

TREATMENT QUALITY ANALYSIS

Objective

The objective of the case is to examine how analytics can model an expert, in this case a physician, in the context of assessing the quality of healthcare patients receive, and use a technique called logistic regression to achieve this objective.

Treatment Quality Problem

D2Hawkeye, a medical data mining company receives claims data. These are data that are generated when an insured patient goes to a medical provider to receive a diagnosis or to have a procedure, for example an x-ray, or to obtain drugs. The medical providers need to get compensated, so the claims data provide the means for them to be paid. An important question is whether we can assess the quality of health care given this claims data. But let's first ask why assessing the quality of healthcare is an important objective. If one identifies patients that have low quality care, one can intervene and improve outcomes for these patients. Moreover, assessing quality correctly can control costs better. Currently, assessing quality is done by physicians who are experts in the health

space using their knowledge, their expertise, and their intuition.

They typically evaluate quality by examining a patient's records, a time consuming and inefficient process. Clearly, physicians cannot assess quality for millions of patients, and D2Hawkeye had, indeed, millions of patients who receive claims data on a monthly basis that the quality of them needs to be assessed. So the key question is as follows. Can we develop analytics tools that replicate expert assessment on a large scale? The goal is to learn from expert human judgment by developing a model, interpret the results of the model, and further adjust the model to improve predictability. The objective is to make predictions and evaluations on a large scale basis, to be able to process millions of assessing the health care quality for millions of people.

About The Data

Medical claims are generated when a patient visits a doctor. Medical claims include diagnosis code, procedures codes, as well as costs. Pharmacy claims involve drugs, the quantity of these drugs, the prescribing doctor, as well as the medication costs. Claims data are

electronically available, they are standardized, and they use well-established codes. However, since humans generate them, they are not 100% accurate. And often, under-reporting is common in the sense that it's a tedious job to record these claims, and as a result, often people under-report them. Also, claims for hospital visits can be vague. In creating a data set, the objective was to assess quality, health care quality. Using a large health insurance claims database, sample of randomly selected 131 diabetes patients were selected for the study. The ages ranged between 35 and 55 and the costs were in the neighborhood of 10,000 to 10,000 to 20,000. The period in which these claims were recorded was September 1, 2003 to August 31, 2005. An expert physician reviewed the claims and wrote descriptive notes, like "ongoing use of narcotics"; "only on Avandia, not a good first choice drug"; "had regular visits, mammogram, and immunizations"; "was given home testing supplies". After this review, expert physician rated the quality of care on a two-point scale, poor or good. Typical comments included, I'd say care was poor. Poorly treated diabetes. Not an eye exam, but overall I'd say high quality. So based on these comments, the variables were extracted. The dependent variable was the quality of care. The independent variables involve the ongoing use of narcotics; only on Avandia,

not a good first choice drug; had regular visits, mammogram, and immunizations; was given home testing supplies. Overall, the independent variables involved diabetes treatment variables, patient demographics, health care utilization, providers, claims, and prescriptions. The dependent variable was modeled as a binary variable -- 1 for low-quality care and 0 for high-quality care. This is by its nature a categorical variable. It only takes two possible values. We have seen linear regression as a way of predicting continuous outcomes. Of course, we can utilize linear regression to predict quality of care here, but then we have to round the outcome to 0 or 1. We will use logistic regression, which is an extension of linear regression, to environments where the dependent variable is categorical (In our case, 0 or 1)