LIN, MICHAEL
03/04/2014
STAT 497R
FINAL PROJECT PROPOSAL - DRAFT v1.0

Improving Self-Pay Patient Collections

Introduction
The U.S. healthcare industry is currently going through a major transition. To the hospitals, payments from the less profitable government funded programs, i.e. Medicare and Medicaid, are projected to surpass that of the more profitable commercial insurance programs. Hospital CEO and CFO are increasingly pressured to improve the operations of their organizations in order hold on to their bottom line and profit margins. One way to do this is to improve self-pay patient collections. According to HFMA, one of the major issues among healthcare organizations is that "90% to 95% of self-pay patient bills (after government programs or charity assistance) are written off as bad debt," and if the hospitals can improve their self-pay patient collections, e.g. charging patients identified as unlikely to pay their medical bills at the point of registration or check-in, those additional revenues will drop directly to hospitals' bottom lines. Hence, it would be tremendous valuable for the CEO and CFO to be able to identify whether a patient is likely to pay their own portion of the medical bills as early as possible.

Data Description:
The data includes about 5,500 self-pay patient's records from the demo site of my company's revenue cycle management solution. The data include information from hospital's patient accounting systems (patient/guarantor info, insurance plan, self-pay balances, amount paid), 3rd party scoring (demographic information, credit scores), and much more. The data include all patients admitted to the hospital in the month of May 2010, and the snapshot of this data is taken at the date of May 29th, 2011; hence, the outstanding balances within these accounts have not been paid off for at least or more than 365 days – marking them highly unlikely to be collected by the hospital.

A quick note about the data: the data has been "scrubbed" to ensure compliance to HIPPA policies. This means that none of the information can be used to identify individual information. All the records have been masked with fake information (i.e. fake names, fake address, fake SSN, etc.), and the dates and dollar amounts have been altered so that the they are no longer equivalent to the original values, but they are still within the approximate range of the original values so that it can be used to conduct meaningful analyses and trend observations. I also have to ask Professor Nord to destroy the raw data once the project is graded.

Project Goals:
- Explore, clean up, and transform the data to make it useful for R analyses
- Find out what variables are useful to identify whether a patient is likely or unlikely to pay their medical bills. I imagine there will be correlation analysis, linear regression analysis, and/or ANOVA. (I can't decide yet until exploring the data more).
- The deliverables of the project will recommend ways for hospitals to identify patients with low payment likelihood.