<u>Portal</u>　　<u>Workers Compensation Law</u>　　<u>Recent Cases, News, Trends & Developments</u>
<u>Predictive Analytics and Big Data in the Workplace and Beyond: The Science Behind the Crystal Ball</u>

11-25-2014 | 04:30 PM　　│　**Author:** <u>Karen C. Yotis</u>

Predictive Analytics and Big Data in the Workplace and Beyond: The Science Behind the Crystal Ball

Karen C. Yotis, Esq., a Feature Resident Columnist for the <u>LexisNexis Workers' Compensation eNewsletter</u>　*, provides insights into workplace issues and the nuts and bolts of the workers' comp world.*

An awful lot of ink has been devoted lately to Predictive Analytics and Big Data—two very hot new terms that are causing unprecedented change (and equal amounts of consternation) throughout the insurance and workers' compensation industries. Predictive analytics refers to the use of applications and techniques to build, test, refine, and apply algorithms in an effort to "predict who will click, buy, lie, or die."[fn1] And predictive analytics depends of course on big data—those vast quantities of information characterized by volume (incredibly high level), velocity (real time, quick access), and variety (as broad, large and wide as the Texas sky). Predictive Analytics unleashes the power of Big Data, and Big Data moves predictive analytics away from a swami's crystal ball and into the realm of science.

BIG DATA IS THE NEW COAL

Big Data arises from the convergence of cloud, mobility and the <u>Internet of Things</u>　(IoT), shifts that are reordering and leading enterprise businesses throughout the United States. This big data is creating a perfect storm, which is resulting in a second information age that has developers writing apps for data (instead of writing them strictly for device platforms) in order to mobilize actions for better outcomes. The type of Big Data we're talking about here is SO BIG that IBM Chairman and CEO Virginia Rometty recently analogized it to the 'new coal,' calling Big Data "the world's natural resource for the next century."[fn2] And with Big Data of course comes Big Power.

Gregory R. Wagner M.D. traced the outer boundaries of the power that can be derived from Big Data in his NIOSH Science Blog titled "<u>Can Predictive Analytics Help Reduce Workplace Risk</u>　?" when he wrote:

"Predictive analytics is at work each time Amazon informs a customer they may be interested in a particular product; each time Netflix recommends a movie; when life insurance companies risk-stratify and set rates for potential policy holders; and when some credit card companies target potential customers with specific incentives to acquire their card. Harrah's casino has been using predictive analytics to promote customer loyalty. UPS is said to employ predictive analytics to analyze sensor data to target fleet maintenance more economically and to adjust delivery routes to reduce traffic-related delays and fuel consumption. Software predicting whether airline ticket prices will rise or fall in the next week can guide purchasing decisions on line."

MANIPULATING BIG DATA TO CHANGE BEHAVIOR

Some seasoned industry folks (the ones who probably never got too excited about social media) may nevertheless wonder what the hullabaloo is all about. After all, using numbers to gain insight which can assist in the decision-making process is nothing new for workers' compensation, for sales, or for other forms of research. Plus it's no secret that prediction of human behavior has often been used to combat financial risk, fortify healthcare, toughen crime fighting, and boost sales. What IS revolutionary about predictive analytics and big data as time marches on towards the year 2015—*and this is key*—is the use of predictive modeling to manipulate big data in order to help define the next generation of tools and techniques that are able to *translate* that data into meaningful information that someone can do something with. This is all about going beyond data that simply helps to forecast behavior to <u>data that works to actually change behavior</u>. And this momentous shift (that in effect has the WORLD engaging in behavior modification) is hitting the workers' compensation industry simultaneously with remarkable sea-changes such as opt-out legislation, drug formularies and evidence-based medicine.

Thus, as society has progressed from the industrial revolution through the digital revolution into today's new information age, we have extended intelligence to the edge with wearable computing, drones, sensors, and new user experiences. And this digital-physical blur, which has the potential to extend across the workforce, has begun to alter workers' compensation in a number of fascinating (and troublesome) ways.

CUTTING THROUGH THE TECHNICAL MUMBO JUMBO

To understand how something complex works, it's often best to begin with clear terms. The Gartner IT Glossary describes <u>predictive analytics</u>　as:

"[A]ny approach to data mining with four attributes:

> An emphasis on prediction (rather than description, classification or clustering)

> Rapid analysis measured in hours or days (rather than the stereotypical months of traditional data mining)

> An emphasis on the business relevance of the resulting insights (no ivory tower analyses) and

> An increasing emphasis on ease of use, thus making the tools accessible to business users"

Whereas predictive analytics refers to the way data is conceived, predictive modeling is the actual *process* of analyzing data to create a statistical model of future behavior. Thus, predictive modeling solutions are a form of data-mining technology that (<u>again according to Gartner</u>　):

"works by analyzing historical and current data and generating a model to help predict future outcomes. In predictive modeling, data is collected, a statistical model is formulated, predictions are made, and the model is validated (or revised) as additional data becomes available."

Dr. Wagner demystified the intricate inter-workings of predictive analytics and predictive modeling in a workers' compensation context when he explained:

"While traditional epidemiology searches for the determinates of disease and injury over time in populations, predictive analytics focuses on the prediction of events or effects in individuals or other affected "units" (such as particular production lines or workplaces) during a specific time window. Epidemiology is used to establish exposure—response relationships using careful measurements of both exposures and health effects while controlling for population variability. Predictive analytics often uses available historical data reflecting the endpoint of interest, for example a five-year history of mine injuries reported to MSHA, then divides the data into a training and test set. Using 'machine learning' approaches, an algorithm is developed from the test set using a wide range of available, potentially relevant data, to fit the test set data. That algorithm is then applied to the test set of data to assess how well the algorithm predicts the results, and then is further refined if necessary. When a good algorithm is developed, it can be applied as new data are gathered. In this example, an algorithm that identifies mines where serious injuries are likely to occur could stimulate operators to adopt preventive practices and also help direct mine inspectors."

Dr. Wagner also points to prior safety experience, a workers' age and time in a specific job; time during a shift and hours worked during a prior day, week or month; geographic location; how recently a workplace inspection occurred and what the results were; season; enterprise profitability; the presence of an injury prevention program; and union representation of the workforce as just a few of the plentiful examples of data that is potentially relevant to predicting injury or disaster.

With respect to Big Data (which we've clearly characterized as really, really HUGE), it's also helpful to qualify limitations. First, because TPAs and carriers from multiple companies that come together over time run on multiple systems that don't always talk to each other, being in possession of big data doesn't necessarily mean that a predictive model can be populated with information that makes sense. Further, predictive models may incorporate data elements from ancillary providers that may not even be present in claims systems. The same is true of medical providers—selective availability of data may limit the predictive value of a particular model where all of the usable legal/medical/claim information (and it's ALL valuable in the big data context) is unavailable simply because it hasn't been translated into a form of text that can be used. Simply put, a predictive model can gather and incorporate only the data that people are willing to input and enter.

While the computing power to process big data is presumed in a predictive model, other essential technology elements must be carefully planned and implemented. The efficacy of your predictive model starts with the type of data you are capturing. The myriad of information generated during the intake and life of a claim in notes and from medical personnel and attorneys must all be accounted for and incorporated. Another key element is the information repository—the digital place where big data is stored. The location of information and the manner in which that information is integrated and managed are critical elements of the predictive process. The final element that provides the basis of your data's technology component is access to big data. Inaccessible data has no value.

Big Data is also useless if it is inaccurate. Keith M. Higdon, VP Claims Data Analytics for ACE Group Claims talked about how to best utilize big data and predictive modeling to drive actions that affect workers' compensation claims management recently at the National Workers' Compensation Defense Network's 2014 Fall Conference. During his presentation Higdon stated that quality is the number one problem within claims when it comes to data. Thus, in spite of all the algorithmic complexity at play, Higdon referred to predictive modeling as being an art form of sorts, a balancing act of deciding how far back to go, how wide to cast your net, and what elements to include. According to Higdon, determining the data you are comfortable with that will produce the best results is "part science and part heart."

Higdon supplemented his conference comments when he told this author:

"The biggest challenge stemming from the use of predictive modeling focuses on measuring program success. Models cannot stand alone—they are designed to drive action and it is that action that drives outcome. In order to measure an outcome, the action must be monitored for consistent application. Too often there is a rush to complete the model, and implement action without enough thought around the necessary data collection and monitoring metrics essential to assess outcomes."

James D. Stephens, Assistant Director of the Illinois Department of Insurance echoed concerns about inaccurate data when he told this author:

"I would think another challenge would be finding credible data. If you don't have credible data it won't matter what your calibration is, results can come out wrong."

A study by the Coalition Against Insurance Fraud   (with the help of big data analytics software maker SAS Institute)—which points to the lack of IT resources as the biggest challenge that insurers face when implementing anti-fraud technology—reports that many companies continue to struggle with the deployment of proactive predictive analytics tools because of IT resource constraints. In addition, while 43 percent of insurance companies deployed predictive modeling technologies to combat fraud, data integration and poor data quality are among the additional cited challenges.
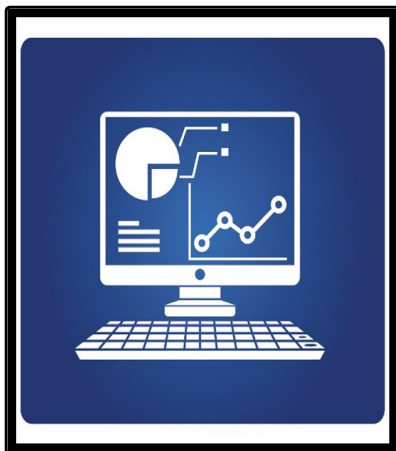
WALKING THE WALK (AND NOT JUST TALKING THE TALK)

To understand these concepts and how they operate, we must also consider the inherent implementation challenges that come into play with predictive analytics, predictive modeling and big data.

Business Issues

Assistant Director Stephens underscored implementation concerns when he told this author:

"To me the biggest challenge for predictive modeling is execution. I am afraid that implementation of predictive modeling would be seen as an effort to replace underwriters. They will see this as being forced on them rather than a tool that will supplement what they already add to the equation."

Various general business issues, as well as issues that are specific to insurance companies, also present challenges in the development and implementation of predictive analytics and/or a predictive model. In a Property Casualty 360 blog titled "The Challenges of Implementing Advanced Analytics   ," Kevin Bingham, John Lucker, Laura Ward and Stacey Peterson outline the various implementation concerns that industry leaders are beginning to observe. Main points of contention include: the necessity of early buy-in from senior leadership along with a clear corporate strategy for integrating predictive models; early IT involvement required to bring models to life in an organization's daily operations; fashioning the correct project management structure with a clear cadence of project milestones and pre-identified ownership of project deliverables; adequate change management that extends beyond a model's roll out; and finding a balance between building a precise statistical model with the ability to explain the model and how it produces results. Additional insurer-specific hurdles to implementation include executing communications and training plans across often siloed organizations; competing initiatives and systems; financial, information, IT and

human resource constraints; and buy-in and use of a model by the end user, which may be more complicated in smaller organizations because of sensitivity around expectations for a particular model.

Competition

Competition is also a big issue when it comes to the secret algorithms that form the analytical basis of a predictive model. Competitors are after this information to determine if there is a variable that hasn't been considered that another company views as predictive. Consumer advocates want the algorithms so they can challenge predictive outcomes and educate consumers about what they've opted into. However, an algorithm has no value without the big data to plug into that algorithm, so access issues presented as a result of market competition raise challenges across the spectrum. Cost is spurring a cottage industry that is selling various predictive models, and creative solutions are finding their way to the marketplace now in an effort to make all of this feasible. During a session titled The Impact of Automobile Technology, Innovations, and Other Neat Stuff on Insurance offered at the Association of Insurance Compliance Professionals' 2014 annual conference, Chris Ziance from The Progressive Group of Insurance Companies discussed the boatload of money that his company has spent developing programs and collecting data and emphasized Progressive's intention to continue fighting to protect its proprietary materials. In fact, in a major effort to guard its proprietary algorithms, Progressive won't file in states that require disclosure without first obtaining simultaneous trade secret protection.

Privacy

Predictive Modeling and Big Data also raise privacy concerns. Injured workers and insureds want to know if employers and insurers are going to be using the data to determine other things that are incidental and extraneous to the handling of a particular claim. Another privacy issue is how (and whether) others get access to Big Data. During his presentation at the AICP annual meeting, Ziance also outlined how Progressive, which is using telematics from car devices to gather behavioral information about insured drivers, offers a great deal of disclosure about what is collected, how collected data is stored, and who can access collected data in order to address privacy concerns. It appears that consumer acceptance is growing all the time (at least while the telematic device results in a premium discount as opposed to a behaviorally supported premium increase), but this same acceptance may not occur in the inherently adversarial (and increasingly cost-conscious) workers' compensation context.

Discrimination

Predictive modeling of psycho-social characteristics may be the juncture where the tremendous potential of predictive modeling gives way to some more troublesome developments. For no matter how telling or informative a predictive model may be, when it comes to regulated industries discriminatory procedures that result in a disparate impact on a particular demographic traditionally tend to pose serious concerns. For example, panel members at the aforementioned AICP conference session revealed to attendees that Birney Birnbaum, executive director of the Center for Economic Justice, Austin, Texas, and an accredited NAIC consumer representative who has voiced concerns about the predictive analytics arising out of telematic devices being hooked up to insureds' automobiles, is clamoring for the NAIC Auto Study Group to analyze each factor considered and that factor's impact by economic status and area to determine whether predictive analytics based upon the speed, distance and time of day that an insured drives will result in an unusual impact on low income areas.

At the NWCDN conference, Higdon suggested another way to look at the application of social data (in the claims as opposed to a pricing context). Higdon reminded attendees that workers' compensation claims at its most basic is a business that involves people—people who are hurt; people who provide care; and people who manage risk. From a social perspective, Higdon compared what's going on with claims with the process that occurs when Amazon suggests the next best thing that someone should buy. In Higdon's view, it's these same motivations in people's behavior that can help us to understand claims.

Eric Nordman, Director of Regulatory Services and Director of the Center for Insurance Policy and Research at the National Association of Insurance Commissioners, cut through to the bottom line of possible discrimination with predictive analytics when he told this author:

> "Predictive modeling and big data are clearly changing the world we live in. Predictive modeling offers tremendous potential for improving the accuracy of prices charged for insurance; however, there are also concerns from regulators about the unintended consequences of the modeling efforts. Regulators have a more difficult time understanding all the nuances of predictive modeling and explaining to the consumer how the complex pricing algorithm results in the ultimate price to the consumer. Some regulators have concerns as to whether the predictive model has identified a non-offensive risk classification factor that is so closely predictive of a societal offensive risk classification factor that its use should be prohibited."

PRIVATE POSSIBILITIES AND PUBLIC BENEFITS

In spite of the concerns and challenges, the insight provided through the predictive modeling of big data in achieving the goals of program effectiveness and cross savings, and in the guiding of the decision-making process, is too valuable to be set aside. This is why conference planners that look to the future for the next best trends and issues have given predictive analytics front and center billing on their session schedules this year.

To supplement the insights offered by the aforementioned sessions at this year's AICP and NWCDN events is the input from speakers at the October 2014 California Workers' Compensation & Risk Conference in Dana Point who delved into the various aspects of predictive modeling and big data in the context of workers' compensation insurance market trends for CEOs and stakeholders. When asked how they found predictive modeling useful, these thought leaders gave a variety of responses that position predictive analytics and big data squarely within the processes and procedures put in place to address cost containment, claims analysis and adjudication, litigation strategy and pricing.

According to the Dana Point panel, analyzing the characteristics of claims that develop—along with the attributes that each workers' compensation claim has at 3, 6 and 12 months out from the date of injury—is one of the major ways the industry is using predictive modeling to get a handle on claims. Analytics are also useful to price out particular accounts, by looking at the characteristics of certain claims that make them expensive, and developing procedures that get those expensive claims into the hands of a company's best claims handlers. Panel members also stated that actionable analytics are important, and that it is useful to look at the proportion of claims that involve specific as opposed to cumulative trauma as well as how many of those claims included litigation. The speakers also revealed that analytics on the relationship between management and labor can be very telling. According to the panel of experts, it's generally best to obtain a predictive modeling analysis as early as possible in the life of a claim in order to achieve optimal predictive results from your revelatory numbers.

The Dana Point panel also talked about the excellent work that TPAs are doing with predictive modeling of psycho-social questions to get an early jump on issues to prevent developmental claims. According to these experts, identifying psycho-social characteristics has been one of the best predictive pieces with respect to identifying co-

morbidities and mental health. Panelists cautioned that these items must be identified very early in the life of a claim, and claims personnel must be extremely candid with the results.

Higdon supplied some additional post-conference insight about promising innovations on the horizon in 2015 relative to predictive modeling and big data when he told this author:

> "Significant advancements in text mining are enabling greater access and analysis of claim data sets. Challenges commonly occur with claims data due to the limitations of structured fields. Diving directly into unstructured data such as claim handler notes provides valuable detail and insight into both claim and claimant characteristics not otherwise available. Text mining is a new analytics instrument that enhances predictive models, allows drilldowns into cost drivers, and can even provide an early glimpse into emerging risks and exposures."

THE MULTI-MILLION DOLLAR QUESTION

Predictive analytics and predictive models can accomplish many things. In addition to informing decisions about which product to purchase, which store to frequent, and which movie to watch, analytics are being used by police departments to make specific patrol assignments in order to improve crime prevention, by health insurers to tailor health screening and modify treatments to those most likely to succeed with a specific patient, and by political strategists to reduce the cost of their voting drives in order to concentrate on constituents most likely to support a candidate.

But can predictive analytics reduce risk in the workplace and prevent employees from getting hurt on the job?

The jury is still out.

During his NWCDN presentation, Higdon was careful to qualify the value and usefulness of predictive modeling. Higdon bluntly stated that a model won't shorten claims duration, prevent litigation or increase closure rates. Rather, predictive modeling will help to identify opportunities that highlight pathways of potential action that help you to leverage those opportunities to arrive at a different (and hopefully better) outcome. In Higdon's view, to improve the odds of arriving at a desired outcome (say , for example, the complete eradication of workplace injuries), the algorithm geniuses must find a way to infuse the model so it lives within a system, similar to the way information feeds back into Amazon's infused model every time you make a purchase. So at least from a claims perspective, this is the direction in which predictive modeling needs to travel.

NIOSH is also aware of various innovators' forays into the business of supplying predictive models built around predictive analytics relating to the reduction of workplace injuries. Dr. Wagner's blog mentions at least one consulting firm that is marketing prediction to reduce workplace injury, others who are working to identify 'leading indicators' that can be measured and are associated with good OSH performance; and a handful of others who are employing predictive analytics in private industry and academia with an OHS/prevention focus.

There are very smart people out there working on this conundrum given the tremendous potential (and profit) available from predictive analytical models. Without doubt, when someone figures this all out, the licensing opportunities will be limitless. But while we wait for the next killer application that will help us to predict workplace injuries out of existence, I encourage readers to take a long, hard look at their smart phones. You all are holding the next Big Data collection device right in the palm of your hand.

Disclosure: LexisNexis® Legal & Professional Solutions is affiliated with LexisNexis® Risk Solutions, which has developed C.L.U.E.® Commercial.

Publisher's Note: This article was revised Dec. 3, 2014 to include more quotes from Keith M. Higdon.

Footnotes:

1. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die   , by Eric Siegel, is a primer that reveals the power and perils of predictive analytics, which incorporates a discussion about why organizations predict death, including one health insurance company. Siegel's book demonstrates how the practice and process of drawing predictions from big data is the driver that makes practically everything—from science, business, and finance to sports, fashion and politics—tick.

2. CNBC Exclusive: CNBC Transcript: IBM Chairman and CEO Virginia Rometty Sits Down with CNBC's David Faber Today on "Squawk on the Street," 13 May 2014, press release and unofficial transcript available here: http://www.cnbc.com/id/101669112

_____

Tags: _discrimination_   _privacy_   _workers compensation_   _workers compensation insurance_   _social media_   _predictive analytics_   _injury prevention_   _fraudulent claims_   _big data_   _predictive modeling_   _algorithm_   _app_   _behavior_   _competition_   _data mining_   _trade secrets_   _psycho-social_   _smart phones_