	SDATE	EDATE	DAYS
1	0	1	1	1	1	1	01JAN2006	31DEC2006	364
2	0	1	1	1	1	1	01JAN2006	31DEC2006	364
3	0	1	1	1	1	1	01JAN2006	31DEC2006	364
4	1	1	0	1	1	1	27DEC2006	27DEC2006	0
5	0	1	1	1	1	1	01JAN2006	31DEC2006	364
6	1	1	1	1	1	1	20JAN2006	31DEC2006	345
7	0	1	1	1	1	1	01JAN2006	31DEC2006	364
8	1	1	1	1	1	1	01FEB2006	31DEC2006	333
9	1	1	1	1	1	1	30AUG2006	31DEC2006	123
10	0	1	1	1	1	1	01JAN2006	31DEC2006	364

Figure12. Survival data for 2006

In 2005, in order to get an accurate conclusion, we filtered out the beneficiaries by Medicare payments for physician visits rather than by the Medicare expenditures on drugs. We first sorted out the enrollees whose Medicare payments are greater than 0 according to the ICD 9 diagnosis code. The SAS code is shown below.

```
/*To sort out the diabetes patients*/
PROC SQL;
   CREATE TABLE SASUSER.OBDIA05 AS
   SELECT t1.DUPERSID, t1.OBICD1X, t1.OBICD2X, t1.OBICD3X, t1.OBICD4X
   FROM SASUSER.FILTER_FOR_QUERY_FOR_FILTER_FOR_ AS t1
WHERE t1.OBICD1X = '250' OR t1.OBICD2X = '250' OR t1.OBICD3X = '250'
OR t1.OBICD4X = '250'; QUIT;
/*To delete duplicates*/
PROC SORT DATA=SASUSER.OBDIA05 OUT=SASUSER.UNIOBDIA05 NODUPKEY;
BY DUPERSID; RUN;
```

We got the results shown in Figure 13, and we used the same method to get another data set about diabetes patients in the outpatient visit file.

	DUPERSID	OBICD1X	OBICD2X	OBICD3X	OBICD4X
1	30078019	250	-1	-1	-1
2	30121012	401	250	185	530
3	30180024	250	-1	-1	-1
4	30192012	250	-1	-1	-1
5	30201026	250	-1	-1	-1
6	30206025	250	401	-1	-1
7	30363015	401	250	272	716
8	30392041	250	-1	-1	-1
9	30450010	590	429	250	-1
10	30494013	401	250	716	530

Figure13. Diabetes patients in office-based visit

Finally, we used the SQL horizontal join to get all the diabetes beneficiaries. The code is as follows.

```
PROC SQL;
CREATE TABLE SASUSER.DIA05 AS
SELECT DUPERSID FROM SASUSER.UNIOBDIA05
UNION
SELECT DUPERSID FROM SASUSER.UNIOPDIA05;
QUIT;
```

Next, we used these patient ID and anti-diabetic drugs to find out all the Medicare beneficiaries with diabetes. After that, we repeated the same methods as we used for the year 2006.

Now, the two new data sets are ready for survival analysis. Since the data sets contain a large number of records, we used the life table method, setting the interval at 10 days and stratifying the data by CHMED. The SAS code and some results are shown below.

```
PROC LIFETEST DATA=SASUSER.ST06 OUTSURV=SASUSER.GP06 ALPHA=0.05
METHOD=LIFE WIDTH=10;     STRATA CHMED; TIME DAYS*STATUS(0); RUN;
```

Results in Tables 2 and 3 show that the medications are divided into 7groups in each year by CHMED. In 2005, since the number of prescriptions of rosiglitazone is one, we did not include it. The censored percentages of glyburide- metformin and tolazamide are 100 per cent, which means that it is hard for the patients to change such medicines once they begin taking them. In 2006, we also discarded prandin due to one-time use. The censored rates of glyburide and insulin are 100 per cent, and the rate of metformin use is 83 per cent; all of them demonstrate that the three drugs can seldom be replaced by other medicines. In conclusion, the metformin and insulin uses are stable in both years. Glyburide itself is unstable in 2005, but stable in 2006. Moreover, the average censored rate in 2005 is a little higher than that in 2006, indicating that the usage of prescribed drugs is more stable in 2005.

Summary of the Number of Censored and Uncensored Values					
Stratum	CHMED	Total	Failed	Censored	Percent
1	GLYBURIDE	4	4	0	0.00
2	GLYBURIDE_METFORMIN	4	0	4	100.00
3	INSULIN	8	3	5	62.50
4	METFORMIN	90	10	80	88.89
5	PRECOSE	3	1	2	66.67
6	ROSIGLITAZONE	1	0	1	100.00
7	TOLAZAMIDE	2	0	2	100.00
Total		112	18	94	83.93

Table2. Summary of censored/uncensored values for 2005

Summary of the Number of Censored and Uncensored Values					
Stratum	CHMED	Total	Failed	Censored	Percent Censored
1	GLYBURIDE	2	0	2	100.00
2	INSULIN	7	0	7	100.00
3	METFORMIN	100	17	83	83.00
4	PIOGLITAZONE	2	2	0	0.00
5	PRANDIN	1	1	0	0.00
6	PRECOSE	5	2	3	60.00
7	STARLIX	3	1	2	66.67
Total		120	23	97	80.83

Table3. Summary of censored/uncensored values for 2006

Next, we estimated the differences of survival cases among various drugs by survival functions. To plot the survival distribution function, we needed to define a temporary variable, _TYPE_ by the trim function and the concatenate operator; we also sorted out the data by CHMED and days. In the graph, we overlaid the strata variable in a single plot. The following SAS code is used.

```
DATA SASUSER.GP06SUR (DROP=_SURVIVAL_); SET SASUSER.GP06;
FORMAT _TYPE_ $50.; _OBSERVATION_ = _N_; _SURVIVAL_ = SURVIVAL;
_TYPE_ = '1: SDF ' || TRIM(LEFT(CHMED));
 IF _CENSOR_ = 1 THEN SURVIVAL =.; ELSE SURVIVAL=_SURVIVAL_; OUTPUT; RUN;
PROC SORT DATA=SASUSER.GP06SUR
OUT=SASUSER.SORTGP06; BY CHMED DAYS; RUN;
TITLE1 "Survival Distribution Function";
SYMBOL1 I=JOIN C=BLUE L=1 WIDTH=1 V=NONE…;
PROC GPLOT DATA=SASUSER.SORTGP06;
LABEL DAYS = 'Survival Time';
AXIS1 MINOR=NONE MAJOR=(NUMBER=6) LABEL=('Survival Distribution Function');
AXIS2 MINOR=NONE MAJOR=(NUMBER=6) LABEL=('Survival Time');
PLOT SURVIVAL * DAYS = _TYPE_ / OVERLAY LEGEND=LEGEND1 DESCRIPTION="SDF of DAYS"
FRAME CAXIS=BLACK VAXIS=AXIS1 HAXIS=AXIS2 HMINOR=0 NAME='SDF'; RUN;
```

The survival distribution function (SDF) in 2005 demonstrates that none of the drug, tolazamide, is switched to the other medicines throughout the whole year. The survival rate of metformin decreases little by little from 100 percent to 89 per cent at the end of the year. During the three periods, 30th – 40th, 120th – 130th and 280th – 290th days, the prescriptions of insulin largely decreases; but in the other time, they remain unchanged. The sharp decrease of precose use appears between the 190th day and the 200th day, but before and after that period the usage is stable. A large number of beneficiaries switch their drugs from glyburide to the other medicines during the following periods, the 40th day – 50th day, the 90th – 100th day and 110th – 120th day, which means that the survival rate of the drug decreases to 20 percent at the end of the year. Therefore, the glyburide usage is very unstable in 2005. The SDF in 2006 shows that insulin and metformin survive longer than the other drugs since the survival rates are higher than that of any other drug throughout the year. None of prescriptions of insulin are changed to another medicine until the end of the year. Only less than 14 per cent of the prescriptions of metformin are switched to the other drugs. Between the 90th day and the 110th day, large quantities of prescriptions of pioglitazone are changed to other drugs; however, after that, no more changes happen. The survival rate of precose goes down to 80 percent around the 220th day, to 60 percent around the 280th day and then remains unchanged until the end of the year. The survival rate of starlix sharply decreases on the 320th day and then stabilizes. In general, the metformin and insulin usages are more stable than those of the other medicines.

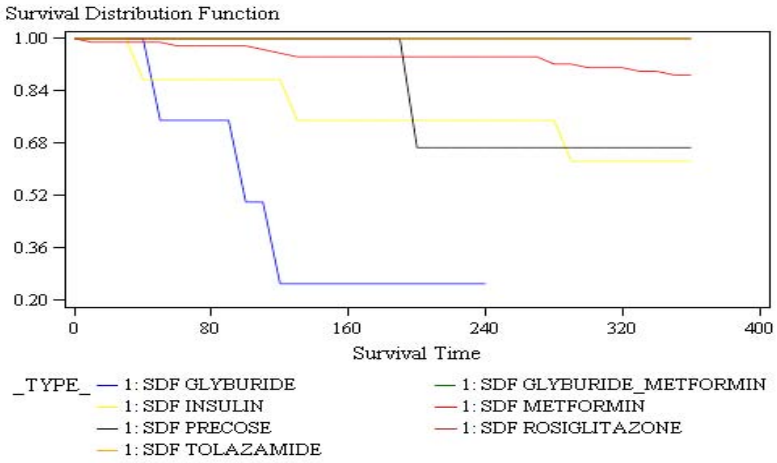


Figure14. Survival distribution function for the year 2005

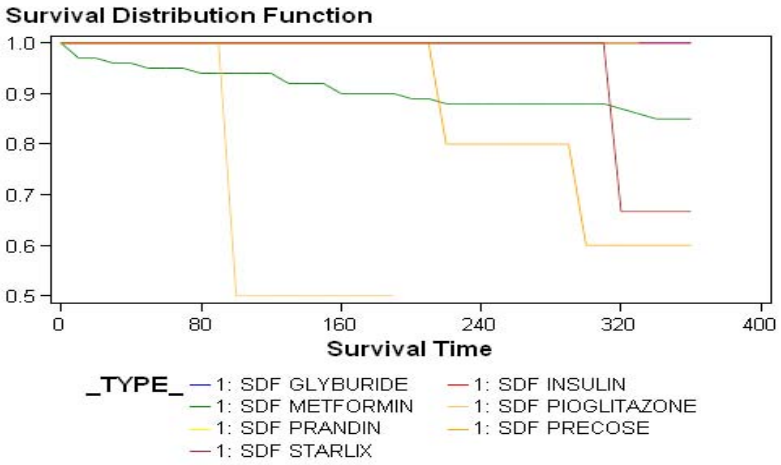


Figure 15. Survival distribution function for the year 2006

CONCLUSION

After the study, we could draw the conclusion that Medicare, part D indeed greatly increases the expenditures of Medicare on diabetes medications from the year 2005 to the year 2006. The prescription drug plan itself reduces the choices of the medicines for diabetes for each year. We also discovered that generally, the usages of insulin and metformin are always more stable than the other drugs. However, glyburide usage is very unstable in 2005 but stable in 2006. In addition, more drugs are switched into the other medicines in 2006, which indicates that the usage of the drugs is less stable than those in 2005. It is also discovered that in 2005, the female patients spend more on insulin and supplies, while the males spend more on metformin. In 2006, the female beneficiaries pay more for metformin. In the future, we will need more information for further study such as the specific dates on which the prescriptions are made to analyze whether the diabetic patients have sufficient treatment under the Medicare program. We also need the data for the year 2008 and the year 2009 to see whether the Medicare reform will benefit more diabetes patients.

REFERENCES

Agency for Healthcare Research and Quality. *Medical Expenditure Panel Survey: Survey Background*, Retrieved from http://www.meps.ahrq.gov/mepsweb/about_meps/survey_back.jsp

Allison, P. (1995) *Survival Analysis Using SAS®: A Practical Guide.* Cary, NC, USA. SAS Institute Inc.

Anoymous, *Annual Average Indexes 2006(Tables 1A-23 A)* Retrieved from http://www.bls.gov/cpi/cpid06av.pdf

Anonymous.*(*Mar.2009), A Status Report on Part D for 2009 in *Report to the Congress: Medicare Payment Policy*

(Chapter 4) Retrieved from http://www.medpac.gov/chapters/Mar09_Ch04.pdf
Centers for Medicare & Medicaid Services (Nov.2008) *Medicare Coverage of Diabetes Supplies & Services*. Retrieved from http://www.medicare.gov/Publications/Pubs/pdf/11022.pdf
Scott, D.(1992) *Multivariate Density Estimation: Theory, Practice and Visualization*. John Wiley & Sons,Inc.
Silverberg, A.& Ligaray,K. (Aug.2008) Oral Diabetic Medications and the Geriatric Patient *Clinics in Geriatric Medicine,Vol. 24*. Retrieved from http://www.geriatric.theclinics.com/issues/contents?issue_key=S0749-0690%2808%29X0003-5
Veeravalli, V. *Detection and Estimation Theory [PDF document].* Retrieved from Lecture Notes Online Web site*:* http://www.stat.ucla.edu/~yuille/courses/Stat153/EMtutorial.pdf

ACKNOWLEDGMENTS

The author thanks Dr. Patricia Cerrito, for guidance and direction in editing this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Xiao Wang
Ph.D. Candidate in Applied Mathematics
Department of Mathematics
University of Louisville
Louisville, KY 40292
Work Phone: 502-852-6022
E-mail: x0wang16@louisville.edu.