In this lecture, we'll examine how analytics can model an expert, in this case a physician, in the context of assessing the quality of healthcare patients receive, and introduce a technique called logistic regression to achieve this objective.

From the early 2000s, I was a member of the board of a company called D2Hawkeye, a medical data mining company.

The company received 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.

However, defining quality is a complex, not well-defined task.

For example, consider what is involved when we talk about the quality of a book.

It is not a well-defined, algorithmically understood task of defining such a quality.

Currently, assessing quality is done by physicians who are experts in the health space using their knowledge, their expertise, and their intuition.

So how do physicians assess quality?

Physicians are, of course, humans who are limited by memory and time.

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.

So the lecture is a story of using analytics in identifying poor quality care using claims data.