

Lifestyle-based analytics—using data available from an abundance of consumer sources—is one of the best ways to help actuaries and underwriters price health and life insurance.

# you are what you eat

## USING CONSUMER DATA TO PREDICT HEALTH RISK

BY CHRIS E. STEHNO AND CRAIG JOHNS

**H**OW GOOD is medical predictive modeling? “It’s getting better and better all the time,” said one medical director not too long ago. “But it’s not as if we’ll ever know that you just ate a double cheeseburger for lunch.”

Five years ago this statement might have been true. Not today. Big brother exists, and not only does he know you ate that double cheeseburger for lunch, he knows about the large fries, too. The massive consumer datasets that were once used for marketing and sales activities are now being used for a variety of new applications, including the identification of an individual’s health risks.

The Centers for Disease Control (CDC) estimates that lifestyle-based chronic diseases account for 75 percent of the nation’s \$1.4 trillion medical care costs. While this high rate of preventability is surely good news, there’s a catch. Current techniques in widespread use throughout the insurance and health care industries — such as historical medical underwriting, fluid testing, Body Mass Index, etc.— have proved largely ineffective at identifying or predicting the probability of these lifestyle-based diseases in the early or pre-stages.

Newer methods, including predictive medical modeling, prescription drug claims analysis, and post review risk adjustors are having some, but limited success. The problem lies in the diseases themselves. Lifestyle-based diseases are typically not hereditary and have limited or no associated medical precursors.

### **The Age of Consumer Data**

The analysis of lifestyle-based data, otherwise known as lifestyle-based analytics, offers enormous promise to patients, doctors, and health and life insurers. Existing and widely available consumer data reflect individuals’ lifestyles and can be analyzed with an eye toward early disease detection. The analysis of this data either with or without medical data offers advanced techniques in the identification of morbidity and mortality risks.

The abundance of consumer data now available has dramatically changed the way most organizations interact with customers and prospects. From financial services organizations to direct retailers to supermarkets, companies are using advanced statistical data analysis to help with everyday decisions.

An approximate measure of the average amount of information stored for any particular individual is estimated by disk storage per person (DSP) measured in megabytes/person/year. In 1985, DSP was estimated at 0.02, grew to 28 in 1996, skyrocketed to 472 in 2000 and is projected to be an astronomical 3,500 in 2005.

Where is this data coming from? It’s coming from a wide variety of sources: the U.S. Census, public records (deed and registration), financial services (banking, credit, and debit cards, and mortgages), warranty and registration cards, Internet transactions, affinity programs and transaction cards (supermarket cards and frequent-buyer programs), and a vast variety of other sources.





**A report by the U.S. surgeon general has concluded that fully 70 percent of the diseases and subsequent deaths in the United States today are lifestyle related and highly preventable.**

Currently, more than 95 percent of the households in the United States have significant amounts of consumer data tied to their addresses. Initially, this data was stored in aggregate as household information. More recently, new data collection techniques and elaborate algorithms have been developed to distribute the data characteristics across individuals within a household.

Finally, the acquisition costs have dropped dramatically. Just five years ago, the cost of acquiring 100 data elements on an individual averaged around \$3.00 per person and took a great deal of work to line up consumers with their associated data elements. Today, data is available from integrated sources, which saves the user the match-up time and effort. In addition, 300 fields of data now cost less than \$0.25 per person. The fields of data needed for most lifestyle-based health risk applications can cost less than \$0.10 per record. As a comparison, medical or pharmacy data frequently used today to estimate health risks can cost more than \$10.00 per head.

The emergence of this consumer data is providing a first-time look at the lifestyle and psychographic trends in our marketplace. Financial services companies were some of the first organizations to take advantage of this data, primarily because they owned a large portion of it. The most obvious use of this data is demonstrated by the large amount of credit card and mortgage offers sent via bulk mail. These companies have profiled recipients not only as good credit and interest rate risks but also as likely candidates to switch to their product at the given incentive level (at least worth the \$0.50 cost to produce and mail the offer).

The majority of consumer data usage in the health care marketplace today has centered on marketing and sales applications. Direct-to-consumer marketing efforts, as well as agent-directed sales efforts, are using consumer data in numerous ways. For example, one insurance company specializing in senior products has developed its own internal database that contains the consumer data metrics for every U.S. citizen over the age of 50. This database is constantly tested, mined, and re-evaluated with each mailing or direct agent contact recorded. The company has increased prospect response rates by over tenfold.

The use of this consumer data outside the marketing and sales departments is just beginning to take hold in health care corporations. Customer service departments and legal operations are also starting to use this data in order to improve customer services and fraud detection.

### **Lifestyles and Diseases**

The U.S. Department of Health and Human Services reports that one third of the years of potential life lost before the age of 65 is due to chronic disease. The most exciting potential use for this data is in underwriting, disease management, and wellness applications.

Many of the choices people make every day, whether conscious or unconscious, have an impact on their risk for disease. Obviously, cigarette smoking is highly correlated with cancer of the lung, larynx, oral cavity, and esophagus. However, tobacco use also has strong ties to cardiovascular diseases, respiratory diseases, reproductive effects (infertility, low birth weight, and other complications), and a variety of other diseases including cataracts, hip fractures, and low bone density.

Similarly, obesity correlates to heart attacks, congestive heart failure, angina, diabetes, various cancers, sleep apnea, arthritis, complications of pregnancy, gall bladder disease, depression, etc. Alcohol consumption is related to liver disease, hepatitis, cirrhosis, heart disease, cancer of the esophagus, mouth, throat, and larynx, and pancreatitis.

Environmental pollutants are tied to cancer, asthma, and cardiovascular problems. Poor nutrition and inactive lifestyles tie to obesity and related diseases. Poor nutrition and inactive lifestyles, however, are also correlated to osteoporosis, osteoarthritis, rheumatism, lower back pain, and others.

In addition, stress indicators such as financial problems, family difficulties, and occupation have strong ties to depression, back pain, obesity, cardiovascular, and other diseases and medical conditions. Pregnancy, a medical condition rather than a disease, is yet another highly predictable circumstance and is based on factors such as age, family size, ages of current children, family status, and financial indicators.

### **Lifestyle-Based Analytics**

Many of the 300-plus data elements that can be found in the marketplace today revolve around lifestyle-based descriptors. Examples of data fields available include food purchases (fast food, diet food, vegetarian, gourmet), self improvement (health/fitness, dieting/weight loss), fitness activities (aerobics, running, walking, tennis, golf), physical inactivity (television time, computer time, board games, stamp collecting), stress indicators (financial problems, family size and status, occupation) tobacco preferences, alcohol consumption, travel, vehicle type, etc.



Advanced statistical methods, sometimes referred to as data mining, are able to find connections among these data fields and medical conditions. The connections can be used to develop predictive models for health risks.

Lifestyle elements are described as correlated when behaviors lead to a lifestyle-related disease. An example of such choices and the consequence is the high probability of the combination of obesity and an inactive lifestyle leading to diabetes. On the other hand, causality refers to lifestyle behaviors that change because of a particular disease or condition. For example, once diagnosed with diabetes an individual is more likely to purchase diet foods.

Depending on the application for lifestyle-based analytics, the correlation and causality intensities will vary. Put simply, when applied to health underwriting, the model will consider both correlated events and causal events. This makes it easier to identify those individuals who currently have a disease or medical condition or who are at high risk of contracting one in the near future.

Similarly, when lifestyle-based analytics is used for wellness applications, the models also include both correlated and caus-

al events. In this situation, however, the model searches for the opposite or negatively causal events. The greatest benefit for wellness applications isn't finding the people who already have a disease or condition, but finding those in the pre-disease or early onset stages where wellness initiatives have the best chance in preventing the condition.

Lifestyle-based models used for life insurance focus more on the correlated side of the equation. This produces greater predictive capabilities than the health application, which focuses on current or near-term disease identification. For life underwriting, the identification of people who are on a collision course with diabetes, cancer, or a cardiovascular event is paramount.

### **Emerging Applications**

Currently, one of the most obvious and studied uses of lifestyle-based analytics is to augment the underwriting process. Current medical underwriting techniques provide an accurate picture of an individual's past medical conditions and act as a good predictor of future issues associated with those conditions.





**The U.S. Department of Health and Human Services estimates that if current trends prevail, by 2011 our nation will spend \$2.8 trillion on health care, thereby doubling our current annual expense in just seven years.**

While medical underwriting is unable to predict future risks associated with lifestyle-based diseases, lifestyle-based analytics excels in the prediction of diseases, particularly those that aren't hereditary and have no associated medical precursors. When combined with marketing, lifestyle-based analytics can be used as a pre-qualifier for the insurability in direct-to-consumer or agent-directed marketing efforts. Similarly, it can be used as a pre-qualifier for simplified-issue policies.

In a more integrated fashion, lifestyle-based analytics can augment traditional medical underwriting. Lifestyle-based underwriting scores can be incorporated directly into the traditional medical underwriting systems or used as a stand-alone score.

An example of a stand-alone tool is in the mid-sized group-health marketplace where data provided by the prospect is usually limited to historical total claims and an employee census. Lifestyle-based underwriting can provide a statistically viable representation of the group's health as well as the health of the individuals within the group. In addition, when used as a stand-alone model, it complies with the Health Insurance Portability and Accountability Act (HIPAA) because it doesn't use medical information and thus eliminates the need for individual applications and authorizations.

Traditional medical underwriting was developed as a means of discovering unhealthy individuals. Its ability to discover healthy people is often questioned, implicitly casting doubt on the merit of the traditional method. Lifestyle-based analytics, on the other hand, is adept at discovering both healthy and unhealthy individuals. Accordingly, lifestyle-based analytics is now being used to place individuals in preferred categories and/or to accelerate and simplify the application process for healthy individuals by defining and confirming jet applications.

Organizations that use tele-underwriting can integrate the use of lifestyle-based analytics to help direct an underwriter's time. First, lifestyle-based analytics can identify which candidates are likely to have high health risks and direct the underwriter to address specific concerns suggested by the potential risks. Furthermore, lifestyle-based analytics gives companies a fresh look at the application process and allows the data to inform their judgments about the type of questions that need to be added or removed. It also gives them the scoring value of a response in the current application process.

### **Disease Management**

Less obvious, but even more beneficial, is the use of lifestyle-based analytics in the wellness and disease management arenas. Current disease management predictive modeling techniques rely on correlated medical events to trigger an intervention. In a typical scenario, it takes five or more medical events to occur before a medical disease

management model can predict a result. Unfortunately, in the majority of cases this technique doesn't detect a disease until it has progressed to a point beyond which early detection can make a significant difference.

Lifestyle-based analytics can speed up the identification and reliability of disease prediction. The combination of lifestyle-based data with medical data allows for earlier detection of diseases. With the combination of data types, the prediction can be made after two or three medical events. This earlier detection results in significant savings through current disease and case management techniques in place today.

In some wellness applications, lifestyle-based analytics can predict an individual's propensity to contract or have a disease in the early-onset or pre-stage, before medical conditions present themselves. It's well-established that intervention at this early point is extremely beneficial to the affected individual, his employer, and the insuring organization.

Unlike health risk assessments, lifestyle-based analytics doesn't require applicants to fill out a form or employers to provide incentive programs to encourage the application process. In addition, it's not subject to the accuracy of a self-reported application involving personal information. Since lifestyle-based analytics doesn't require any medical information, it's HIPAA compliant and therefore needs no authorization by the individual.

### **Conclusion**

In a study by the American Cancer Society, various lifestyle and nonlifestyle-based elements were attributed to the risk of developing cancer. In aggregate, 82 percent of the risks were lifestyle-based conditions, including diet, smoking, sexual behavior, occupation, alcohol, and sun radiation.

Overall, lifestyle-based diseases account for more than 70 percent of the disease in the United States today and represent 75 percent of the total medical dollars spent. While historical datasets and techniques have proved relatively unsuccessful at predicting health risks for lifestyle-based diseases, new lifestyle-based datasets and predictive modeling techniques are proving highly effective.

Whether used in an underwriting context or for disease and wellness management, use of consumer data and lifestyle-based analytics as a health risk measurement technique will accelerate quickly over the next few years as companies gain significant competitive advantages through its use.

CHRIS STEHNO, MBA, is a consultant in Milliman's Denver office and is an expert in lifestyle-based analytics. CRAIG JOHNS, PhD, is a statistician in Milliman's Denver office and an expert in predictive modeling.