A Model for Estimating Medicare/Supplemental Security Income Fraction for 340B Drug Pricing Program Qualification

Steve Hillmer and Prakash P. Shenoy

Center for Business Analytics Research
University of Kansas School of Business
1300 Sunnyside Ave, Summerfield Hall, Lawrence, KS 66045-7601
hillmer@ku.edu, pshenoy@ku.edu

Abstract

We describe a model for estimating the Medicare/supplemental security income (SSI) fraction for a metropolitan hospital for qualification as a distressed share hospital in the 340B Drug Pricing Program. Our model is a mixture of two models. The first model is a probability model that computes the probability of SSI eligibility for each patient visit based on SSI eligibility on a previous visit for repeat patient visits. The second model is a logistic regression model based on categorical factors such as gender, race, median block income of the patient's residence, and distance of the patient's residence from the hospital, for first-time patient visits. These probabilities are then used to estimate the SSI fraction, and compute a 95% confidence interval for the estimate, for each fiscal year. The model is validated using a holdout sample. The model has been used by a hospital for reporting the SSI fraction for two consecutive fiscal years.

Key Words: 340B drug discount program, Medicare supplemental security income fraction, hybrid model, logistic regression model.

1. Introduction

The 340B Drug Pricing Program is a U.S. Federal government program created in 1992 that requires drug manufacturers to provide outpatient drugs to eligible health care organizations at significantly reduced prices. The 340B Program is administered by the Office of Pharmacy Affairs, located within the Health Resources and Services Administration of the Department of Health and Human Services (http://www.hrsa.gov/opa/). Eligible health care organizations include Medicare/Medicaid disproportionate share hospitals that meet the requirements of 42 USC 256b(a)(4)(L).

One of the requirements of 42 USC 256b(a)(4)(L) is that the hospital, for the most recent cost reporting period, has a disproportionate share adjustment percentage (as determined under section 1886(d)(5)(F) of the Social Security Act [42 U.S.C. 1395ww(d)(5)(F)]) *greater than* 11.75%. The disproportionate share adjustment is determined using two fractions. The first fraction, referred to as the Medicare/SSI (or simply SSI) fraction, measures the percentage of all Medicare patients (regardless of means) who are entitled to supplemental security income benefits. The second fraction, referred to as the Medicaid fraction, accounts for the number of Medicaid patients—who, by definition, are low income—not entitled to Medicare.

For most hospitals, estimating the Medicaid fraction is a fairly straightforward task. However, estimating the SSI fraction is difficult as a hospital has little or no information about SSI eligibility of its patients (at the time the fraction has to be estimated). In order to compute the SSI fraction the hospital must submit information on patients admitted during the most recent cost period to the *Center for Medicare Services* (CMS). Subsequently CMS provides the hospital with a file called *Medicare Provider Analysis and Review* (MedPAR) that has data about the SSI eligibility of all the patients in the cost period. However, the MedPAR file is only available retrospectively, two to three years after date of discharge, and thus cannot be used to determine SSI eligibility for the most recent fiscal year patient visits.

Our goal is to build a model that estimates the SSI fraction for the most recent fiscal year using information of patients that a hospital collects such as health insurance card number (HICN), social security number (SSN), gender, race, residential address, etc., and also from information in past MedPAR files. By using this model, hospitals will be able to estimate the SSI fraction for the current cost period, and in turn this will allow the hospital to compute the current disproportionate share adjustment percentage.

2. Data

We obtained data from a hospital, ABC Hospital, that was participating in the 340B Drug Discount Program. The data spanned 8 Federal fiscal years (FFY). These data were from the following two sources.

1. MedPAR files from CMS for FFY1 through FFY8. These files contain 61,922 patient-visit records for FFY1 through FFY8 and 49,996 patient-visit records for FFY3 through FFY8. Each patient-visit record consists of *HICN*, *Admit date*, *Discharge date*, *Length of stay* (in days), *Unit* (hospital, psych, or rehab), *Covered Medicare days*, *SSI eligible days*, *Medicare advantage*, and *Fiscal year*. There are no missing values.
2. The ABC Hospital patient visit database for FFY3 through FFY8 contains 58,862 patient visits. Each patient-visit record includes (among other fields) *HICN*, *Discharge date*, *Address* (Street, City, State, Zip), *Gender*, and *Race*. There are missing values.

We merged information from the ABC Hospital patient visit database with the MedPAR files using *HIC number* and *Discharge date*. After the merge process, for FFY3 through FFY8, there were a total of 49,996 patient-visit records, but 4,985 of these were missing information about address, gender, and race. These were patient-visit records in the MedPAR file that did not match the corresponding records in the ABC Hospital Database (probably because of missing *HICN* or *Discharge date* or non-match of one of these two fields).

The data were separated into a *training* set and a *test* set. The training set consisted of patient-visit records from the merged dataset for FFY3 through FFY7. The test set consisted of patient-visit records in FFY8. The training set was used to develop and evaluate the model. The model that was developed was then used in the test set to predict the SSI fraction for FFY8 for out-of-sample testing, which is necessary to prevent over-fitting.

3. A Model for Estimating the SSI Fraction

The goal is to develop a model that can be used to predict the fraction of SSI eligible patient days to the total hospital patient days for each fiscal year. In a given fiscal year, assume that there are a total of n distinct patients that are admitted to the hospital. For each such patient, define a random variable $Y_i = 1$ if that patient is SSI eligible and $Y_i = 0$ if that patient is not SSI eligible. Notice that the random variable associated with each patient, Y_i, has only two possible outcomes, either the patient is SSI eligible, or the patient is not SSI eligible. The random variable Y_i is said to have a Bernoulli distribution with parameter π_i, where $0 < \pi_i < 1$, π_i denotes the probability that $Y_i = 1$. It is well known (Larsen and Marx, 2006) that the expected value of Y_i, denoted by $E(Y_i)$, is $E(Y_i) = \pi_i$, and the variance of Y_i, denoted by $V(Y_i)$, is $V(Y_i) = \pi_i(1 - \pi_i)$.

If the length of stay (in days) for the i^{th} patient is denoted by LOS_i, then the SSI fraction for a fiscal year is the total length of stays for the patients who are SSI eligible divided by the total length of stays for all patients admitted in the fiscal year. If the SSI fraction is denoted by R, then,

$$R = \frac{\sum_{i=1}^{n}(LOS_i)(Y_i)}{\sum_{i=1}^{n} LOS_i}. \tag{3.1}$$

It is assumed that the random variables Y_i , for $i = 1$ to n, are mutually independent. This assumption will not be true for a patient who is admitted to the hospital more than once during the fiscal year. However, this can be corrected by combining any multiple visits by any patient and treating the multiple admissions as a single admission. This means that for all patients with multiple admissions to the hospital, the variable LOS_i will be the total length of stay for all the combined admissions for that patient in the fiscal year.

Since R is a linear combination of n random variables Y_i, then it follows (Larsen and Marx, 2006) that the expected value of R, denoted by $E(R)$, is:

$$E(R) = \frac{\sum_{i=1}^{n}(LOS_i)(\pi_i)}{\sum_{i=1}^{n} LOS_i},$$

and the variance of R, denoted by $V(R)$, is:

$$V(R) = \frac{\sum_{1=1}^{n}(LOS_i)^2\pi_i(1-\pi_i)}{\left(\sum_{i=1}^{n} LOS_i\right)^2}.$$

Thus, in order to calculate estimates of $E(R)$ and $V(R)$ we need to compute estimates of π_i for each patient that was admitted to hospital ABC during the current FFY. Let $\hat{\pi}_i$ denote the estimated value of π_i for patient i. Then, an estimator of the SSI fraction R, denoted by $\hat{R}$, is:

$$\hat{R} = \frac{\sum_{i=1}^{n}(LOS_i)(\hat{\pi}_i)}{\sum_{i=1}^{n} LOS_i}, \tag{3.2}$$

and an estimate of the variance of the SSI fraction is:

$$V(\hat{R}) = \frac{\sum_{1=1}^{n}(LOS_i)^2\hat{\pi}_i(1-\hat{\pi}_i)}{\left(\sum_{i=1}^{n} LOS_i\right)^2}. \tag{3.3}$$

Thus, in order to compute the estimated SSI ratio, we need a method to estimate π_i for each patient admitted.

Some of the patients who were admitted in the current fiscal year had been admitted to the hospital in previous fiscal years for which we have MedPAR data. For such patients, it is known whether or not they were SSI eligible on their previous visit. It is convenient to partition the set of all patients into the following three mutually exclusive groups: patients who were admitted prior to the current fiscal year and who were SSI eligible at their previous admission (denote this set as *TY*), patients who were admitted prior to the current fiscal year and who were not SSI eligible at their previous admission (denote this set as *TN*), and patients who were not admitted during *FFY*1 to *FFY*7 (denote this set as *F*). Let $P(TY)$ denote the probability a patient is in group *TY*, $P(TN)$ denote the probability that a patient is in group *TN* and denote $P(F)$ the probability that a patient is in group *F*. It follows that (Larsen and Marx, 2006):

$$\pi_i = P(Y_i = 1) = P(Y_i = 1|TY)\,P(TY) + P(Y_i = 1|TN)\,P(TN) + P(Y_i = 1|F)\,P(F).$$

Table 1 gives a cross-tabulation for the patients being admitted in *FFY*3 to *FFY*7 who had been admitted during a previous FFY so that their SSI eligibility from the previous admission is known.

Table 1. Cross tabulation of SSI eligibility for previously admitted patients

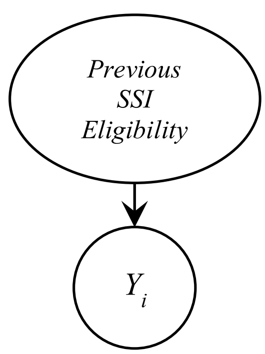
	SSI eligible in previous admission		
SSI eligible in current admission	*no*	*yes*	Totals
no	19,468	147	19,615
yes	68	2,088	2,156
Totals	19,536	2,235	21,771

From this table, it is possible to estimate two probabilities: $P(Y_i = 1|TY) = 2{,}088/2{,}235 = 0.9342$, and $\mathrm{P}(Y_i = 1|TN) = 68/19{,}536 = 0.0035$. Thus, we get:

$$\hat{\pi}_i = P(Y_i = 1) = 0.9342\ P(TY) + 0.0035\ P(TN) + P(Y_i = 1|\mathrm{F})\ P(F). \quad (3.4)$$

Eq. (3.4) can be used to determine $\hat{\pi}_i$ for a subset of the patients admitted in the current FFY. For all patients who were admitted to the hospital prior to the current admission, and who were *SSI* eligible at that visit, then $P(TY) = 1$, $P(TN) = P(F) = 0$, and $\hat{\pi}_i = 0.9342$. For all patients who were admitted to the hospital prior to the current admission, and were not SSI eligible at that visit, then $P(TN) = 1$, $P(TY) = P(F) = 0$ and $\hat{\pi}_i = 0.0035$. For all patients who were not admitted previously, then $P(F) = 1$, $P(TY) = P(TN) = 0$, and $\pi_i = P(Y_i = 1|F)$. In the last case, the value of π_i must be estimated from demographic information of the patient that is typically available in hospital records. Figure 1 shows a graphical model (Koller and Friedman, 2009) for the case where a patient has visited the hospital earlier, and the SSI eligibility during the previous visit is known.

Figure 1: A graphical model for SSI eligibility of patients who have visited earlier



Thus, we need to develop a model to estimate π_i for patients that have not been previously admitted to the hospital based on data that can be derived from information in the hospital data base. The information that is available about each patient that may be relevant to the probability that the patient is SSI eligible includes their age, gender, race, whether they are covered by Medicaid, and whether they were admitted to the emergency room. SSI eligible patients normally have very low incomes; however, the patient's income is not a part of the available patient information. In addition, since ABC Hospital was located near some neighborhoods with a low socio-economic status, it was conjectured that patients who live close to the hospital might be more likely to be SSI eligible than patients who live further from the hospital. Low-income patients will tend to go to the hospital in an emergency, and in this case will tend to go to a hospital close to where they live.

While information about the income is not directly available from the hospital records, one can derive some information about a patient's income from their address, which is available. GIS programs, such as ArcGIS from ESRI (http://www.esri.com), can take an address and match that address to defined geographical blocks that contain median income for that block, estimated from US Census data. In addition, such GIS programs can determine whether or not a given address is within a specified geographic region, like within 15 miles of the hospital. Therefore, for each patient, we determined the median income of the residents in the geographic block that the patient resided in, and the distance from the hospital. Thus, the following variables were used as potential inputs to a statistical model to estimate π_i:

- *Female* (F) = 1 if the patient is a female, and 0 if a male;
- *White* (W) = 1 if the patient is white or Caucasian, and 0 otherwise;
- *Age* (A) = age of the patient at dismissal;
- *Medicaid* (M) = 1 if the patient is Medicaid eligible, and 0 otherwise;
- *ER Admit* (ER) = 1 if the patient is admitted to the emergency room, and 0 otherwise;
- *Residence within 15 miles* ($Within15m$) = 1 if the patient's address is within 15 miles of the hospital, and 0 otherwise;
- *Low Income* (LI) = 1 if the median annual income of the block containing the patient's address is $30,000 or less, and 0 otherwise.

For patients who have not been previously admitted, we are interested in finding a model to determine π_i based on patient inputs. For these patients, the random variable Y_i has a binary response, it only can take on the two values 1 and 0. A common approach in this situation is to use a logistic regression model (Agresti, 1996). The logistic regression model allows for π_i to be a function of the known patient inputs. In particular, the odds associated with each patient is the ratio of the probability that the patient is SSI eligible and the probability that the patient is not SSI eligible or $\frac{\pi_i}{(1-\pi_i)}$. In a logistic regression model, the natural log of the odds ratio is assumed to be a linear function of the input variables:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{1,i} + \cdots \beta_p X_{p,i}. \tag{3.5}$$

There are p input variable X_1 to X_p, and the corresponding parameters $\beta_0, \ldots, \beta_p$ can be estimated from observed data. Once the parameters have been estimated, the input variables can be used to estimate $\ln\left(\frac{\pi_i}{1-\pi_i}\right)$, and this can then be converted to an estimate of π_i as follows:

$$\hat{\pi}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \cdots \hat{\beta}_p X_{p,i}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \cdots \hat{\beta}_p X_{p,i}}}. \tag{3.6}$$

The data in the test set were used to estimate a logistic regression model using all 7 input variables. The results are given in Table 2. Based on these results, the null hypothesis that the coefficients for the variables *Age*, *Medicaid*, and *ER Admit* are equal to zero cannot be rejected at 5% level of significance, as the associated p-values are all greater than .05. This suggests that these three variables are not statistically significant (at the 5% level of significance), and can be dropped from the logistic regression model.

Table 2. Logistic regression results with all input variables

Input Variable	*Estimated Coefficient*	*p-value*
Female (*F*)	.4743	< .0000
White (W)	–.7386	< .0000
Age (*A*)	–.0009	.7518
Medicaid (*M*)	–.0172	.8466
ER Admit (*ER*)	.0981	.1617
Within15m	.2779	.0003
Low Income (*LI*)	.8864	< .0000

The model was then re-estimated using the four remaining variables. The results are provided in Table 3. All of these variables are significantly different from zero (at the 5% significance level) as their associated p-values are all less than .05. The corresponding equation for the log odds is:

$$\ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = -2.61 + .40\,F - .77\,W + .20\,Within15m + .92\,LI \qquad (3.7)$$

Table 3: Logistic regression results with four input variables

Input Variable	*Estimated Coefficient*	*p-value*
Female (*F*)	.3859	< .0000
White (*W*)	–.7735	< .0000
Within 15 miles (*Within15m*)	.1972	.0052
Low Income (*LI*)	.9163	< .0000

One issue we should consider is whether or not the effect of the input variables is stable over the different FFY periods. In the training set, there are 5 FFY's, so that one way to evaluate if this is true or not is to allow the coefficients in Eq. (3.7) to change over the different FFY's. We will illustrate how this was done with the variable *Female* (*F*). Let the variable F_3 take the same value as variable *F* if the data are in FFY3, and take the value 0 if the data are not in FFY3. Define the variables F_j for *j* = 4 to 7 in a similar manner for FFY's 4–7. Then consider the following model:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 F_3 + \cdots + \beta_5 F_7 + \beta_6 Within15m + \beta_7 LI + \beta_8 W. \qquad (3.8)$$

The model in Eq. (3.8) has the impact of the variable *F* in FFY3 on the log odds to be β_1, the impact of F in FFY4 to be β_2, the impact of F in FFY4 to be β_3, the impact of F in FFY6 to be β_4, and the impact of F in FFY7 to be β_5. We want to evaluate whether or not the impact of the variable F on the log odds is the same for each FFY. One way to do this (Larsen and Marx, 2006) is to test the null hypothesis that $\beta_1 = \cdots = \beta_5$ in the model in Eq. (3.8). If we cannot reject this null hypothesis, then we can conclude that the coefficient for the variable *Female* is not changing over the different FFY's. On the other hand, if this null hypothesis is rejected, we would conclude that the impact of *Female* is changing over the different FFY's.

To test the null hypothesis that $\beta_1 = \cdots = \beta_5$, we fit the full model in Eq. (3.8), and subsequently fit the reduced model in which the null hypothesis is true. We then use the likelihood ratio statistic based on the difference between the deviance in the reduced model and the full model. This difference has a chi-squared distribution in large samples that can be used to compute the *p*-value for testing the null hypothesis (Agresti, 1996). This approach was repeated for each of the 4 input variables. The results are given in Table 4. Since the *p*-values for the variables *Female*, *Within 15 miles*, and *Low Income*, are all larger than 0.05, the null hypothesis cannot be rejected for these three variables (at the 5% level of significance). On the other hand, for the variable *White*, the null hypothesis is rejected.

Table 4: *p*-values for testing the null hypothesis of stability over FFY's

Variable	*p-value for testing null hypothesis*
Female	0.1090
White	0.0423
Within 15 miles	0.1899
Low Income	0.1191

Table 5 gives the estimated values for the coefficients for the variable *White* for the five different FFY's. From Table 5, the estimated values for FFY4 through FFY7 are nearly the same so it appears that the coefficients for these years could be assumed to be the same. We subsequently verified that the null hypothesis that the coefficients for FFY 4–7 were all equal could not be rejected.

Table 5: Estimated coefficients for the variable White for different FFY's

Year	*Estimated coefficient*
FFY3	–.5305
FFY4	–.8086
FFY5	–.8215
FFY6	–.8081
FFY7	–.8959

To summarize, the data suggest that the coefficient for the variable W is different for FFY3 than it is for fiscal years FFY 4–7. There is no evidence that the coefficients for the other variables changes over the different fiscal years.

The resulting model for fiscal FFY3 is as follows (all coefficients are rounded to two decimal places for display purposes):

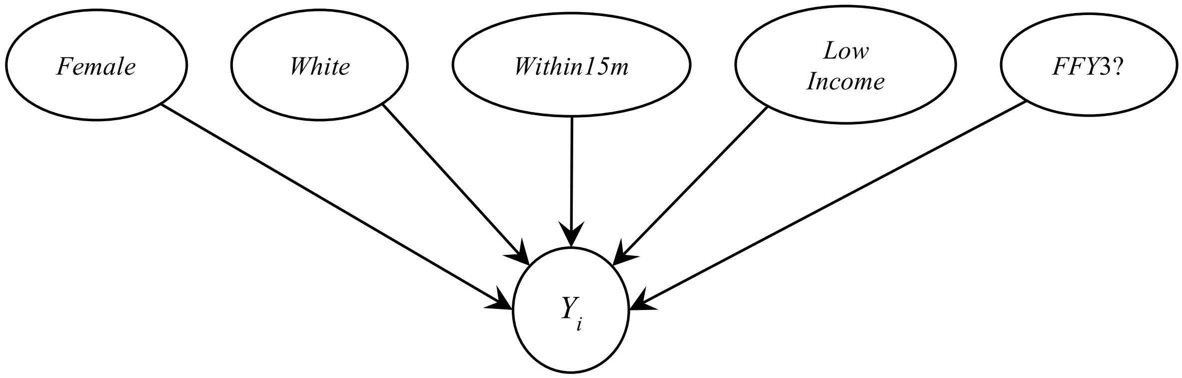
$$\ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = -2.61 + 0.38\,F + 0.20\,Within15m + 0.92\,LI - 0.53\,W, \quad (3.9)$$

and the model for fiscal years FFY4 through FFY7 is as follows:

$$\ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = -2.61 + 0.38\,F + 0.20\,Within15m + 0.92\,LI - .84\,W. \qquad (3.10)$$

Figure 2 displays a graphical model for SSI eligibility for first-time patient visits.

Figure 2: A graphical model for SSI eligibility of first-time patient visits



Note that if the input variables change causing $\ln(\hat{\pi}_i/(1-\hat{\pi}_i))$ to increase, then $\hat{\pi}_i$ will also increase, and if the variables change causing $\ln(\hat{\pi}_i/(1-\hat{\pi}_i))$ to decrease, then $\hat{\pi}_i$ will also decrease. Thus, the value of $\hat{\pi}_i$ will increase if the patient is a female, if the patient's address is within 15 miles of the hospital, if the median income of the block in which the patient lives is $30,000 or less, and if the patient's race is not white or Caucasian. This combination will result in the largest estimated probability of being SSI eligible of 0.24. The lowest estimated probability of being SSI eligible, 0.03, will occur for a white male patient living more than 15 miles from the hospital in an area where the median income is greater than $30,000. Thus, depending upon the characteristics of the patient, the estimated probability of being SSI eligible will change quite a bit.

Based on our model, the probabilities of SSI eligibility for each patient are estimated as follows:

1) If the patient was admitted to the hospital prior to the current visit, and if at the prior admission the patient was SSI eligible, the estimated probability of the patient being SSI eligible for the current admission is 0.9342. By prior visit, we mean discharged on or after the start of FFY1, the earliest fiscal year for which we have data of SSI eligibility from the MedPAR file.
2) If the patient was admitted to the hospital prior to the current visit, and if at the prior admission the patient was not SSI eligible, the estimated probability of the patient being SSI eligible for the current admission is 0.0035.

3) If the patient is being admitted for the first-time since FFY1, then the probability that the patient is SSI eligible is estimated from the logistic regression models in Eqs. (3.9) and (3.10).

If we know the values of $\hat{\pi}_i$ for each patient, the estimated SSI ratio can be computed from Eq. (3.2), and the variance of the estimated SSI fraction can be computed from Eq. (3.3). The standard deviation is the square root of the estimated variance. The estimated SSI fraction, and its standard deviation, can be used to compute a 95% confidence interval for the SSI fraction as follows: Estimated SSI fraction $\pm$ (1.96)*(Standard Deviation). The 1.96 constant is the 97.5 percentile of the standard normal distribution, and normal distribution assumption of $\hat{R}$ is justified based on the large values of n, and the central limit theorem (Larsen and Marx, 2006).

4. Results

For each of FFY3 though FFY8 we used the data available for each patient visit during that particular FFY to compute $\hat{\pi}_i$ for each patient according to our model. We then combined the $\hat{\pi}_i$ with the known LOS_i (the length of stay for the patient visit) to compute the estimated SSI fraction, the standard deviation of the estimate, and a 95% confidence interval for the SSI fraction. Since we have the MedPAR files for these FFY's, we can compute the actual SSI fraction for each FFY. The estimated results for the test year FFY8 were computed only using the data from the training set, FFY3 through FFY7. The results are provided in Table 6.

For each fiscal year, the estimated SSI fraction is very close to the actual SSI fraction. The largest difference occurs in FFY8, the test set, where the actual is .0039 less than the estimated SSI fraction. Also, all of the 95% confidence intervals contain the actual SSI fraction. Based on these results, it appears that our model does an excellent job of estimating the SSI fraction.

Table 6. Results from using our model to estimate the SSI fraction

Fiscal year	*Actual SSI Fraction*	*Est. SSI Fraction*	*|Actual – Est.|*	*Std. Dev. of Est.*	*95% Confidence Interval of Estimate*
FFY3	0.0865	0.0901	0.0036	0.0041	(0.0821, 0.0981)
FFY4	0.0799	0.0766	0.0033	0.0037	(0.0693, 0.0839)
FFY5	0.0748	0.0754	0.0006	0.0030	(0.0695, 0.0813)
FFY6	0.0662	0.0690	0.0028	0.0029	(0.0633, 0.0747)
FFY7	0.0764	0.0769	0.0005	0.0028	(0.0714, 0.0824)
FFY8	0.0675	0.0714	0.0039	0.0025	(0.0665, 0.0763)

Summary & Conclusion

We have developed a model that can be used to estimate the SSI fraction for a given FFY in a particular hospital. This model has been used by ABC Hospital to estimate the SSI fraction for two subsequent Federal fiscal years. This fraction was one input in determining the disproportionate share adjustment percentage for ABC Hospital in order to justify the hospital's eligibility in the 340B Drug Pricing Program. The ability to estimate the SSI fraction soon after the Federal fiscal year ends makes it possible for ABC Hospital to document that it met the drug pricing program threshold, and was qualified for drug reimbursement. Without our model, ABC Hospital would have to wait 2 to 3 years to get the MedPAR data from CMS that could then be used to compute the SSI fraction. Our results show that the SSI fraction can be accurately estimated with currently available data.

While the details of the model we developed for ABC Hospital will change for other hospitals, our general approach can be used for other hospitals. What is required is historical MedPAR data on the SSI eligibility of past hospital patients, and information in current hospital records. The manner in which we deal with patients whose SSI status is known from a previous visit will be the same. For first-time patients, the data available to estimate a logistic regression model may change for other hospitals; however, the general approach of using logistic regression can be applied.

Acknowledgements

The authors would like to acknowledge help from many individuals including: Catherine Shenoy (for formulas in Excel to compute previous SSI eligibility), Alan Halfen (for geo-coding patient addresses to blocks and computing median block income, and distance from the hospital), Andrew Chen, Paul Koch, and Yi Tan (for helping with merging files using Microsoft Access and SAS), and Janahan Ramanathan (for details of MedPAR files). We are grateful to ABC

Hospital administrators (CEO, CFO/Senior VP, and Director of Reimbursement, who are unnamed to protect the identity of the hospital) for introducing us to the problem and providing the relevant data. Finally, we are grateful to Neeli Bendapudi, Dean of University of Kansas School of Business, for introducing us to ABC Hospital administrators.

References

Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, John Wiley and Sons, New York, NY.

Koller, D. and N. Friedman (2009), *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, Cambridge, MA.

Larsen, R. and M. Marx (2006), *An Introduction to Mathematical Statistics and its Applications*, Pearson Prentice Hall, Upper Saddle River, New Jersey.