Predictive analytics is a valuable tool and is being applied to a growing number of areas in auto insurance operations. In recent months, I have been asked questions about the benefit of applying predictive analytics to auto physical damage data and medical data to identify questionable injuries. To be sure, predictive analytics has shown benefits in claims operations by improving fraud referrals, identifying subrogation opportunities and “right tracking” claim assignments. Consequently, on the surface, the approach sounds appealing. However, a closer look into what predictive analytics can offer and the constitution of the data employed reveals a problematic landscape.

In the book “Competing on Analytics” authored by Tom Davenport and Jeanne Harris, analytics is described as “the extensive use of data, statistical and quantitative analysis, explanatory and predictive models and fact-based management to drive decisions and actions”. Further, in the table below, they outline a succinct progression of business analytics and intelligence (note the table presented in “Competing on Analytics” was adapted from a graphic produced by SAS and used with permission). Each method presented requires increasing sophistication (e.g., optimization is a more sophisticated level than predictive modeling).

<table>
<thead>
<tr>
<th>Method</th>
<th>Business Questions Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Analysis</td>
<td>Why is this happening?</td>
</tr>
<tr>
<td>Forecasting/extrapolation</td>
<td>What if these trends continue?</td>
</tr>
<tr>
<td>Predictive Modeling</td>
<td>What will happen next?</td>
</tr>
<tr>
<td>Optimization</td>
<td>What’s the best that can happen?</td>
</tr>
</tbody>
</table>

In the deployment of analytics, most will concur that the usefulness of results will depend greatly on the quality of the data, the appropriateness of the data analysis and the quality of assumptions employed. It is also important to note that when models are properly deployed, they don’t give answers; they yield information about a tighter distribution on possible outcomes.

Some have suggested that the opportunity to use predictive analytics to identify problematic or questionable injury claims is analogous to providing a claims organization a smoke alarm. The analytics alert one to a problem early before it becomes a bigger problem. This analogy, while inaccurate, is actually instructive regarding the proper use of predictive analytics. A smoke alarm detects smoke, which is an outcome that is caused by an actual fire. And, it hopefully alerts one of an actual fire before it becomes a bigger fire. Alternatively, effective predictive analytics applications are more consistent with alerting one to conditions that are favorable for a fire taking place. One would want to investigate
Predictive Analytics with Predictable Problems

the actual existence of the fire before taking action (e.g., calling 911, activating a fire suppression system, etc.) because there can be considerable costs to such actions or decisions when there are false alarms. There still may be value in knowing favorable conditions exist early, but an investigative step or process is usually recommended to confirm the actual existence of the problem before decisions are made. This is why in fraud analytics applications, fraud investigations are usually conducted after potentially fraudulent claims are flagged and before final decisions about fraud are made. In subrogation applications, subrogation opportunities are normally further evaluated before being acted upon.

Let’s examine why further investigative steps might be required when basing analytics, in part, on medical data. Medical data usually available to the auto insurance industry is accumulated from injury claims presented by a claimant, their attorney or their medical care provider. This information is often audited for reasonableness and appropriateness, in both first and third party injury claims. The entire process of collecting the data assumes that the basis for the claim, specifically the auto accident, actually caused the claimed injury and the need for treatment.

The assumption that causation exists is actually perpetuated by medical providers. Physicians, for example, are classically trained to take a history from the patient, conduct a physical examination and then incorporate the findings of the previous two steps into his/her diagnosis and treatment recommendations. So, if someone claims a condition is related to an auto accident, the physician factors this information into his/her treatment plan and cites the auto accident as the cause. This problem is further compounded when the treatment plan is tailored to the severity of the injury as perceived by the medical provider, not by an analysis of the severity of the collision and the imparted stresses and strains. Subsequently, treatment guidelines are often used, regardless of the severity of the impact (in terms of collision energy), to identify when treatment is reasonable and appropriate. These real world practices begin to illuminate the flaws in the medical data maintained by auto insurance companies and their service providers. Medical treatments are often not prescribed or subsequently evaluated in light of the physical events required to produce the need for them. This flaw is inherent in the data maintained by utilization review or medical audit providers. From a scientific perspective, injuries are actually caused by a very specific physical stress or strain, or specific combinations of stresses and strains, which are unique to the injury. These requisite stresses and strains can be found to exist or not exist based on an analysis of the physics of the accident, position of the occupant in the vehicle, use of restraint systems and other factors. Additionally, when requisite stresses and strains exist, they still must exceed an individual’s tolerance to same, before the injury can be caused.

Without such a scientific analysis, how is one to determine which medical data in a data set is questionable versus not questionable? Wouldn’t this knowledge be required to build a sound model to identify questionable claims so auto physical damage relationships can be developed for both groups?

From a holistic view, it would also seem there would be much learned about questionable injury claims by understanding data in instances when injuries are not claimed as a result of an auto collision. The good news is that accidents that do not cause injuries occur very frequently. This outcome is
Predictive Analytics with Predictable Problems

corroborated by human subject testing in low speed crash tests. Over 75 percent of the approximately 4,000 scientific human subject test exposures known to the author produced no injury. The bad news is that little claim data is typically collected by an insurer on an uninjured passenger when a claim is made. Sometimes no data is available because no claim is made as a result of an accident. Wouldn’t some information about the condition and physical attributes of the individual not making an injury claim, his or her seating position in the vehicle, use of available restraint systems, etc., be relevant to a reliable model that predicts questionable injury claims? Other times, injuries occur but data collected is incomplete because there is a determination that there is either no coverage or liability. Again, wouldn’t more complete information under these circumstances be important to a reliable model?

Structurally, one can begin to see how trying to apply predictive analytics to auto physical damage data and medical data to identify questionable injuries can be problematic. To illustrate, let’s examine a hypothetical TMJ injury claim from a low energy frontal collision from a statistical and scientific perspective. Statistically, data will exist that involve accidents with varying degrees of physical damage to the claimant’s automobile accompanied by a TMJ injury claim. Most likely, these injuries will be observed at a low to very low incident rate. Data will also likely show reasonable and customary costs for treatment of the TMJ injury and likely reflect deviations from these standards. So, does the statistical model suggest the injury should be questioned? If so why? Because the treatment period was too long or not properly coded? Because the medical provider was not a physician? Because of the location of the clinic? What gives the adjuster the basis to make a decision or take an action and defend it?

Scientifically, the answer is straightforward. To traumatically injure the TMJ, there must be contact between the mandible and an object with sufficient force to create the stresses and strains to cause injury. Simply, if there is no mandible contact with an object, there is no opportunity for a TMJ injury (none of the previously referenced human subject test exposures experienced a mandible strike or a TMJ injury). Can an occupant in a vehicle involved in a frontal collision strike his/her mandible on an object? It is certainly possible. The answer becomes clearer once we know where the occupant was seated in the vehicle and whether or not he or she was restrained. Incidentally, both of these critical facts are typically not found in the medical data.

Scientific analysis as just described can be applied a wide variety of injuries (e.g., neck, back, shoulder, knee) to actually determine when questionable injuries were or were not caused from a collision. To use a popular example illustrated in the book Moneyball by Michael Lewis, there was an important difference in the predictive outcome in baseball games between the “old school” use of batting average to evaluate players and newer analytics utilizing on-base percentage. While admittedly not a perfect analogy, using medical data in the proposed approach in lieu of scientific analysis certainly has many “old school” attributes.

So what can be done? First, understanding the limitations of the data and the resulting implications are critical. Use of statistical methods for dealing with unknowns and data limitations is common in predictive analytics. However, when these methods are used in claims applications, they should be followed by investigative processes that resolve the related unknowns and validate the assumptions.
Predictive Analytics with Predictable Problems

employed. Such a step will help avoid calling the fire department when conditions are favorable for a fire, but there is actually no fire. Secondly, there is an opportunity to improve the data collected when there is no injury claim made in an accident. Based on the TMJ injury example previously provided, collection of scientifically relevant facts during a claim investigation can, over the long term, help offset many of the significant limitations found in the medical data. Thirdly, consider utilizing analytics that, while not considered as only predictive analytics, actually define and describe the event in an accurate and defensible way. For example, referring back to the book “Competing on Analytics”, one form of analytics described was employed by VisViva Golf Inc. The company uses nanotechnology embedded in golf clubs that is connected to Bluetooth radio to calculate and measure technical aspects of a golfer’s swing (swing speed, acceleration, deceleration, etc.) as well as use the data with predictive analytics to provide guidance to improve the golfer’s swing. Today, automated analytics are available for claims organizations which scientifically determine acceleration and deceleration of vehicles in accidents and the resulting implications to injury potential or, in other words, scientifically identify questionable injuries. More specifically, the scientific analytics described in the previous TMJ example can be systematically applied to claim data such as repair estimate information and injury claim information.

In conclusion, using medical data in predictive analytics applications to identify questionable claims could be, well, questionable. While all the numbers and formulas associated with today’s analytics suggest objectivity, experienced managers understand that the "garbage in, garbage out" phenomenon has never been more true. Realizing the power of analytics requires being realistic about what models can and cannot do, improving the quality of the data feeding models, and creating the appropriate managerial processes around them. In the context of injury claims, models should be accompanied by a well defined investigative process that will provide an adjuster with actionable information and the basis for defensible decisions. Otherwise a claim organization may find itself in a position of systematically creating more fires than they are trying to avoid. Lastly, the use of predictive analytics in this approach begs the following question: Why not use defensible analytics in the identification of questionable claims? Would this not eliminate an unnecessary step?