

JUNE 2018

BIG DATA AND THE ROLE OF THE ACTUARY

American Academy of Actuaries
Big Data Task Force



AMERICAN ACADEMY *of* ACTUARIES

Objective. Independent. Effective.™

ACTUARY.ORG

The American Academy of Actuaries is a 19,000-member professional association whose mission is to serve the public and the U.S. actuarial profession. For more than 50 years, the Academy has assisted public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.



AMERICAN ACADEMY *of* ACTUARIES

Objective. Independent. Effective.

AMERICAN ACADEMY OF ACTUARIES
1850 M STREET NW, SUITE 300, WASHINGTON, D.C. 20036
202-223-8196 | WWW.ACTUARY.ORG

© 2018 American Academy of Actuaries. All rights reserved.

Big Data and the Role of the Actuary

JUNE 2018

American Academy of Actuaries Big Data Task Force

Bob Beuerlein, MAAA, FSA, FCA, CERA—*Chairperson*

Dorothy Andrews, MAAA, ASA, CSPA—*Vice Chairperson*

Mary Bahna-Nolan, MAAA, FSA, CERA

Elena Black, MAAA, EA, FCA, FSA

Elizabeth Brill, MAAA, FSA

Patrick Causgrove, MAAA, ACAS

Robert Curry, MAAA, FCAS

Ian Duncan, MAAA, FSA, FCIA, FIA

Seong-Min Eom, MAAA, FSA

Andy Ferris, MAAA, FCA, FSA

Audrey Halvorson, MAAA, FSA

William Hines, MAAA, FSA

Liaw Huang, MAAA, EA, FCA, FSA

Robert Miccolis, MAAA, FCA, FCAS

David Sandberg, MAAA, FSA, CERA

Martin Snow, MAAA, FSA

Jim Toole, MAAA, CERA, FSA

Jason VonBergen, MAAA, FSA

Contents

Executive Summary: Current and Emerging Practices	1
What Is Big Data?	1
The American Academy of Actuaries' Role	2
Data Analytics Techniques and Methodologies	3
Regulatory Considerations	4
Professionalism	6
Section I: Current and Emerging Practices	8
What Is Big Data?	8
The American Academy of Actuaries' Role	10
Data Analytics Techniques and Methodologies	10
Application of Predictive Analytics	12
Practice-Area-Specific Applications	14
Considerations in the Use of Predictive Analytics	17
Data Sources	19
Section II: Regulatory Considerations	23
Benefits and Challenges to Insurers, Regulators, and Consumers	23
Existing Regulatory Framework	25
Emerging Regulatory Developments	28
Section III: Professionalism	36
Actuarial Professionalism	37
Ethical Considerations	40
Algorithms, Techniques, Correlation, and Causality	41
Role of the Actuary	44
Appendices	45
Appendix 1: InsurTech	45
Appendix 2: Actuarial Standards of Practice (ASOPs)	51

Executive Summary

Current and Emerging Practices

Remarkable advances have been made over the past decade in the use of Big Data, including the Internet of Things, machine learning, cognitive computing, and artificial intelligence, and the field continues to evolve. These advances have led to the development of a multi-billion-dollar industry referred to as InsurTech, the innovative use of technology in insurance, which is expected to have a significant impact on insurance and the work that actuaries perform.

While the use of Big Data in the property and casualty insurance area is more developed than in some of the other areas of actuarial practice, significant advances have been made in recent years in the use of Big Data in health and life insurance. Similar advances in the pension area have not been as noticeable. However, it can be expected that over the next decade, all areas of actuarial practice will be significantly impacted by the use of Big Data.

What Is Big Data?

“Big Data” has become a common term and topic of discussion throughout the world. A glance at any news outlet will likely find a story that describes some facet of the Big Data phenomenon.

Broadly speaking, Big Data refers to the collection of extremely large data sets that may be analyzed using advanced computational methods to reveal trends, patterns, and associations.

Big Data can support numerous uses, from search algorithms to InsurTech. The definition of Big Data generally includes the “5 V’s”:

- Volume** Large amounts of data are collected and require processing.
- Velocity** Data is available and must be processed at lightning speed, frequently instantaneously.
- Variety** The data being used comes in different forms.
- Veracity** The reliability of the data is not uniform.
- Value** The data being extracted must be usable or be able to be monetized.

Big Data is not only about data. New, advanced tools are available that enable Big Data to be processed and utilized in ways that were not previously possible. These tools include data handling capabilities and computational techniques such as predictive analytics and advanced algorithms that have significantly increased data speed and storage capacity.

With the rapid advances in the availability of data and the development and proliferation of advanced data analytics techniques, the insurance industry's interest in Big Data analytics capabilities has grown commensurately. InsurTech is the use of recent technology to bring efficiencies and innovation to the insurance industry. It has led to new products, new distribution channels, new risks for insurance companies, and changes to claims handling methods. It also can lead to greater emphasis on market conduct examinations, potential jurisdictional arbitrage, and a more complex regulatory environment. As the utilization of Big Data becomes a potential disruptor for the insurance industry, the need for professionals who are bound by a code of conduct, adhere to standards of practice and qualification, and subject to counseling and discipline if they fail to do so, will become more apparent.

The American Academy of Actuaries' Role

The focus of the American Academy of Actuaries regarding Big Data has been and will continue to be around the concepts of professionalism and public policy. From a public policy standpoint, the Academy continues to work with regulatory bodies on how these complex issues impact the public through the regulation of insurance and governance of retirement systems. The American Academy of Actuaries continues to work with policymakers and regulators to address and refine regulatory frameworks in which Big Data work may appropriately be governed.

From the perspective of the U.S. actuarial profession, the pillars of actuarial professionalism—the Code of Professional Conduct, actuarial standards of practice, and U.S. Qualification Standards—provide a framework for actuaries to perform actuarial services related to Big Data.

Data Analytics Techniques and Methodologies

With regard to advanced data analytics techniques for Big Data, four types exist:

- Descriptive: What happened?
- Diagnostic: Why did it happen?
- Predictive: What will happen?
- Prescriptive: What should I do?

Most insurers have a long history of performing descriptive and diagnostic analytics. Included in diagnostic analytics are traditional statistical inference techniques that seek to characterize the relationships between variables or elements. Recently, there has been a significant increase in the use of predictive analytics that differs from traditional inferential statistics in that it is not concerned with proving the “why” behind what’s driving a relationship but only with whether variables help predict a given outcome objective. Determining the optimal action to take considering these analytics is the function of prescriptive analytics.

Descriptive data analysis and feature extraction/selection, as well as data visualization, use sophisticated mathematical tools, including principal component analysis, ridge and lasso regressions, and clustering algorithms. Understanding the data and the relationships between variables is of utmost importance before engaging with the models designed to predict. Visualization tools such as box-plots, histograms, scatter diagrams, and scatter matrix are used for this task.

When using these techniques, actuaries need to consider that it is not always possible to develop a precise and definitive formula where complex human behavior is involved. Accordingly, actuaries need techniques in addition to predictive analytics to significantly increase their understanding of anticipated behavior or events and support their strategies and decisions. This becomes a professionalism issue for actuaries.

Regulatory Considerations

Benefits and Challenges to Insurers, Regulators, and Consumers

Despite its potential, there are a number of concerns regarding Big Data that impact insurers, regulators, and consumers.

Insurers

The use of predictive analytics can lead to a better understanding of risk than traditional methods. New sources of data not only increase dimensionality of data dramatically, but also allow for the use of more direct indicators of individual risk. New methodologies allow for a potentially better understanding of risk drivers and relationships between them, as well as detecting potential fraud. The benefit of a better understanding of risk is protection against adverse selection and improved reserve adequacy, such as with health care models that can be used to more accurately predict utilization of health care services.

Potential drawbacks of new insurance models driven by predictive analytics include disruptions in the fundamental pricing principles of the industry, such as the collapse of the law of large numbers, disruptions in risk peaks and subsequent difficulty in assessing short-term risk, and premium inadequacy resulting from both new pricing models and substantial upfront build costs.

Regulators

Regulators may benefit from better advance knowledge of outcomes and could apply some predictive analytics techniques directly to their review processes. Potential benefits for regulators include the enabling of a more streamlined process for approval of pricing and rate filings as well as scanning of annual statement filings to detect previously unknown patterns. Regulators can also use predictive analytics to detect fraud.

Reviewing predictive analytics can be a challenge to regulators given the amount of data used to develop a model, the complexity of the techniques, and limited regulatory resources. Regulators also may have difficulty explaining complex models to consumers and other interested parties who are trying to understand the impact of the models on insurance rates. The NAIC Big Data (EX) Working Group is proposing additional support for regulators for reviewing new models that contain predictive analytics capabilities.

Consumers

Analytics can lead to more competition and more competition can lead to more options for consumers. Predictive analytics can result in quicker decisions on underwriting, where allowable, because of the use of external data. Claim settlement can also be accelerated using predictive analytics. Analytics also can result in better offerings by insurers to policyholders from the use of external data that can help inform decisions regarding better fit of coverage.

The main challenge to consumers is lack of transparency: trying to understand the data and analytics being used to determine their eligibility for products and the price they are being charged. It may not be clear to the consumer how they are being underwritten or what behaviors they can modify or steps they can take to get a better rate. A potential issue with pricing based on predictive analytics is that it can lead to more granular pricing, which benefits some consumers but not others. This broader distributed range of prices could be perceived as unfair. Privacy issues are also a concern for consumers because of a lack of transparency regarding how data is collected and used.

Existing Regulatory Framework

The legal and regulatory requirements that potentially govern the use of Big Data by insurers at the state, federal, and international levels fall into two categories: 1) those designed to protect consumers in general; and 2) those intended to prohibit discrimination against certain protected classes of individuals.

Emerging Regulatory Developments

NAIC Activity (NAIC Big Data (EX) Working Group)

Advances in statistical modeling techniques and evolving sources of data are challenging existing regulatory processes. Methods, such as those used to calculate premiums, are more complex than ever before. Current algorithms and models are not as easy to understand and follow as traditional algorithms. In addition, with the exploding availability of data, including consumer data, insurers are utilizing types of data not previously incorporated into advanced modeling techniques. Moreover, for many aspects of the insurance business, companies differ in methods and approaches employed and in their documentation and explanation of such methods and approaches.

The complexity and evolution of the methods and approaches used by insurers is threatening to outpace the rate at which regulators can educate themselves on these new methods and approaches. To address these issues, the NAIC has increased training opportunities, such as the predictive model training that was organized by the American Academy of Actuaries at the 2017 Summer NAIC Insurance Summit, and information-sharing forums to address current gaps in knowledge. The NAIC also formed a Big Data Working (EX) Group (the Big Data WG).

Regulatory Sandboxes

“Regulatory sandboxes” have recently received significant attention from regulators, companies, and start-ups active in the financial services industries. Although the concept can take a variety of forms, a regulatory sandbox is generally a discrete regulatory environment designed to encourage innovation in a regulated industry. Depending on the context, a sandbox might function primarily as a forum for encouraging earlier and more frequent engagement between innovators and regulators, without necessarily allowing for waivers of existing law. Alternatively, a sandbox can relax regulatory requirements, effectively creating an alternative, less restrictive regulatory regime for proposed innovations. Given the regulatory issues involved, it is not difficult to imagine this concept being applied to insurance companies in the context of Big Data.

Professionalism

Actuaries have professional obligations to uphold the reputation of the actuarial profession and fulfill the profession’s responsibility to the public in the emerging area of Big Data. An important part of this responsibility is to comply with the law. In many situations, actuaries also have unique insights into the results and implications of the use of Big Data and must be willing and capable to explain such insights, where appropriate, to the key stakeholders of the work, such as regulators, consumers, company management, auditors, etc. The value of the actuaries’ work is enhanced through adherence to the Code of Professional Conduct, actuarial standards of practice, and U.S. Qualification Standards. A key attribute of the applicable standards is the requirement for actuaries to provide explanations and rationales for their conclusions.

Professional judgment from actuaries is critical in the utilization of Big Data in actuaries’ work. Actuaries provide added value to Big Data work in their ability to “connect the dots” through a deep understanding of the subject matter. In exercising professional judgment, it is important for actuaries to be cognizant of the fact that without performing proper

analyses or validation, the results of Big Data can be misleading. A combination of a good understanding of the context in which the data was obtained and avoidance of unthoughtful adherence to the results of a model can aid in better Big Data outcomes.

There are many professionalism issues that may be encountered when working with Big Data and predictive analytics. The work of actuaries is governed by the Code of Professional Conduct (Code) and must comply with applicable actuarial standards of practice (ASOPs). The Code and ASOPs provide a framework for dealing with issues of professionalism that might arise in the work of actuaries. While actuaries have traditionally dealt with large volumes of data and a variety of modeling techniques, Big Data may pose new challenges that differ from those that actuaries encountered in the past. In addition, actuaries historically have built analyses and models based on traditional inferential statistical methods (descriptive and diagnostic analytics); however, predictive analytics techniques offer unique and different challenges to consider.

Role of the Actuary

In many applications of Big Data in businesses in which actuaries are employed, multidisciplinary teams are utilized to efficiently and effectively complete the project. The teams are commonly composed of statisticians, computer scientists, data scientists, and actuaries. Actuaries on these teams may be thought of as the subject matter experts. But actuaries may be positioned to be the quarterbacks of the Big Data teams. With the proper background, an actuary can understand and direct the work of the Big Data multidisciplinary team based on their professionalism requirements and subject matter expertise.

As the evolution of Big Data continues in the areas of practice in which actuaries provide services, the professionalism and technical expertise provided by actuaries are essential elements upon which the public and regulators can place reliance. The professionalism requirements of actuaries provide guidance for the proper application and disclosure of Big Data assumptions and methodologies. They require actuaries to adhere to the high standards of conduct, practice, and qualifications of the actuarial profession, thereby supporting the actuarial profession in fulfilling its responsibility to the public.

Section I

Current and Emerging Practices

Remarkable advances have been made over the past decade in the use of Big Data, including the *Internet of Things*, *machine learning*, *cognitive computing*, and *artificial intelligence*, and the field continues to evolve. These advances have led to the development of a multi-billion-dollar industry referred to as InsurTech, the innovative use of technology in insurance, which is expected to have a significant impact on insurance and the work that actuaries perform.

Section I of this monograph provides examples of current and emerging applications of Big Data in the various practice areas of actuarial work. While the use of Big Data in the property and casualty insurance area is more developed than in some of the other areas of actuarial practice, significant advances have been made in recent years in the use of Big Data in health and life insurance. Similar advances in the pension area have not been as noticeable. However, it can be expected that over the next decade, all areas of actuarial practice will be significantly impacted by the use of Big Data.

What Is Big Data?

“Big Data” has become a common term and topic of discussion throughout the world. A glance at any news outlet will likely find a story that describes some facet of the Big Data phenomenon.

Broadly speaking, Big Data refers to the collection of extremely large data sets that may be analyzed using advanced computational methods to reveal trends, patterns, and associations.

Big Data can support numerous uses, from search algorithms to InsurTech. The definition of Big Data generally includes the “5 V’s”:

- Volume** Large amounts of data are collected and require processing.
- Velocity** Data is available and must be processed at lightning speed, frequently instantaneously.
- Variety** The data being used comes in different forms.
- Veracity** The reliability of the data is not uniform.
- Value** The data being extracted must be usable or be able to be monetized.

Big Data is not only about data. New, advanced tools are available that enable Big Data to be processed and utilized in ways that were not previously possible. These tools include data handling capabilities and computational techniques such as predictive analytics and advanced algorithms that have significantly increased data speed and storage capacity. The value of the data in the absence of these tools might be orders of magnitude less than it is currently. Within the context of this monograph, Big Data refers to both the data and the associated analytics applied to the data.

With the rapid advances in the availability of data and the development and proliferation of advanced data analytics techniques, the insurance industry’s interest in Big Data analytics capabilities has grown commensurately. InsurTech is the use of recent technology to bring efficiencies and innovation to the insurance industry. It has led to new products, new distribution channels, new risks for insurance companies, and changes to claims handling methods. It also can lead to greater emphasis on market conduct examinations, potential jurisdictional arbitrage, and a more complex regulatory environment. InsurTech is discussed in depth in Appendix 1 of this monograph. As the utilization of Big Data becomes a potential disruptor for the insurance industry, the need for professionals who are bound by a code of conduct, standards of practice, and qualification standards will become more apparent.

This monograph describes some uses of Big Data and predictive analytics in the work of insurance and pension actuaries. The primary focus is on the regulatory and professionalism aspects and the roles of actuaries who work with Big Data.

The American Academy of Actuaries' Role

The focus of the American Academy of Actuaries regarding Big Data has been and will continue to be around the concepts of professionalism and public policy. From a public policy standpoint, the Academy continues to work with regulatory bodies on how these complex issues impact the public through the regulation of insurance and governance of retirement systems. The American Academy of Actuaries continues to work with policymakers and regulators to address and refine regulatory frameworks in which Big Data work may appropriately be governed.

From the perspective of the U.S. actuarial profession, the pillars of actuarial professionalism—the Code of Professional Conduct, actuarial standards of practice, and U.S. Qualification Standards—provide a framework for actuaries to perform actuarial services related to Big Data.

Data Analytics Techniques and Methodologies

With regard to advanced data analytics techniques for Big Data, four types exist:

- Descriptive: What happened?
- Diagnostic: Why did it happen?
- Predictive: What will happen?
- Prescriptive: What should I do?

Most insurers have a long history of performing descriptive and diagnostic analytics. Included in diagnostic analytics are traditional statistical inference techniques that seek to characterize the relationships between variables or elements. Recently, there has been a significant increase in the use of predictive analytics that differs from traditional inferential statistics in that it is not concerned with proving the “why” behind what’s driving a relationship but only with whether variables help predict a given outcome objective. Determining the optimal action to take considering these analytics is the function of prescriptive analytics. This monograph primarily discusses predictive analytics.

Any data analytics project starts with data, and a variety of techniques are used to reconcile, scrub, and pre-process that data. A common rule of thumb for most predictive analytics projects is that 80 percent of the time is devoted to ensuring data quality, understanding the data and relationships within, and extracting/selecting features (or predictors).¹ Many techniques exist to ensure quality of modeling data, including reconciliation using a range of data sources, dealing with data issues such as missing values, and reducing data dimensionality, if necessary.

Descriptive data analysis and feature extraction/selection, as well as data visualization, use sophisticated mathematical tools, including principal component analysis, ridge and lasso regressions, and clustering algorithms. Understanding the data and the relationships among variables is of utmost importance before engaging with the models designed to predict. Visualization tools such as box-plots, histograms, scatter diagrams, and scatter matrix are used for this task.

Most data analytics techniques and tools use various forms of optimization and statistical algorithms, as well as machine learning methods. Model inferences are then evaluated and analyzed. The tasks of implementation and documentation, as well as the purpose of the predictive or data analytics model, provide for additional considerations for model selection, including the level of transparency, ease of use, update, and feedback loop execution.

Some models are harder to interpret than others, and precise formulas and causal relationships are not always discernable. To this end, other techniques are typically used to supplement and explain models' results, such as expert opinions, customer questionnaires, existing relevant industry research, and research from other industries.

Commonly used modeling techniques include the following:

- Generalized linear modeling
- Linear discriminant analysis
- Time series analysis
- Survival analysis
- Association algorithm
- Sequence analysis
- Clustering algorithms
- Classification algorithms
- Neural network analysis
- Decision trees
- Random forests
- Gradient boosted machines
- Support vector machines
- Naïve Bayes analysis
- Bayesian estimation
- Ensemble models
- Text mining
- Behavioral economics models

¹ "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says"; *Forbes*; March 23, 2016.

When using these techniques, actuaries need to consider that it is not always possible to develop a precise and definitive formula where complex human behavior is involved. Accordingly, actuaries need techniques in addition to predictive analytics to significantly increase their understanding of anticipated behavior or events and support their strategies and decisions. This becomes a professionalism issue for actuaries. See Section III on professionalism for more information.

Application of Predictive Analytics

The following are examples, by function, of how insurers use predictive analytics:

Marketing: Insurers use predictive analytics to market to consumers. Companies can observe consumer behavior in a variety of forms and build targeted advertisements to appeal to customers. Companies can gather information about consumers using cookies or other mechanisms. Companies also can build “propensity to buy” models to target consumers who are more likely to make a purchase. These activities can reduce marketing costs, leading to overall cost reductions or the reallocation of marketing funds for other purposes.

Engagement: After an insurance purchase has been made, companies engage with targeted customers using customer-specific methods, as research shows that an engagement focus by the company leads to more future sales and better retention as compared to a transactional focus. These targeted customers and engagement methods are selected using predictive analytics. Customer value propositions should improve, as should internal performance management. However, companies should recognize that this increased engagement can offset some or all the cost reductions achieved through more efficient marketing.

Underwriting: Predictive analytics can improve underwriting processes where this is permitted by regulation. Streamlined application processes and shorter underwriting wait times can improve company placement rates. The elimination of costly underwriting methods, such as the use of bodily fluids in life insurance, can significantly reduce expenses. These enhanced risk assessment processes can then reduce the cost of the policy through an improved ratio between mortality and expenses.

Product Development: Insurance companies can use predictive analytics techniques to find new markets and design new products for it. Companies can offer a better-fitting product line to the market by analyzing prior history data on insurance, driving history, health records, and lifestyle.

Claims and Reserving: Claims management for fraud detection is another area where predictive analytics can be useful, as are process efficiency, cost reduction, fast-tracking, and principle-based reserves (PBR) assumption-setting for life insurance.

Decision-Making Analytics: Predictive analytics can be used to mimic human decision-making, to produce decision-making rules that are better than those used previously, and to map potential outcomes more quickly and with more accuracy. Each of these can provide major benefits, but also come with certain constraints. For example, the matching of human decision-making means that human biases will be preserved. Producing decision-making rules requires an investment of significant effort, and the mapping of potential outcomes requires vast quantities of data.

Behavior Analytics: Acquiring a comprehensive understanding of customer behaviors and needs is important so that insurers can anticipate future behaviors, offer relevant products, and appropriately segment their business. For example, analytics systems can spot if a customer is likely to lapse by detecting a large number of calls to a customer service center.

Customer Satisfaction and Upselling: In addition to providing predictions about when a customer is likely to lapse, gaining customer insight with predictive analytics also can help insurers to develop trusted relationships and engage customers with accurate information. As a result, insurers can be more successful in achieving positive outcomes such as solving customer problems in real time and upselling and cross-selling products.

Targeted Marketing: Developing a more complete understanding of customer behavior allows insurers to become more efficient in targeting products and services. This can be accomplished by offering personalized services, contacting the customer for special offers when they are likely to lapse, or offering a package for a family life cycle event.

Practice-Area-Specific Applications

Big Data and predictive analytics are used in each of the four actuarial practice areas.

Life Insurance

Some predictive analytics offerings for accelerated underwriting develop scores from biometric information. Others look at predictors less commonly used in traditional underwriting, such as public records, social media activity, motor vehicle reports, credit information, and wearable devices. Examples of predictive models that are used include: triaging individual requirements (e.g., determining if blood is needed), best classification model, multiple classification model, and a true mortality prediction model.

For life insurance, the application of predictive analytics to actuarial assumption development, such as mortality or lapses, sometimes starts with term insurance and then is expanded to permanent coverages. Predictive analytics techniques are applied to term insurance to improve term conversion rates (the rates at which customers convert their term policy into a whole life policy). For annuities, predictive analytics is used to develop mortality assumptions, improve longevity analyses, and to model policyholder behavior under guaranteed riders such as Guaranteed Minimum Withdrawal Benefits (GMWB).

GMWBs and Guaranteed Minimum Income Benefits (GMIB) are two common product features where predictive analytics are used in the setting of policyholder behavior assumptions. Specifically, companies are using predictive analytics to model how policyholders exercise guaranteed benefit options. For example, policyholders can wait longer than the initial waiting period to gain additional guarantees. Predictive analytics will examine such things as policy size, funding level, asset allocations, percent of guaranteed amount withdrawn, and prior withdrawal history to predict the likelihood of a future withdrawal.

Companies can utilize apps and wearables that enable the proactive tracking of their customers, while helping the customers to manage their health. For example, a company may make post-issue changes in underwriting classification based on health-related data from wearables.

Property and Casualty Insurance

Historically, property and casualty insurance companies have utilized predictive analytics for purposes such as pricing, especially personal lines; underwriting; claims management; quoting; fraud detection/prevention; premium audit; and agent selection/retention.

While these techniques have been used for many years, more sophisticated and broader applications continue to evolve. These advanced methods are often used for rating and underwriting, risk management, targeted marketing, behavioral analytics, and product development.

The types of analyses that can be developed, including entire new rating algorithms, new classification plans such as territory structures that incorporate geographical elements, or scoring algorithms such as insurance scoring, may not look like traditional simple rating steps. These complex algorithms and models may be difficult for a reviewer to follow and understand. As a result, such algorithms and models, if used for rating and underwriting purposes, may attract additional scrutiny from regulators as the regulators seek to understand the new and emerging practices.

Analytics also can assist risk management efforts by providing feedback on unsafe actions or conditions and generating alerts for potential fault or failure situations. The “Internet of Things” enables sensors to provide continuous monitoring and feedback. Telematics can provide information on driving actions or conditions that may be used to provide discounts for safe drivers.

Health Insurance

An important application in health care modeling is the task of risk adjustment, utilizing risk scoring models that can be both predictive and descriptive. Risk adjustment in health insurance became prevalent in the 1990s, before the widespread use of predictive modeling. The models employed (often referred to as grouper models) were developed using linear regression to predict resource utilization in a period from a set of covariates (frequently age, sex, and diagnoses). They are referred to as “grouper models” because they group together diagnostic International Classification of Disease (ICD)-9 (15,000) or ICD-10 (80,000) codes² into a smaller number of hierarchical codes consisting of similar diagnoses. Grouper models are powerful tools for both risk adjustment and for predictive modeling because they significantly reduce the dimensionality of predictive modeling without significant loss of accuracy.

² The International Classification of Diseases, Tenth Edition (ICD-10) is a clinical cataloging system that went into effect for the U.S. health care industry on Oct. 1, 2015. Accounting for modern advances in clinical treatment and medical devices, ICD-10 codes offer many more classification options compared to those found in its predecessor, ICD-9.

Risk adjustment is widely used as a predictive analytics tool in reimbursements of many government health systems (Medicare Advantage; Medicaid; ACA exchanges) and private insurance contracts, in reimbursement for providers taking on risk under capitation and risk-sharing arrangements, and for determining the effectiveness of providers in building high-performance networks.

Early applications of predictive modeling in underwriting and “case finding” (identification of high-risk patients for management) used grouper models because these were frequently available (being required by insurers for risk adjustment). However, in the early 2000s purpose-built predictive models began to proliferate, often for case finding purposes for patients with specific conditions (cardiovascular disease, diabetes, mental illness, orthopedic, etc.), as well as specific problem areas such as hospital readmissions following the Centers for Medicare and Medicaid Services’ (CMS) introduction of penalties for excessive readmissions. Currently, predictive analytics is widely used in case finding for medical management programs.

All health insurers; many provider groups; many hospitals, pharmacies, and pharmacy benefit management (PBM) companies; and all medical management companies employ predictive modeling in some form or another to identify high-cost or high-risk patients. Predictive modeling was used, prior to the passing of the Affordable Care Act, to predict high-cost members of insurance pools for underwriting, rating, and pricing. It may still be used in rating and pricing for blocks of business but not at the individual level. Its use is often limited for underwriting, although it may be used to price an entire group under large group lines of business and/or association business in some states.

Pensions

The use of Big Data and data analytics in the pension area currently is limited, but its use is growing with the emergence of new roles for pension actuaries. One notable use is mortality improvement assumptions for pension valuations. These assumptions often are derived via extensive mortality data analysis, graduation to smooth out random noise, trend identification, and pattern extrapolation. This also is an example of data analytics used to set actuarial assumptions.

Pension actuaries have begun to analyze and model embedded options in employer benefit programs and potentially suboptimal choices made by plan participants. This is an emerging area for the use of predictive analytics in pension practice to set appropriate participant behavior assumptions. A related emerging use of predictive analytics is in the fields of pension risk transfer and longevity risk management.

Considerations in the Use of Predictive Analytics

Business Considerations

Before developing, implementing, and employing predictive analytics and other Big Data analytics models, companies need to carefully assess what their objectives are, what barriers they are likely to face, and how best to proceed. Barriers that need to be overcome include how to build the necessary infrastructure and synchronize it with existing infrastructure. Other issues that also need to be addressed include obtaining the expertise to effectively use predictive analytics, data availability, potentially conflicting priorities, and cost.

Because predictive analytics may involve multiple business functions and objectives, companies may wish to consider the benefits of developing a comprehensive and integrated strategy to support their efforts and a cost-effective means to test their strategy. Companies will need to develop robust sets of data principles to govern enterprise-wide handling of data. For example, it can be extremely challenging to harmonize, cleanse, and certify data from multiple internal (often legacy) and external systems. This is a critical step, especially when using data to derive assumptions used in financial reporting or for key company decision-making.

As companies aggregate data into data warehouses (often structured, more traditional data for reporting) and data lakes (often unstructured data combined with structured data), investments in data infrastructure are needed. Companies also will need to consider what else may change because of the use of predictive analytics. For example, if underwriting were to be streamlined, would changes to the application process be needed?

The evaluation of predictive models, important in actuarial professionalism, typically includes retrospective studies to measure model effectiveness and to establish criteria for when the new methods are used alone or in combination with old methods. Scenario analysis can aid in the determination of criteria that best align with companies' goals. Sensitivity tests can be used to assist in understanding variations in contributing variables and how interactions among those variables impact model outcomes.

After implementation, the model must continue to be monitored to measure its continued fit to new data. Does the model meet the objectives? Are the emerging results consistent with the projections based on the historical data on which the model was built? Is there any change in the strategy that may require the model to be adjusted? Must traditional methods be maintained to supplement some or all the new methods?

Companies also will need to address legal, regulatory, professional, ethical, and privacy concerns. These considerations typically are factored in before models are built, but, at a minimum, before implementation. Regulators may have questions about how predictive analytics-generated assumptions were demonstrated to be credible. Predictive analytics may be found to give more efficiently generated, evidence-based assumptions than traditional methods.

Model Development Considerations

There are many considerations in developing a predictive analytics model. Many of the considerations also apply when using more traditional analytical methods. The questions that might be asked include:

- Is the model appropriate for the situation for which it is being used?
- What are the evaluation criteria used to assess accuracy, effectiveness, and statistical appropriateness of the model?
- Is the data used in the analytical method acceptable to regulators? Some variables may not be allowed by current regulation.
- Is the data verifiable and credible?
- Is there a way for the policyholder to challenge and correct values?
- Is the relationship between predictor variables and the target variable intuitive? While causation is not a requirement of the actuarial standards regarding classification plans (there are generally four classifications for life insurance: preferred plus, preferred, standard plus, and standard), an attempt is generally made to explain the rationale for the relationship.
- Is the new variable replacing a previously used variable? When a new variable is replacing a historical variable, an explanation as to why this replacement is an improvement is generally developed. An example of such an improvement is the use of actual driving patterns from telematics devices replacing variables like age and gender. Clearly, the use of the actual driving experience is a better match to the expected claims than the historical rating variables of age and gender that have acted as proxies for driving behavior.
- Could the data variable be considered a proxy for a disallowed variable? Insurers are not permitted to use certain variables, such as race and nationality. However, there is a possibility that some other variables might be proxies for disallowed variables. Caution should be exercised to avoid using variables that may be considered as proxies for data elements not permitted, although determination of proxy status may not be feasible.
- How are missing values handled in the preprocessing stage of the data and/or in the modeling?

- What steps have been taken to ensure quality of the modeling data?
- How frequently will the values be refreshed? From an implementation standpoint, the modeler must decide on how frequently the model will be measured against new data to determine if the model needs to be “refreshed” or “rebuilt.” Refreshing a model involves updating the model with parameter estimates that result from running the algorithm on new data. A complete rebuilding of the model may become necessary if there are major changes in company underwriting, risk, or if environmental and behavior factors impact the level of loss experience.

Data Sources

The insurance industry has long relied on multiple sources of data. Emerging sources of data utilized in Big Data often are external to a company or can be internal data that previously was not available or difficult to extract. In legacy systems, for example, inconsistent sources and historical infrastructure may have created barriers to utilizing data. The explosion of structured and unstructured data availability, computing power, and new methods of data extraction provide for new opportunities regarding data collection.

Many observers believe that social media and consumer data may hold promise, but their lack of structure and the significant prevalence of missing data make them more difficult to process.

Specific data sources by area of practice are summarized below. In many instances, data sources are common among multiple areas of practice.

Life Insurance

Traditional data sources used for life insurance include the following:

- Experience study data, much of it coming from companies’ internal administrative systems, including the policyholder’s age, gender, account value, face amount, and other key customer and policy data. Policyholder use of elective benefits, death, withdrawal, and surrender/lapse data are also included in this category.
- Underwriting data that includes the policy application, attending physician statements, bodily fluids test results, Medical Information Bureau (MIB) information, and motor vehicle reports (MVRs).

Emerging data sources used for life insurance include the following:

- Data captured by sales and marketing to target customer segments, as well as customers within those segments.

- Electronic inspection reports for accelerated underwriting (AU) programs (i.e., underwriting without invasive testing such as fluids and exams).
- Other emerging data for underwriting includes public records such as bankruptcy filings and criminal history, demographic data, genetic information, credit scores, electronic medical records (EMRs), prescription histories, and lifestyle and behavioral data captured from wearables like Fitbit devices. Some of these are used for pre-policy-issue analytics, while others are used for ongoing monitoring. Some are used as part of formal underwriting and others highlight the need for additional analysis.
- Social media interactions including website clickstreams used both to verify underwriting data and as a lead-generation tool. For example, underwriters may check social media outlets, such as Facebook, Instagram, and Snapchat, for signs of nicotine use and other health-related information.
- Facial analytics and facial visuals to assist with identifying elements that were previous difficult to verify, e.g., smoking status.
- Income and wealth information for risk classification, marketing, and to assist with identifying lapse propensity.

Property and Casualty

The application of analytics for predictive purposes in the property/casualty (P&C) area of practice has been commonplace for some time and has become an important aspect of underwriting, ratemaking, and reserving. The data used for most P&C lines includes location and claims loss history, while other data is used specifically for the personal or commercial lines.

Traditional sources for P&C insurance include the following:

- For personal and/or commercial auto insurance—age and gender of the driver, type of vehicle, miles driven, as well as DMV information.
- For property insurance—type of construction, fire protection (e.g., smoke detectors, sprinklers), distance to water, and age of roof or utilities.
- For commercial liability insurance—the type of business being insured.

Emerging data sources for P&C insurance include the following:

- For some personal lines models, data sources that reflect more specific personal information. However, these variables are finding disfavor with some regulators, due to potential discrimination issues.
- For all lines of insurance, non-insurance information like weather data, crime statistics, population density, traffic density, and census information that might be predictive of claims.

- Telematics devices in cars that make detailed information about driver behavior easier to obtain. Telematics data has started to be used in rating and underwriting for personal and commercial auto.
- For many lines of P&C business, cellular technology, the Internet of Things and other advanced technologies, and new sources of data like home telematics and social media offer new insights into risk.

Health Insurance

Traditional data sources for health insurance include the following:

- Physician referral information or medical chart information, which can be useful in identifying diagnosis codes and other information about a patient.
- Enrollment information, including effective dates of coverage.
- Medical claims information, including diagnosis codes.
- Prescription drug claim information to provide additional insight into a patient's condition.
- Laboratory results information for understanding member outcomes, status, and morbidity.
- Self-reported data, such as from health risk assessments (although possibly not reliable because it is self-reported).

Emerging sources for health insurance include the following:

- Device-reported information, such as from wearable devices or home use devices.
- Electronic medical records, which are emerging as highly valuable information and often are used for risk adjustment supplemental information and audits. This data may be in a standard format or of an unstructured nature.
- Consumer and social media data, such as web searches.

Pension

Traditional data sources for pensions include the following:

- For pension plan design purposes—company-specific, proprietary, and confidential data, such as participant information.
- For projects involving new plan designs, assumption setting, and risk management—company- or client-provided proprietary data on plan participants.

- For models that use macroeconomic or geographic input—data from the Census Bureau and the Department of Labor, data from household surveys conducted by other government agencies such as the Centers for Disease Control and Prevention, longitudinal studies, as well as tax statistics available from the Internal Revenue Service.
- Company data on other retirement plans, such as 401(k), to be aggregated with traditional pension plan data for benefit adequacy analysis.

Emerging data sources for pensions include the following:

- Data available to a company from different benefit programs or from a different part of its business. For example, a data warehouse consisting of payroll and human resources data, pension administration information (defined benefit and defined contribution plans), and medical, dental, and disability claims can be constructed. Aggregating various existing data sources allows more patterns and relationships to be found via data analytics.
- Plan participant behavior, preferences, and the level of participant satisfaction from participant surveys or pension plan administration data. Pension plan administration data provided by record-keepers can include data across different employers, not just a company's own employees. Also, behavioral economists, who study the impact of psychological, social, cognitive, and other non-rational factors in the economic decisions of individuals, conduct research to identify factors influencing participant choice. The results of this research are useful in identifying attributes to use in predictive analytic models. A company can look for data associated with such attributes from its own data warehouse or from other data vendors.
- Consumer data, such as credit scores or consumer purchase patterns, and other forms of digital data, such as social media data, background checks, motor vehicle records, or facial analytics, for participant behavior modeling.
- Mortality data from broader public sources.
- For pension risk-transfer business—age, gender, benefit amounts, and actuarial assumptions associated with the group of plan participants in question. The emerging practice is to use other data available from a company's data warehouse or information from similar employers (usually provided by pension administrators or benefit plan consultants) to better assess the mortality experience of a group of plan participants, as well as the benefit options likely to be elected by plan participants.

Section II

Regulatory Considerations

Benefits and Challenges to Insurers, Regulators, and Consumers

Despite its potential, there are a number of concerns regarding Big Data that impact insurers, regulators, and consumers.

Insurers

The use of predictive analytics can lead to a better understanding of risk than traditional methods. New sources of data not only increase dimensionality of data dramatically, but also allow for the use of more direct indicators of individual risk. New methodologies allow for a potentially better understanding of risk drivers and relationships between them, as well as detecting potential fraud. The benefit of a better understanding of risk is protection against adverse selection and improved reserve adequacy, such as with health care models that can be used to more accurately predict utilization of health care services.

Potential drawbacks of new insurance models driven by predictive analytics include disruptions of the fundamental pricing principles of the industry, such as the collapse of the law of large numbers, disruptions in risk peaks and subsequent difficulty in assessing short-term risk, and premium inadequacy resulting from both new pricing models and substantial upfront build costs.

Regulators

Regulators may benefit from better advance knowledge of outcomes and could apply some predictive analytics techniques directly to their review processes. Potential benefits for regulators include the enabling of a more streamlined process for approval of pricing and rate filings as well as scanning of annual statement filings to detect previously unknown patterns. Regulators can also use predictive analytics to detect fraud.

The main regulatory rate standard in P&C rate making is that rates not be “excessive, inadequate or unfairly discriminatory.”³ Analytics that result in a premium that is more closely correlated with the future expected cost could assist regulators in ensuring that this standard is met. Additionally, the not unfairly discriminatory standard could be addressed with a more granular classification model that is supported by analytics. Increased solvency could result, to the extent that the analytics supporting a classification plan result in a better match of price to risk. The use of analytics may also increase competition resulting in better service (coverage options, claims settling, etc.) to policyholders.

However, risk pooling requirements in health insurance may not necessarily result in this type of additional benefit in rate setting due to restrictions in pricing and underwriting based on individual member characteristics.

Reviewing predictive analytics can be a challenge to regulators given the amount of data used to develop a model, the complexity of the techniques, and limited regulatory resources. Regulators also may have difficulty explaining complex models to consumers and other interested parties who are trying to understand the impact of the models on insurance rates. The NAIC’s Big Data (EX) Working Group is proposing additional support for regulators for reviewing new models that contain predictive analytics capabilities.

Consumers

Analytics can lead to more competition, and more competition can lead to more options for consumers. Predictive analytics can result in quicker decisions on underwriting, where allowable, because of the use of external data. Claim settlement can also be accelerated using predictive analytics. Analytics also can result in better offerings by insurers to policyholders from the use of external data that can help inform decisions regarding better fit of coverage.

The main challenge to consumers is lack of transparency: trying to understand the data and analytics being used to determine their eligibility for products and the price they are being charged. It may not be clear to the consumer how they are being underwritten or what behaviors they can modify or steps they can take to get a better rate. A potential issue with pricing based on predictive analytics is that it can lead to more granular pricing, which benefits some consumers but not others. This broader distributed range of prices could be perceived as unfair.

³ For example, see the NAIC’s Model Rating Laws (Property and Casualty Model Rating Law – File and Use Version (NAIC Model 1775); Property and Casualty Model Rate and Policy Form Law Guideline (NAIC Model 1776); Property and Casualty Model Rating Law – Prior Approval Version (NAIC Model 1780)).

Privacy issues are also a concern for consumers because of a lack of transparency regarding how data is collected and used. Consumers also may object to the use of some data either because they do not believe it is related to the cost of providing insurance, does not fairly distinguish risk, or because they do not believe the data is accurate. For example, the use of credit-related data in ratemaking for private passenger auto insurance is an example of data to which some consumers have objected, resulting in a variety of treatments from regulators ranging from complete prohibition in some states to allowing certain credit-related data in rating and underwriting in others.

Existing Regulatory Framework

The legal and regulatory requirements that potentially govern the use of Big Data by insurers at the state, federal, and international levels fall into two categories: 1) those designed to protect consumers in general; and 2) those intended to prohibit discrimination against certain protected classes of individuals.

Given the wide span of potentially applicable requirements, the following is a high-level overview of the legal and regulatory landscape. It is not intended to provide a comprehensive legal analysis of any laws or regulations.⁴

Consumer Protection Requirements

Consumer protection requirements cover a broad span of laws and regulations designed in a variety of areas. These requirements can be divided into privacy protections and general protections.

The collection and use of personal data by insurers is governed by privacy requirements that fall under regulatory review. In general, consumers have control over how their protected financial and health information, and other sensitive personal information, is shared by insurers with third parties. In addition, insurers may use consumer reports (as defined in applicable laws and regulations) only for specified permissible purposes. The increasing variety, velocity, and native digital format of available personal consumer data also are increasing focus on cybersecurity regulations and their connection to privacy concerns.

⁴ In 2017, the NAIC Big Data (EX) Working Group reviewed a summary of the “current regulatory frameworks used to oversee insurers’ use of consumer and non-insurance data,” focused primarily on the P&C insurance industry. See NAIC Big Data (EX) Working Group 2017 Summer Meeting Materials.

In terms of general protections, insurers overall must notify and explain adverse underwriting decisions to consumers. In addition, regulations exist that prohibit P&C and health insurers from charging excessive, inadequate, or unfairly discriminatory rates. Regulations also exist that prohibit life insurers from unfair rate discrimination between individuals of the same class and equal life expectation..

Examples of potentially relevant consumer protections include:

- **The Gramm-Leach-Bliley Act (GLBA).**⁵ Title V of GLBA includes specific rules governing how insurers may share and disclose consumers' personal information, including consumer reports and protected health information. The NAIC Privacy of Consumer Financial and Health Information Model Regulation⁶ implements the requirements of GLBA as they apply to insurers. Specifically, insurers are required to provide consumers with an annual privacy notice explaining the information collected, how such information is used and shared, and how it is protected. Subject to certain exceptions, consumers have the right to opt out of having their protected financial information shared with unaffiliated third parties and must opt in before their protected health information can be shared.
- **The Fair Credit Reporting Act (FCRA).**⁷ The FCRA regulates the use and dissemination of consumer reports. Users of consumer reports are subject to certain requirements under the FCRA, such as notice requirements for adverse actions with respect to insurance transactions based upon consumer report information.
- **European Union General Data Protection Regulation (EU GDPR).**⁸ The EU GDPR effective as of May 2018 is intended to simplify the regulatory environment across the EU and give more control to consumers over how their personal data is used by businesses. Companies governed by the GDPR, including companies based in the EU as well as companies collecting/processing data on EU residents, will have an obligation to erase data when customers ask to exercise their "right to be forgotten" and withdraw their consent to storing or using their personal data. The GDPR also requires companies to obtain explicit consent before collecting personal data.

⁵ 15 U.S.C. § 6801 et seq.

⁶ NAIC Model 672.

⁷ 15 U.S.C. § 1681.

⁸ Regulation (EU) 2016/679.

- **NAIC Insurance Information and Privacy Protection Model Act (the Model Privacy Act).**⁹ The Model Privacy Act governs insurers’ collection, use, and disclosure of consumer information in connection with insurance transactions. Among other things, it provides access to personal information and the consumer’s right to verify and correct such information. The Model Privacy Act also requires insurers to provide consumers with notice of the reasons for an adverse underwriting decision (or notice that such reasons can be requested).
- **Rate Regulation.** In the P&C space, state insurance laws and regulations ensure that premium rates—which can be developed using several different data sources—are not excessive, inadequate, or unfairly discriminatory. Additional requirements regarding the use and review of predictive models in determining rates vary widely by state and context. For example, certain states require P&C insurers to file predictive models used to determine premium rates, rating classes, etc.¹⁰ In addition, state and federal rate regulations in health insurance also limit the ability to use certain variables for rating, particularly in the individual and small group markets. And finally, the NAIC Model Unfair Trade Practices Act prohibits life insurers from unfair discrimination between individuals of the same class with equal life expectation.
- **Cybersecurity Regulation.** In early 2018, the New York State Department of Financial Services issued a first-in-the-nation regulation setting forth minimum requirements for covered entities to address cybersecurity risks. Covered entities must establish cybersecurity programs that address encryption, access controls, and limitations on data retention.

Anti-Discrimination Requirements

Anti-discrimination laws are meant to prohibit discrimination with respect to protected classes of people. State insurance laws include anti-discrimination requirements, and there are several federal anti-discrimination laws that could be relevant to insurers’ use of Big Data. Potentially applicable anti-discrimination requirements include, but are not limited to, the following:

- State insurance law anti-discrimination requirements: These laws prohibit unfair discrimination.¹¹

⁹ NAIC Model 670.

¹⁰ To illustrate, see NH Rev Stat § 412:16(II) (2016) (“Every insurer shall file with the commissioner every manual, predictive models or telematics models or other models that pertain to the formulation of rates and/or premiums, minimum premium, class rate, rating schedule or rating plan and every other rating rule, and every modification of any of the foregoing which it proposes to use.”); and NV Insurance Bulletin 17-001 (2017) (“The Division issues this bulletin to remind insurers that any mathematical model used in underwriting or rating of any personal line of property and/or casualty insurance, or other line of property and/or casualty insurance subject to regulation of rates pursuant to NRS 686B.030, must be filed with the Division for prior approval pursuant to NRS 686B.110.”).

¹¹ For example, N.Y. Ins. Law § 4224(a)(1) provides:

(a) No life insurance company doing business in this state and no savings and insurance bank shall

(1) *make or permit any unfair discrimination between individuals of the same class and of equal expectation of life, in the amount or payment or return of premiums, or rates charged for policies of life insurance or annuity contracts, or in the dividends or other benefits payable thereon, or in any of the terms and conditions thereof* (emphasis added).

- Discrimination based on sex, marital status, race, religion, and national origin also is generally prohibited.¹² In addition, certain state-specific requirements may apply.
- **Federal Laws:**
 - Equal Credit Opportunity Act: This prohibits any creditor from discriminating against any applicant based on race, color, religion, national origin, sex, marital status, or age. Title VII of the Civil Rights Act prohibits discrimination by covered employers based on race, color, religion, sex or national origin. Americans with Disabilities Act (ADA) extends the coverage of the Civil Rights Act of 1964 to Americans with disabilities.
 - Age Discrimination in Employment Act (ADEA): This forbids employment discrimination under certain circumstances against anyone at least 40 years of age in the United States.
 - Fair Housing Act (FHA): This makes it unlawful to refuse to sell, rent to, or negotiate with any person because of that person's inclusion in a protected class.
 - Genetic Information Nondiscrimination Act (GINA): This prohibits the use of genetic information in health insurance and employment.

Emerging Regulatory Developments

NAIC Activity (NAIC Big Data (EX) Working Group)

The evaluation of insurers' compliance with state law and regulation relies, in large part, on the information that is provided to regulators. This information can come from various sources, including financial statements, financial and market conduct examinations, filings, specific requests and data calls, or from statistical agencies.

Advances in statistical modeling techniques and evolving sources of data are challenging existing regulatory processes. Methods, such as those used to calculate premiums, are more complex than ever before. Current algorithms and models are not as easy to understand and follow as traditional algorithms. In addition, with the exploding availability of data, including consumer data, insurers are utilizing types of data not previously incorporated into advanced modeling techniques. Moreover, for many aspects of the insurance business, companies differ in methods and approaches employed and in their documentation and explanation of such methods and approaches.

¹² See NAIC Model 880, which prohibits "[r]efusing to insure, refusing to continue to insure, or limiting the amount of coverage available to an individual because of the sex, marital status, race, religion or national origin of the individual."

The complexity and evolution of the methods and approaches used by insurers is threatening to outpace the rate at which regulators can educate themselves on these new methods and approaches. Insurance regulators may choose to educate insurance department staff on these new techniques or employ external resources versed in techniques to evaluate of these new methods. From an insurer perspective, any delay on the review of new methods due to expertise limitations could result in reduced speed to market of innovations and new products, which could create a non-level playing field, allowing some companies to exploit regulatory shortfalls.

To address these issues, the NAIC has increased training opportunities, such as the predictive model training that was organized by the American Academy of Actuaries at the 2017 Summer NAIC Insurance Summit, and information-sharing forums to address current gaps in knowledge.

The NAIC also formed a Big Data (EX) Working Group (the Big Data WG). The Big Data WG's charges are to:

- “Review current regulatory frameworks used to oversee insurers’ use of consumer and non-insurance data. If appropriate, recommend modifications to model laws/regulations regarding marketing, rating, underwriting and claims, regulation of data vendors and brokers, regulatory reporting requirements, and consumer disclosure requirements.
- Propose a mechanism to provide resources and allow states to share resources to facilitate states’ ability to conduct technical analysis of, and data collection related to, states’ review of complex models used by insurers for underwriting, rating, and claims. Such mechanism shall respect and in no way limit states’ regulatory authority.
- Assess data needs and required tools for regulators to appropriately monitor the marketplace and evaluate underwriting, rating, claims, and marketing practices. This assessment shall include gaining a better understanding of currently available data and tools and recommendations for additional data and tools as appropriate. Based upon this assessment, propose a means to collect, house, and analyze needed data.”¹³

This Big Data WG recently proposed the exploration of a predictive analytics team staffed by the NAIC to provide predictive analytics modeling, insurance, and actuarial expertise to the states. The suggestion is that state regulators could rely on the expertise of the team to assist them in the review of advanced modeling techniques presented in insurance company models. The team would not opine on compliance with state laws or regulations but would serve in a technical advisory role at the request of a state regulator.

¹³ Big Data (EX) Working Group 2018 Charges; National Association of Insurance Commissioners.

Another recent proposal by the Big Data WG proposes the creation of a Predictive Analytics Working Group (PAWG). The PAWG would develop guidelines and processes to govern how state regulators would work with the team. An example of such a guideline would be a versioning system for company models, which would allow for the identification of company models previously submitted for a technical review. The objective is to have a more flexible and cost-effective resourcing approach for the states, bringing increased technical understanding to model reviews for the evaluation of state-specific laws and regulatory compliance.

Some of the concerns raised thus far include whether the NAIC will be able to obtain the necessary staff for such a team and the legality of housing such an organization within the NAIC; such concerns are currently under review. Beyond staffing and legal concerns, there are additional concerns regarding a centralized organization's ability to manage model versions, data security, models based on machine learning, and the protection of intellectual property.

Permitted Uses of Big Data

As regulation of Big Data evolves, defining what is and is not allowable—and what parameters and restrictions should apply under what circumstances—for insurance modeling and other uses of Big Data will be key decisions for legislators and regulators. An outstanding question from a regulatory perspective is whether, and to what extent, legislators and regulators will adopt different approaches with respect to:

- new uses of traditional data elements, such as using new types of models for mortality assumptions as opposed to a traditional actuarial actual-to-expected approach; and
- the introduction of new data elements, such as data from online shopping, social media, or telematics, into the insurance decision-making process.

The regulatory issues associated with the use of new data elements are potentially more complex. For example, driving telemetry data could include information on the specific roads traveled by an individual and the time at which they were traveled, which could pose issues from a privacy perspective.

Data Ownership, Transparency, and Portability

As the use of new data sources and analytic techniques increases and evolves, lawmakers and regulators will face difficult issues when crafting rules around how and when data can be owned, accessed, and transported.

Various models for governing the collection and dissemination of consumer data exist in different jurisdictions. For example, in the United States, consumers generally have the right to opt out of data collection or sharing activity.¹⁴ In contrast, in the EU, consumers generally must explicitly opt in before data can be collected or shared.¹⁵

Examples of potential regulatory questions with regard to data ownership, transparency, and portability include:

- Are existing privacy protections adequate?
- Should individuals “own” their data? To what degree should individuals have the right to access their own data? Who exactly should be able to access such data?
- Should individuals have the right to challenge, amend and/or correct their own data? Should there be limits on what can be corrected, e.g., medical diagnostic data?
- Should individuals have the right to “blur” their data (while also bearing the consequences of such blurring)? For example, in certain instances individuals can choose to limit their smartphone GPS location to a set radius to maintain their privacy. However, doing so renders pizza delivery and Uber/Lyft requests ineffective. This could have an unintended effect as those individuals willing to share more accurate data could end up with less expensive insurance coverage and/or enhanced benefits.
- Should individuals have the right to “transport” their data? Can an individual with auto coverage with one insurer take the personal data that the insurer has collected to a competing insurer to shop for a better quote? Current pricing is mainly driven by public information (accidents/violations), but if driving habits have been monitored, could that data be transferred? What are the possible effects on anti-selection and cost spirals?
- Are there relevant distinctions among different lines of insurance business that necessitate or justify different regulatory approaches or treatment?

¹⁴ 15 U.S.C. § 6801 et seq.

¹⁵ Regulation (EU) 2016/679.

Regulatory Sandboxes

“Regulatory sandboxes” have recently received significant attention from regulators, companies, and startups active in the financial services industries. Although the concept can take a variety of forms, a regulatory sandbox is generally a discrete regulatory environment designed to encourage innovation in a regulated industry. Depending on the context, a sandbox might function primarily as a forum for encouraging earlier and more frequent engagement between innovators and regulators, without necessarily allowing for waivers of existing law. Alternatively, a sandbox can relax regulatory requirements, effectively creating an alternative, less restrictive regulatory regime for proposed innovations. Given the regulatory issues involved, it is not difficult to imagine this concept being applied to insurance companies in the context of Big Data.

Several regulators have implemented some form of regulatory sandbox, both in the United States and internationally. For example, in the United States, the Consumer Financial Protection Bureau (the CFPB) and the Office of the Comptroller of the Currency (the OCC) each has projects designed to encourage innovation. The CFPB launched Project Catalyst in 2012. This project includes dedicated CFPB staff focused on encouraging innovation that is “safe and beneficial” to consumers. In 2016, the OCC announced a new framework designed to encourage “responsible innovation.” The framework includes the establishment of an OCC Office of Innovation with dedicated staff that will serve as a central point of contact for InsurTech innovators and will conduct outreach and provide technical assistance for InsurTech innovations.

In the context of the U.S. insurance industry, in 2017 the Illinois Department of Insurance proposed legislation that would have created a new “Innovation Division” within the insurance department and granted this division broad authority to support the development of insurance innovations and assist insurers with compliance.¹⁶ As of the publication of this monograph, this legislation has not been acted upon.

A major reinsurance company has proposed a Future Insurance Technology Lab (FITLab) framework to the NAIC. The FITLab is intended to serve as “a ‘safe space’ for open communication between industry and regulators surrounding new innovative efforts.” It would create a confidential forum at NAIC meetings during which companies could discuss and receive feedback on proposed innovations from a working group of state regulators.¹⁷

¹⁶ Illinois SB3451 (2018).

¹⁷ “Proposal: Future Insurance Technology (FIT) Lab, a US regulatory sandbox solution,” Kelsey Brunette, Munich Re America Inc., Nov. 28, 2016.

There are still a number of open questions around the FITLab and the regulatory sandbox concept in general, such as how long the “innovation waiver” would last or how material the innovation needs to be.

In the United Kingdom, the Financial Conduct Authority, the primary financial product and market conduct regulator, launched an innovation project in 2014 and created an “Innovation Hub.” If an innovator demonstrates that it is developing a real innovation that benefits consumers, it can apply to receive dedicated support and feedback from Innovation Hub staff.

Potential Regulatory Disruptions

In any regulated industry, changes in business practices may evolve so quickly that regulators, and regulation, will need to sprint to keep pace. Big Data is already accelerating the pace of change in certain aspects of the insurance business.

The development of accelerated underwriting (AU) in the life insurance industry—made possible in large part by the availability of new data sources and analytic techniques—and the associated reserving implications under the NAIC’s PBR framework are a useful example. Guidance set forth in the initial PBR valuation manual did not anticipate the use of Big Data and the emergence of AU, so it did not address the question of how reserving standards should incorporate AU. Regulators are working on bridge solutions for 2018 and beyond.

In other instances, it is possible that a regulated entity, or possibly a startup, may follow the examples of Uber and Airbnb and bring a new solution to market irrespective of existing regulatory protocols or the fundamental permissibility of the solution. This could create unintended regulatory consequences for traditional insurers.

These events could impact the insurance business model via changes in the distribution model (e.g., robo-advisers, social media advertising, smartphone tie-ins), changes in coverages, changes in premium and claim payment practices, and operational risks, among others. Based on experience in the P&C insurance and other industries, some of the potentially critical success factors for these innovative approaches include the following:

- Are the offerings voluntary?
- Do they create clear value for consumers?
- Do the offerings elicit a groundswell of public support?

Some conceivable examples of potential disruption include the following:

- Offerings may cross regulatory boundaries, such as a FinTech¹⁸ company providing long-term insurance coverage or auto insurance rates based on savings account balances.
- Driverless cars may move regulators to mandate commercial insurance rather than personal insurance coverage.
- Offerings of “all-in-one” risk packages for a major portion of the life cycle may become available.

Insurers will need to consider the regulatory response to their use of Big Data and what level of regulatory risk they are prepared to assume. There is currently considerable uncertainty in the industry around how insurers’ use of Big Data will be regulated. Meanwhile, many companies continue to make significant investments in InsurTech, new models, and Big Data infrastructure. To help limit potential losses and foster the confidence needed for insurers to continue to invest in Big Data, lawmakers and regulators will need to watch these developments carefully and be prepared to respond quickly.

Regulatory Challenges

Regulators will continue to face challenges as they review and respond to insurers’ evolving uses of Big Data. The following highlights important challenges, which often have professionalism considerations as well (outlined in Section III):

- a. **Privacy.** As insurers’ collection and use of data evolve, insurance regulators seek to better understand company algorithms and the types of data used for areas in which regulatory and legal review is necessary. To provide state-of-the-art products, many insurers are investing heavily in data, technology, and related resources. Given the competitive nature of the marketplace, insurers often are reluctant to share data-related intellectual property and market insights with regulators, which can create challenges for regulators trying to understand evolving practices. The degree of protection afforded under state freedom-of-information laws varies substantially by jurisdiction and often does not provide sufficient protections from insurers’ perspectives. Stronger privacy protection for Big Data information might increase transparency and thereby enhance regulators’ understanding of evolving practices and facilitate better regulation.

¹⁸ FinTech stands for financial technology, and in its broadest definition, it is “technologies used and applied in the financial services sector, chiefly used by financial institutions themselves on the back end of their businesses. But more and more, FinTech is coming to represent technologies that are disrupting traditional financial services, including mobile payments, money transfers, loans, fundraising, and asset management.” See: “The Complete Beginner’s Guide to FinTech in 2017”; *Forbes*; Feb. 10, 2017.

In addition, as both the uses and complexity of data grow, consumer opinion may increasingly influence regulators' views and reactions. For example, if individuals believe that the use of certain types of data is inappropriate, regulators may need to understand and account for these expectations of consumer privacy.

- b. **Staffing.** Shortages of staffing and expertise for regulators will likely prove problematic given the increasing complexity of data and approaches. To address this, proposed addition of NAIC staff that could support technically rigorous and data-intensive reviews could facilitate a more efficient use of regulatory resources.
- c. **Correlation vs. Causation.** If individuals and competitors do not know their risk exposure versus others, then large heterogeneous pooling works well. As insurers identify behaviors (or controllable risk drivers) through empirical research or data analytics, insurers can signal to the market how to lower collective risks or appropriately charge those who take on riskier behavior. For example, owners of commercial buildings understand the value of automatic sprinklers, which result in lower insurance premiums and claims. Individuals who smoke are charged for their elected riskier behavior. However, predictive analytics can only uncover correlations among data elements. These data elements may be driven at a deeper level by other factors. Both insurers and regulators will need to ensure that spurious correlations are not driving pricing and coverage decisions. For those events where the true drivers are not known, risk pooling can be used to smooth out the impact of costly events randomly striking members of a group.

The American Academy of Actuaries has historically worked closely with regulators and policymakers in providing objective, unbiased, and nonpartisan insights into issues of an actuarial nature. In these interactions, these parties have relied on the professionalism and technical skills of actuaries to provide clear information for the benefit of the public.

Section III will address professionalism considerations for actuaries working with Big Data. As Big Data continues to evolve, the Academy will continue to work with regulators and the public to provide insights and information to address the challenges that Big Data may present.

Section III

Professionalism

Actuaries have professional obligations to uphold the reputation of the actuarial profession and fulfill the profession's responsibility to the public in the emerging area of Big Data. An important part of this responsibility is to comply with the law. In many situations, actuaries also have unique insights into the results and implications of the use of Big Data and must be willing and capable to explain such insights, where appropriate, to the key stakeholders of the work, such as regulators, consumers, company management, auditors, etc. The value of the actuaries' work is enhanced through adherence to the Code of Professional Conduct, actuarial standards of practice, and U.S. Qualification Standards. A key attribute of the applicable standards is the requirement for actuaries to provide explanations and rationales for their conclusions.

Professional judgment from actuaries is critical in the utilization of Big Data in actuaries' work. Actuaries provide added value to Big Data work in their ability to "connect the dots" through a deep understanding of the subject matter. In exercising professional judgment, it is important for actuaries to be cognizant of the fact that without performing proper analyses or validation, the results of Big Data can be misleading. A combination of a good understanding of the context in which the data was obtained and avoidance of unthoughtful adherence to the results of a model can aid in better Big Data outcomes.

It should be noted also that "spurious correlations" that might be exhibited in a Big Data analysis do not imply causality. There are many examples of two pieces of data that are very closely correlated over a period of time that do not have a causal relationship. While causality is not a requirement for the application of Big Data analytics, users of Big Data should be aware of that these correlations exist.

There are many professionalism issues that may be encountered when working with Big Data and predictive analytics. The work of actuaries is governed by the Code of Professional Conduct (Code) and must comply with applicable actuarial standards of practice (ASOPs). The Code and ASOPs provide a framework for dealing with issues of professionalism that might arise in the work of actuaries. While actuaries have traditionally dealt with large volumes of data and a variety of modeling techniques, Big Data may pose new challenges

that differ from those that actuaries encountered in the past. In addition, actuaries historically have built analyses and models based on traditional inferential statistical methods (descriptive and diagnostic analytics); however, predictive analytics techniques offer unique and different challenges to consider. Some professional organizations, such as the Data Science Association, have codes of conduct that apply specifically to the key elements of Big Data, such as data quality, volume, variety, and associated analytical techniques. For instance, data scientists must “use reasonable diligence when designing, creating and implementing machine learning systems to avoid harm.”¹⁹

This section reviews the professionalism requirements for actuaries working with Big Data and engaging in predictive analytics. Some professionalism and ethical issues that arise in this context are also highlighted.

Actuarial Professionalism

Code of Professional Conduct

In 2001, the five U.S.-based actuarial organizations adopted a consistent Code of Professional Conduct. The Code sets forth what it means for an actuary to act as a professional. It identifies the responsibilities that actuaries have to the public, to their clients and employers, and to the actuarial profession. The purpose of the Code is to require actuaries to adhere to standards of conduct, practice, and qualification. The Precepts of the Code identify the professional and ethical standards with which an actuary must comply to fulfill their responsibility to the public and the actuarial profession. The law (i.e., statutes, regulations, judicial decisions, and other statements having legally binding authority) may impose additional obligations upon an actuary. Where requirements of law conflict with the Code, the requirements of law shall take precedence. Many of the 14 Precepts in the Code will have relevance to work performed related to Big Data.

Several Precepts deal with general conduct issues that apply to every service provided by actuaries, such as acting honestly, with integrity and competence; using titles and designations only as authorized by the relevant actuarial organization; prohibitions against disclosing confidential information; and requirements to cooperate with others. Most of the Precepts focus on the conduct of an actuary when providing actuarial services. The Code defines actuarial services as “Professional services provided to a Principal by an individual acting in the capacity of an actuary. Such services include the rendering of advice,

¹⁹ *Data Science Code of Professional Conduct*; Data Science Association.

recommendations, findings, or opinions based upon actuarial considerations.”²⁰ An actuary will need to consider whether the Code applies to their performance of services that involve Big Data based on whether those services meet the definition of actuarial services and if a particular service involves actuarial considerations. Consider a marketing effort to gain new customers that uses predictive analytics to determine the customers who would be most likely to buy an insurance product. Actuarial considerations for such an effort might include data quality, appropriateness of use, and the accuracy of predictive results.

Actuarial Standards of Practice

Precept 3 of the Code requires an actuary to ensure that actuarial services performed by or under the direction of an actuary satisfy applicable standards of practice. In the United States, the applicable ASOPs are promulgated by the Actuarial Standards Board (ASB). When a question arises about the applicability of a standard of practice, or where no applicable standard exists, an actuary shall utilize professional judgment, considering generally accepted actuarial principles and practices.²¹ When an actuary uses procedures that depart materially from those set forth in an applicable standard of practice, the actuary must be prepared to justify the use of such procedures.

A full treatment of the relevant sections of each of the ASOPs is beyond the scope of this paper. Following are some of the ASOPs that may be relevant to services involving Big Data. Further details regarding these ASOPs are included in Appendix 2.

1. ASOP No. 23, *Data Quality*, provides guidance to actuaries when selecting data, performing a review of data, using data, or relying on data supplied by others in performing actuarial services. It also applies to actuaries who are selecting or preparing data or who are responsible for the selection or preparation of data that will be used by other actuaries in performing actuarial services when making appropriate disclosures regarding data quality.
2. ASOP No. 12, *Risk Classification (for All Practice Areas)*, applies to all actuaries when performing professional services with respect to designing, reviewing, or changing risk classification systems used in connection with financial or personal security systems regarding the classification of individuals or entities into groups intended to reflect the relative likelihood of expected outcomes.

²⁰ American Academy of Actuaries; *Code of Professional Conduct*; Jan. 1, 2001.

²¹ *Ibid.*; Precept 3, Annotation 3.1.

3. ASOP No. 38, *Using Models Outside the Actuary's Expertise (Property and Casualty)*, applies to actuaries who use models that incorporate specialized knowledge outside of the actuary's own area of expertise when performing professional services in connection with property and casualty insurance coverages. This standard applies to the use of all models whether or not they are proprietary in nature.
4. ASOP No. 25, *Credibility Procedures*, applies to actuaries when performing actuarial services involving credibility procedures: a) when the actuary is required by applicable law to evaluate credibility; b) when the actuary chooses to evaluate the credibility of subject experience; c) when the actuary is blending subject experience with other experience; or d) when the actuary represents the data being used as statistically or mathematically credible.
5. ASOP No. 41, *Actuarial Communications*, provides guidance for preparing actuarial communications within any practice area. Included in this guidance are requirements regarding: a) form and content; b) clarity; c) timing of communication; and d) identification of responsible actuary. Additionally, guidance regarding disclosures with an actuarial report, explanation of material differences, oral communications, responsibility to others, and retention of materials are included.
6. ASOP No. 21, *Responding to or Assisting Auditors or Examiners in Connection with Financial Audits, Financial Reviews, and Financial Examinations*, applies to actuaries when performing actuarial services as a responding actuary or as a reviewing actuary in accordance with generally accepted auditing standards or a financial examination for the purpose of oversight of the financial condition of an entity. An actuary needs to be sensitive to the possibility that when Big Data and predictive analytics are used for financial reporting purposes, the responding actuary may have to explain the use of Big Data to the reviewing actuary.

The examples of applicable ASOPs are not exhaustive. Other ASOPs may be applicable depending on the assignment. As the use of Big Data and predictive modeling continues to evolve, it is possible that it will become the basis for developing actuarial assumptions or contribute to the construction of models or be integrally involved in pricing and ratemaking or the evaluation of risks in general. With these innovations, the actuary would be well served to understand the implications, benefits, and considerations in using Big Data and predictive modeling.

Qualification Standards

Precept 2 of the Code states that “An Actuary shall perform Actuarial Services only when the Actuary is qualified to do so on the basis of basic and continuing education and experience, and only when the Actuary satisfies applicable qualification standards.” Annotation 2-2 goes on to state: “The absence of applicable qualification standards for an assignment or for the jurisdictions in which an Actuary renders Actuarial Services does not relieve the Actuary of the responsibility to perform such Actuarial Services only when qualified to do so in accordance with this Precept.”²² The actuary should always reflect on their qualifications, and must be prepared to document their qualifications (USQS Section 6.2) for any project being undertaken, and Big Data/predictive analytics projects are no exception. As an evolving area, it may not always be a clear-cut determination, and professional judgment may need to be applied.

In addition, U.S. Qualification Standards section 4.3²³ addresses emerging or nontraditional areas of actuarial practice. It states that an actuary practicing in an emerging or nontraditional practice area can satisfy the continuing education requirements by maintaining knowledge of applicable standards of practice, actuarial concepts, and techniques relevant to the topic of the Statement of Actuarial Opinion.

Ethical Considerations

Many actuaries are well equipped to integrate innovative analytics with traditional actuarial practices. A new paradigm involves a demand for new skills and can raise a wide range of ethical and professional challenges. The Code and the ASOPs guide actuaries in navigating these challenges, and dealing with new implications, while continuous education and the highly developed quantitative skills of actuaries can aid them in acquiring new skill sets and staying abreast of emerging technologies.

The traditional “look in the mirror” test (which is implied but not spelled out in the Code) means that an actuary objectively examine his or her qualifications (basic and continuing education and experience) and make a professional judgment about whether the actuary can fulfill the actuary’s obligations under the Code to:

- Act honestly, with integrity and competence—perform actuarial services with skill and care (Precept 1); and
- Perform actuarial services only when qualified to do so (Precept 2).

²² Ibid.

²³ Available at actuary.org/usqs.

Algorithms, Techniques, Correlation, and Causality

This section reviews the potential professionalism issues that may surface when using Big Data and predictive analytics in any actuarial area.

Many newly introduced methodologies, whether previously employed in other professions or recently developed, represent sophisticated models that borrow from other areas of science, such as artificial intelligence. Some methodologies involve extremely difficult and complex mathematics that may require someone specifically trained in that area. Other models may be hard to interpret, even if fully understood. This could result in what is perceived as nontransparent outcomes.

To the extent that an actuary employs a model, the actuary's level of effort in understanding and evaluating a model should be consistent with the intended use of the model and its materiality to the results of the actuarial analysis. At times an algorithm or model may lack transparency or may not exhibit a clear connection between the input and output. If the application of an algorithm or model results in an outcome that regulators or others perceive as unfair or unfairly discriminatory, its use may be restricted or disallowed. As noted in Section II, the actuary should be aware of regulators' concerns that a variable could be considered a proxy for, or be correlated with, a prohibited factor.

Actuaries often are asked to lead projects that utilize predictive models. ASOP No. 38, although referenced as a property and casualty ASOP, may provide some guidance beyond P&C work, as it contemplates that actuaries may make use of a model that is outside of their area of expertise.²⁴ In addition, a revised version of ASOP No. 38 is pending that would cover all practice areas. The current ASOP No. 38 requires the actuary to:

1. Determine appropriate reliance on experts;
2. Have a basic understanding of the model;
3. Evaluate whether the model is appropriate for the intended application;
4. Determine that appropriate validation has occurred; and
5. Determine the appropriate use of the model.

Understanding what an actuary's responsibilities are and what roles the actuary plays on the predictive analytics team is key. These are important professionalism questions for the actuary who may not have an explicit role or defined responsibility in the development or use of the models but who nonetheless has some implicit level of professional or ethical responsibility.

²⁴ ASOP No. 1, Section 4.3 states: "An ASOP should not be interpreted as having applicability beyond its stated scope and purpose. ... If no ASOPs specific to the task are applicable, the actuary may, but is not required to, consider the guidance in related ASOPs."

There are currently no ASOPs specifically dealing with Big Data or predictive models that differ in material aspects from traditional actuarial methods, models, and techniques. Consequently, users of such models may choose to look to ASOP No. 38 or, if they are performing services in connection with P&C insurance coverages to which ASOP No. 38 applies, they will need to justify any material deviation from the obligations identified in ASOP No. 38.

For example, in employee benefit plan designs, if an actuary is unfamiliar with the algorithms used to model employee behavior, employee preference, and employee choice and those considerations are material to the actuary's work, ASOP No. 38 may provide useful information in terms of model evaluation, validation, and documentation. The actuary's work product may not involve the creation of such models, but their use could impact the actuary's work, assumptions, or communications.

Applications of Big Data can be useful in identifying correlations based on patterns discovered by analyzing data that tracks well with the behavior of individuals. In some cases, however, the correlation indicated by the data might be coincidental or there may be a confounding factor—i.e., a spurious correlation. This may suggest an algorithm problem. Actuaries working in this area need to ensure that specialists who analyze the data and build the models/algorithms have appropriate training and use the tools and procedures to test and correct for issues such as spurious correlations. For example, following standard model-building practices such as data partitioning with training, validation, and testing sets will most likely identify and eliminate such spurious correlations. Without correcting for spurious correlations, undesirable results may occur.

Underwriting is an area where it is important to understand the distinction between correlated results and causal relationships. While actuarial standards do not require an actuary to establish a causal relationship, many regulators have, for public policy reasons, disallowed the use of underwriting indicators unless it can be shown there is a causal relationship with the insurance claims that might occur under the insurance contract. In some cases, causal relationships are self-evident or can be presumed or explained. In other cases, such causal relationships can be demonstrated with data and analyses. However, there can be cases where the relationship is subject to some uncertainty about the validity or the quantification of the relationship, and the underwriting indication may not be allowed.

Algorithms can be used in the underwriting process to assign a policyholder to a risk class and/or rate class. Generally, such assignments must be objective, transparent, and explainable to regulators and to insureds. There can be regulatory, statutory, or other legal restrictions regarding explanations and justifications of ratings and risk class assignment.

Data analytics also brings the potential benefit of uncovering previously unknown or hidden relationships in highly dimensional data. Once indicated by the data analytics, such relationships or correlations may indicate a need for further investigation. In health insurance, data analytics may suggest that a gap in diagnostic coding of a condition may exist as part of a risk adjustment program, when the condition that appears to be missing a diagnostic code in claims may not actually exist. For example, if prescription drug claims are used to determine potentially missing diagnoses in medical claims and an asthma medication claim is present without a diagnosis in the medical claims data, it may suggest that a gap exists for the asthma condition. However, some asthma medication also is used to treat chronic obstructive pulmonary disease (COPD) and, if this is the case, the model's result may be erroneous. The descriptive and predictive models, consequently, may provide opportunities for identifying potential issues that can be researched through review of medical records, or through a care coordination visit, or further investigation into potential waste, fraud, or abuse. If the method or approach does not result in an unsupportable action, the algorithm can be tested for its ability to be a good predictor, and adjusted as necessary.

Using Big Data for claim/care management outreach may give an incomplete or even an inaccurate picture of the issues a member may have. For care management efforts in health insurance, outreach on asthma education or disease management programs may be inaccurate if the member is using an asthma medication for treatment of COPD. It is important for the actuary to be aware of the correlation of the data to other potential causes before using the information. Often Star Ratings in Medicare Advantage and Prescription Drug and Affordable Care Act business for health insurance are used to measure how well a plan performs in several categories, such as quality of care and customer service, include patient satisfaction scores. If outreach is performed based on an inaccurate result from an algorithm, this can lead to patient dissatisfaction and lower Star Ratings of a plan.

The use of Big Data models is an extension of the traditional work of the actuary governed by the Code and ASOPs. There are several challenges not seen in traditional actuarial work including, but not limited, to:

- Reliance on and the need to supervise the work of other technical experts;
- Drawing conclusions from correlated relationships without clear evidence of a causal relationship; and
- Public policy concerns regarding the use of personal data.

These challenges require the actuary to carefully consider the professionalism and ethical considerations associated with these data models in ways that may not apply in traditional actuarial work.

Role of the Actuary

In many applications of Big Data in businesses in which actuaries are employed, multidisciplinary teams are utilized to efficiently and effectively complete the project. The teams are commonly composed of statisticians, computer scientists, data scientists, and actuaries. Actuaries on these teams may be thought of as the subject matter experts. But actuaries may be positioned to be the quarterbacks of the Big Data teams. With the proper background, an actuary can understand and direct the work of the Big Data multidisciplinary team based on their professionalism requirements and subject matter expertise.

As the evolution of Big Data continues in the areas of practice in which actuaries provide services, the professionalism and technical expertise provided by actuaries are essential elements upon which the public and regulators can place reliance. The professionalism requirements of actuaries provide guidance for the proper application and disclosure of Big Data assumptions and methodologies. They require actuaries to adhere to the high standards of conduct, practice, and qualification of the actuarial profession, thereby supporting the actuarial profession in fulfilling its responsibility to the public.

Appendices

Appendix 1: InsurTech

InsurTech is a blending of the words “insurance” and “technology.” It is the insurance industry analog of the term FinTech, a blending of the words “financial” and “technology.” The application of InsurTech is marked by the innovative use of technology to transform the insurance customer’s buying, underwriting, and in force management experience by replacing traditional constructs of insurance with technology-driven systems that use predictive analytics and are often independent of the traditional approaches.

InsurTech innovations continue to occur at increasing rates of speed throughout the insurance marketplace, ranging from marketing to claims, and including financial management, although the current focus is significantly on marketing and distribution. These innovations are happening in all lines of insurance business.

Below are three examples of ways in which InsurTech is transforming the industry:

- Insurance companies are changing the customer buying experience through InsurTech applications. Under one such app-driven product, underwriting utilizes Big Data-based algorithms to issue policies in less time than consumers have experienced under traditional underwriting. This company primarily targets Millennials, an app-driven generation that cares about causes. The company donates a portion of their revenues to charities insureds elect through the app-mediated application process.
- Life insurance companies are deploying life insurance applications using InsurTech devices and approaches. For instance, one company has deployed InsurTech processes to speed up the issuance of life insurance policies and another introduced a program that integrates InsurTech technologies with its life insurance products.
- Attracting and retaining new customers is a top priority of some insurers using technology-driven devices to transform the customer engagement relationship. InsurTech consulting firms are cropping up in the life insurance space to address the challenges insurers are facing to understand the evolution currently taking place in the marketplace.

Momentum in the industry is growing to increase the capitalization on the benefits of InsurTech both for additional functionalities and in other insurance practice areas.

The “how” of InsurTech, like FinTech, is highly dependent upon Big Data sources and Big Data analytics, such as predictive analytics. The most pervasive examples of InsurTech applications include wearable devices, telematics devices, customer technology apps, data portals, and platforms. Innovative InsurTech applications utilize predictive and artificial intelligence methodologies and technologies that simplify underwriting algorithms, and improve claims management, retention, targeted marketing, and other processes after issue. Companies are measuring the accuracy of traditional models against Big Data-based models and often finding the latter just as accurate, if not more so—and, more importantly, significantly less expensive than traditional models. Additionally, many real-time analyses that previously could not be performed are now performed using predictive analytics.

InsurTech approaches deploy Big Data to manage, expand, and remediate, if necessary, the customer experience and other aspects of insurance transactions, as well as insurance company management and strategy, often with significant savings and efficiencies. However, infrastructure changes to manage Big Data capabilities can involve large investments.

The driving force behind the development of InsurTech companies is the belief that the insurance industry is ripe for innovation and disruption. One force driving this disruption is behavioral. Millennials pursue a different consumer engagement paradigm than prior generations. The following generations will be even more media-enabled, forcing additional evolution in how companies engage consumers, simplify the issuance of policies, and manage those policies after issue.

The offering of ultra-customized policies, social insurance, and new streams of data from internet-enabled devices characterize the market approach of InsurTech companies. In addition to new pricing models, InsurTech startups are testing deep learning-trained artificial intelligence models to handle the tasks of brokers and find the right mix of policies to complete an individual’s insurance coverage. There is interest in the use of apps to pull disparate policies into one platform for management and monitoring, creating on-demand insurance for micro-events like borrowing a friend’s car and the adoption of the peer-to-peer model to both create customized group coverage and incentivize positive choices through group rebates.

The industry may be ripe for these innovations, but incumbent players are sometimes reluctant to adopt them. Insurance is a highly regulated industry with many layers of jurisdictional legal limitations. Regulators are still developing the expertise to regulate the use of Big Data in the context of insurance. Thus, they may be resistant to relaxing

regulations before they fully understand predictive algorithms. Insurance companies may err on the side of caution and shy away from startup ventures rather than risk regulatory challenges.

Many InsurTech startups still require the help of traditional insurers to handle underwriting and manage catastrophic risk. In addition, change always requires a transformative mindset. However, insurance is dependent upon consumers, and as more InsurTech capabilities garner consumer interest with a more refined, tech-enabled, and user-friendly approach, insurers will likely embrace the idea of InsurTech, buying up some of the innovations or creating their own innovations.

Observations

While innovations come with rewards, they also involve risks. There is a need to evaluate the risks these innovations pose to the financial standing of insurance organizations. The following are some key observations of the potential impact of emerging insurance technologies on life, health, pension, and property and casualty insurance.

Observation 1: The distribution of many insurance products is moving away from the traditional and exclusive agent/broker-policyholder relationship toward a more impersonal, internet-based relationship. This will likely benefit insurers in the following ways:

- Provide significant strategic advantage to those companies that effectively, and in a timely manner, deploy its use. InsurTech companies can provide significant guidance as to how insurance companies can market better and more cost-effectively;
- Improve how insurance companies manage their in-force blocks of business; and
- Motivate regulators to develop Regulatory Technology (RegTech) to monitor the use of InsurTech.

Observation 2: For insurers, the key risks associated with the emergence of InsurTech include data privacy, regulatory compliance, product marketing, cyber fraud, and operational, underwriting, and strategic risks.

Observation 3: Insurers adopting and leveraging advanced technologies to deliver innovative insurance products face the risk of conflicting outcomes derived from the used of technologies such as artificial intelligence, machine learning algorithms, and natural language processing techniques. Cloud computing services pose a unique risk associated with unauthorized sharing of consumer data.

Observation 4: The increasing use of third-party data to reduce and simplify traditional underwriting methodologies poses risks to post-claim review processes for insurers, especially within the contestable period. It may also be more difficult to use claims experience as a learning tool for the underwriting process.

Observation 5: InsurTech developments may increase the scrutiny of insurer market conduct and operations by regulators as nontraditional data sources may contain proxies for variables disallowed by regulators. In addition, the technologies will likely undergo scrutiny by regulators to ensure similar outcomes for similar risks.

Observation 6: Regulators will need to augment their skill sets to supervise the use of InsurTech, advanced modeling techniques, and Big Data by insurance companies. Insurers and regulators likely will need to strike a balance between regulatory supervision and industry innovation to deliver an improved level of services to consumers at competitive costs.

The observations provide insight into how InsurTech will likely transform the insurance industry. They do not directly address risks that are a function of how the technologies were developed or the standards by which these technologies are evaluated against model risk and validation criteria. The following outlines considerations for assessing InsurTech vendor risk and developing model risk and validation criteria.

InsurTech Vendor Risk

Many companies (InsurTechs) have been formed in recent years that focus on leveraging technology to address the issues and opportunities presented to insurers. These InsurTechs are vendors to insurance companies as the insurance marketplace and regulators take up these innovations. Considerations for working with InsurTechs follow.

Product Quality

Criteria must be established to assess the quality of the InsurTech startups and the products they can potentially offer insurers. Areas important in assessing quality might include:

- Insurance product expertise;
- Quality of company management;
- Insurance-backed funding sources;
- Knowledge of insurance distribution channels;
- Financial strength to suggest industry sustainability;
- Understanding of the regulatory insurance environment and privacy issues; and
- Demonstrated proficiency developing tech-based customer engagement media.

Integration and Maintenance

A significant problem with any technology is its susceptibility to obsolescence. It can be very costly and resource-intensive for companies to integrate innovative technology with existing company systems. However, the integration of digital technologies can help insurers develop the following:

- Advanced methodologies to exchange data between facilities;
- Advanced machine learning analytics capabilities; and
- Ability to identify and acquire new sources of consumer data.

External Data Dependencies

The main concerns involve the consistency of data from a myriad of sources and how to measure the impact of data inconsistency on models and ultimately the consumer. Specific considerations include:

- The credibility, validity, and traceability of data sources;
- The independent validation and reconciliation of data sources;
- The epoch of data sources and alignment to measures assessed by models; and
- The validity and review of underlying models generating external data sources.

Compliance Standards

The advent of the age of Big Data has challenged regulators with issues that current regulations are not equipped to address. Regulators are rapidly augmenting their education and regulatory tools to deal with the following:

- Privacy issues poised by the inclusion of Big Data sources in models;
- Ethical issues raised using Big Data in models impacting consumers;
- The inclusion of variables in Big Data masking disallowed variables;
- The reconciliation of consumer risk metrics derived from different models and data sources, and across different geographies; and
- The structuring of modeling data sets to assess geographical influences.

Model Risk & Validation

As with any innovation, Big Data represents an unexplored frontier for insurers, regulators, and consumers. Every model poses a certain amount of model risk to an organization.

Model risk can be introduced through such things as:

- Applying models incorrectly;
- Using improper models;
- Developing inaccurate conclusions; and
- Utilizing improper data.

Other forms of model risk can be introduced through items that are uniquely associated with Big Data. InsurTech vendor models use Big Data and technology for driving decisions based on data rather than traditional underwriting methods. However, the validation methodologies of InsurTech technologies are still developing. Some considerations in the development of validation methods might include the following:

- Controls around authorized access and authorized use;
- Controls around the proper operation of InsurTech technologies;
- Assessing controls around data transmission and security from hacking;
- Validation of underlying algorithms and temporal consistency of results; and
- Analytical and surveillance tools to trigger alerts to refresh or rebuild models.

It is unlikely that the use of Big Data will become obsolete. The insurance industry will need to develop model governance policies and standards of practice to monitor the use and application of InsurTech technologies, as well as to collaborate with the regulatory community on issues that these innovations raise.

Appendix 2: Actuarial Standards of Practice (ASOPs)

A full treatment of the relevant sections of each of the ASOPs²⁵ is beyond the scope of this paper. The pertinent sections of some relevant ASOPs are highlighted and commented on in the following. This list is not intended to be exhaustive or all inclusive.

1. ASOP No. 23, *Data Quality*

Section 4.1.g states: [An actuarial communication should disclose when material and relevant] “the existence of results that are highly uncertain or have a potentially significant bias of which the actuary is aware due to the quality of the data or other information relevant to the use of the data, and the nature and potential magnitude of such uncertainty or bias, if they can be reasonably determined...”

Big Data cannot be expected to be completely error-free. Data may come from third-party sources or may require frequent updating in near real time for use in certain applications. Section 4.1.g is just one of the 11 disclosure requirements in the ASOP. The disclosures in ASOP No. 23 tie into ASOP No. 41, *Actuarial Communications*.

2. ASOP No. 12, *Risk Classification (for All Practice Areas)*

Section 3.2.1 states: “The actuary should select risk characteristics that are related to expected outcomes.”

Section 3.2.2 states: “While the actuary should select risk characteristics that are related to expected outcomes, it is not necessary for the actuary to establish a cause and effect relationship between the risk characteristic and expected outcome in order to use a specific risk characteristic.”

Section 3.3.3 states: “When establishing risk classes, the actuary should (a) comply with applicable law; (b) consider industry practices for that type of financial or personal security system as known to the actuary; and (c) consider limitations created by business practices of the financial or personal security system as known to the actuary.”

As noted above, this ASOP says that “...it is not necessary for the actuary to establish a cause and effect relationship between the risk characteristic and expected outcome to use a specific risk characteristic.” However, this cause-and-effect relationship may make it easier to explain the results to policyholders, agents, regulators, underwriters, and management.

²⁵ Available on the Actuarial Standards Board's website.

It should be noted that this ASOP is not confined to pricing and underwriting. A Big Data project to identify liability claims that have a high potential for adverse development would use many data elements, each of which can be thought of as a risk classification. Care should be taken to ensure that the data elements, perhaps in combination, do not result in discrimination that would violate applicable law.

3. ASOP No. 38, *Using Models Outside the Actuary's Expertise (Property and Casualty)*

Section 3.3.1 states: “The actuary should be reasonably familiar with the basic components of the model and have a basic understanding of how such components interrelate within the model. In addition, the actuary should identify which fields of expertise were used in developing or updating the model and should make a reasonable effort to determine if the model is based on generally accepted practices within the applicable fields of expertise. The actuary should also be reasonably familiar with how the model was tested or validated and the level of independent expert review and testing.”

ASOP No. 38 covers topics in the P&C area that may be relevant to reliance on models developed by others, reliance on other actuaries on the modeling team, responsibilities in understanding the model, model structure, and model assumptions and parameters within the limits already discussed.

As of the writing of this paper, the ASB is considering the adoption of an actuarial standard of practice that more broadly addresses the use of models by actuaries in all practice areas. The proposed modeling ASOP has completed its 3rd exposure draft and will be considered by the ASB in June 2018 for a 4th exposure.

4. ASOP No. 25, *Credibility Procedures*

Section 3.5 states: “In carrying out credibility procedures, the actuary should consider the homogeneity of both the subject experience and the relevant experience. Within each set of experience, there may be segments that are not representative of the experience set as a whole. The predictive value can sometimes be enhanced by separate treatments of these segments. The actuary should also consider the balance between the homogeneity of the data and the size of the data set.”

ASOP No. 25 also covers such topics as selecting or developing credibility procedures, selection and blending of experience, and homogeneity of the data. Appendix 1 of ASOP No. 25 contains a section on emerging techniques that discusses generalized linear models and other multivariate modeling techniques. However, there is no express commentary regarding the applicability of this ASOP to Big Data.

5. ASOP No. 41, *Actuarial Communications*

Section 3.2 states: “In the actuarial report, the actuary should state the actuarial findings, and identify the methods, procedures, assumptions, and data used by the actuary with sufficient clarity that another actuary qualified in the same practice area could make an objective appraisal of the reasonableness of the actuary’s work as presented in the actuarial report.”

Section 3.4.4 states: “An actuarial communication should identify the party responsible for each material assumption and method. Where the communication is silent about such responsibility, the actuary who issued the communication will be assumed to have taken responsibility for that assumption or method. The actuary’s obligation for identifying the other party who selected the assumption or method depends upon how the assumption or method was selected.”

ASOP No. 41 also covers topics such as clarity, timing, who the responsible actuary is, the actuarial report, reliance on others for data and other information, responsibility for assumptions and methods, and disclosures, but there is no specific discussion of the applicability to Big Data.

6. ASOP No. 21, *Responding to or Assisting Auditors or Examiners in Connection with Financial Audits, Financial Reviews, and Financial Examinations*

Section 3.5.4 states: “The responding actuary should be prepared to discuss with the auditor or examiner, including the reviewing actuary, the following items underlying those elements of the financial statement or other elements within the scope of the financial audit, financial review, or financial examination for which the actuary is the responding actuary:

- a) the data used;
- b) the methods and assumptions used, and judgments applied, and the rationale for those methods, assumptions, and judgments;
- c) the source of any methods and assumptions not set by the responding actuary;
- d) the models used;
- e) the design and effectiveness of controls around the process, procedures, and models;
- f) any significant risks to the entity considered by the responding actuary; and
- g) the reasoning to support results and conclusions.”

Therefore, where Big Data and predictive analytics are used for financial reporting purposes, the responding actuary should be able to explain the use of Big Data to the reviewing actuary.

Notes

Notes

Notes



AMERICAN ACADEMY *of* ACTUARIES

Objective. Independent. Effective.™

AMERICAN ACADEMY OF ACTUARIES
1850 M STREET NW, SUITE 300, WASHINGTON, D.C. 20036
202-223-8196 | **ACTUARY.ORG**

© 2018 American Academy of Actuaries. All rights reserved.