CENTER for ECONOMIC JUSTICE | FAIR ACCESS FAIR TREATMENT

The Challenges and Opportunities of Big Data: Reforming State-Based Insurance Regulation in the 21st Century

Federal Advisory Committee on Insurance

Birny Birnbaum
Center for Economic Justice

January 5, 2017

The Center for Economic Justice

CEJ is a non-profit consumer advocacy organization dedicated to representing the interests of low-income and minority consumers as a class on economic justice issues. Most of our work is before administrative agencies on insurance, financial services and utility issues.

On the Web: www.cej-online.org

Why CEJ Works on Insurance Issues

Insurance Products Are Essential Financial Security Tools for Individual and Community Economic Development:

CEJ Works to Ensure Fair Access and Fair Prices for These Essential Products and Services, particularly for Low- and Moderate-Income Consumers.

Insurance is the Primary Institution to Promote Loss Prevention and Mitigation, Resiliency and Sustainability:

CEJ Works to Ensure Insurance Institutions Maximize Their Role in Efforts to Reduce Loss of Life and Property from Catastrophic Events and Promote Resiliency and Sustainability of Individuals, Businesses and Communities.

Big Data Defined

Insurers' use of Big Data has transformed the way they do marketing, pricing and claims settlement. Big Data means:

- Massive databases of information about (millions) of individual consumers
- Associated data mining and predictive analytics applied to those data
- Scoring models produced from these analytics.

The scoring models generated by data mining and predictive analytics are algorithms. Algorithms are lines of computer code that rapidly execute decisions based on rules set by programmers or, in the case of machine learning, generated from statistical correlations in massive datasets. With machine learning, the models change automatically.

What's So Big About Big Data?

1. There has been a revolution in insurance pricing, marketing and claims settlement resulting from insurers' use of Big Data -- massive databases of new insurance and non-insurance personal consumer information with associated data mining and predictive analytics and scoring.

2. Insurers' use of Big Data has huge potential to benefit consumers and insurers by transforming the insurer-consumer relationship and by discovering new insights into loss mitigation.

3. Insurers' use of Big Data has huge implications for fairness and affordability of insurance and for regulators' ability to keep up with the changes and protect consumers from unfair practices

4. The current insurance regulatory framework in most states – particularly related to risk classifications and unfair discrimination – does not provide regulators with the tools to effectively respond to insurers' use of Big Data. Big Data has massively increased the market power of insurers versus consumers and versus regulators.

5. Market forces alone – "free-market competition" – cannot and will not protect consumers from unfair insurer practices. So-called "innovation" without some consumer protection and public policy guardrails will lead to unfair outcomes.

6. Regulatory reform emphasizing regulatory collection, analysis and publication of consumer market outcomes – Regulatory Big Data – will yield more efficient and effective regulation for consumers, insurers and producers, will promote more competitive markets and will foster quicker adoption of innovative technologies that benefit consumers and fulfill public policy goals.

Personal Consumer Information in Big Data

- Telematics – Auto, Home, Wearable Devices
- Social Media
- Shopping Habits/Purchase History
- Hobbies and Interests
- Demographics/Household Data/Census Data
- Government Records/Property Records
- Web Tracking
- Vehicle Registration and Service Records
- Facial Analytics
- Mainstream Credit Files: Loans, Credit Cards
- Alternative Credit Data: Telecom, Utility, Rent Payment

Examples of Insurer Big Data Algorithms

Pricing/Underwriting:

- Price Optimization/Demand Models
- Customer Value Scores,
- Telematics,
- Credit Scores,
- Criminal History Scores,
- Vehicle Scores,
- FireLine Rating
- Accelerated Life Insurance Underwriting

Claims:

- Fraud Scores,
- Severity Scores
- Telematics

Example: Price Optimization/Consumer Demand Models

Deloitte Presentation at 2014 CAS Ratemaking Seminar

What is the ultimate goal of price optimization? Increase Profit

Insurance pricing can be classified in three levels of sophistication: Basic Rating Plans, Underwriting Models, and Market Demand Models

Market Demand Models: Customer price elasticity to optimize price

A key advantage of including underwriting components is that the insured's price elasticity and demand behavior **is on the final price at the policy level** and not the coverage and sub-component level. Optimizing price on the sub-coverage level creates a gap between the results and the insured's price behavior.

Example: Pricing Models

EagleEye Analytics Real Time Scoring Model

An insurer testimonial: when our underwriter is sitting at his or her desk, and they're looking at a renewal or quoting new business, there are two scores that pop up: a frequency score a kind of underwriting quality score and also a pricing score we use to help price renewal business and new business quoting. The risk scoring goes way beyond just financial data, it uses all of our characteristics in our data whether used for rating or not and then we are also able to bring in third party data, Census Bureau data, to supplement that.

EagleEye Analytics Real Time Scoring Model

Step 3: Build Elasticity Models. The loss ratio model will provide rate indications greater than you will take in one revision. **Elasticity models provide the data you need to forecast how your policies will respond to price change. Talon Elasticity models have multiple segments, not one curve for the entire portfolio.**

Step 4: Optimize. Using the Loss Ration Models, Elasticity Models and incorporating the aging that understands the new business penalty – it's time to make rate level decisions. **Optimization takes these models and applies them specifically to the policies you want. It's a tool to pick the perfect rate level** and forecast future profitability and growth.

Step 5: **Filing Machine Learning models is becoming common**. . . . EagleEye customers have filed and been approved in 45 states across multiple lines of business. . . .**As a bonus – once you switch to machine learning, your competition will not be able to reverse engineer your rating plan.**

Example: Pricing Model

TransUnion Criminal History Score

"TransUnion recently evaluated the predictive power of court record violation data (including criminal and traffic violations)

"While a court record violation is created during the initial citation, the state MVR is updated later and may be delayed depending on a consumer's response to the citation. For example, if someone pleads guilty and pays a ticket immediately, the state MVR will be updated in approximately two months. If the ticket is disputed in court, in contrast, the state MVR may not be updated for 6–19 months or longer.

"Also, as court records are created when the initial citation is issued, they provide insight into violations beyond those that ultimately end up on the MVR—such as violation dismissals, violation downgrades, and pre-adjudicated or open tickets."

Example: Claim Fraud Scores, Claim Severity Scores

LexisNexis Claim Tools

“LexisNexis (LN) seeks to provide a Single Point of Entry for delivering all of information directly back into a carrier’s system whether from a marketing standpoint, underwriting process or especially the claims part.

“LN has over 10,000 data sources that feed into its infrastructure each month and has contributed information from the industry.

“Claims Data Fill” – deliver data and analytics directly into claims system in the claims process regarding parties, vehicles and carrier information. Used to verify information provided to insurers and provide indicators beyond the data to identify whether a social security number is an indicator of fraud or whether an address provided is a good address. Has an analytic component at first notice of loss and throughout the claim, constantly monitoring the claim looking for fraudulent activities. Real time data verification and enhancement with fraud scoring and attributes

LexisNexis Claim Tools (con't)

"Example, insured calls in, rear-ended, all I got was license plate:

"Claims Data Fill takes that license plate, reach out to DMV to get vehicle registration to get VIN number, we have policy database and get the carrier and policy information, take the registered owner, go out to public records, pull back their address, date of birth, telephone number, social security, wrap that into a package and put it back into our system, 88% of the time done in less than 5 seconds.

"Take minimum information provided at first notice of loss, provide a fraud score at the initial notice of loss. Daily monitoring of claim every time new information comes in, able to run various scores: fraud scores, severity score."

Example: Fraud Scores

LexisNexis: "Severity Focus"

"Identify claims with the potential to become severe: SeverityFocus utilizes advanced predictive modeling to identify claims with the potential to become severe as they develop claims that otherwise would go undetected until much later.

"SeverityFocus does not constitute a "consumer report" as that term is defined in the federal Fair Credit Reporting Act, 15 USC 1681 et seq. (FCRA)."

Example: Claims Scoring

StatSoft's Predictive Claims Flow™

"A predictive analytics and reporting solution for property and casualty insurance companies can help you reduce loss ratios and improve bottom-line profitability, often within a few months of implementation. StatSoft's Predictive Claims Flow™ solution incorporates predictive modeling at every stage of an insurance claim. This closed loop system has a unique scoring system that rates each claim at its inception on its propensity for fraud and then continually rescores the claim as it goes through each step of a claim's lifecycle."

Example: Fraud Scores

Infosys Social Network Analysis

"The SNA tool combines a hybrid approach of analytical methods. The hybrid approach includes organizational business rules, statistical methods, pattern analysis, and network linkage analysis to really uncover large of amounts of data to show relationships via links. When one looks for fraud in a link analysis, one looks for clusters and how these clusters link to other clusters. Public records such as judgments, foreclosures, criminal records, address change frequency and bankruptcies are all data sources that can be integrated into a model. Using the hybrid approach, the insurer can rate these claims. If the rating is high, it indicates the claim is fraudulent."

Example: Fraud Scores

Infosys: Social Customer Relationship Management

“Social CRM is neither a platform nor a technology, but rather, a process. It is important that insurance companies link social media to their CRM. Social CRM . . . gathers data from various social media platforms. It uses a “listening” tool to extract data from social chatter,. . . .The reference data along with information stored in the CRM is fed into a case management system. The case management system then analyzes the information based on the organization’s business rules and sends a response. The response, from the claim management system as to whether the claim is fraudulent or not, is then confirmed by investigations independently, since the output of the social analytics is just an indicator and should not be taken as the final reason to reject a claim.”

Accelerated Underwriting for Life Insurance

AUW may look like an expanded simplified issue process but with mortality that aligns more closely with fully underwritten business

AUW is often modeled using predictive modeling and complex algorithms

New data sources include

- Enhanced application with use of behavioral economics
- Credit profiles
- Applicant Candor
- Criminal History
- Facial Analytics
- Use of Wearable Devices

AUW is a process that is dynamic in that non-medical and medical information gathering may be customized to the individual applicant.

Source: Mary Bahna-Nolan presentation to NAIC Life Actuarial Task Force, August 2016

A 20th Century Regulatory Framework for 21st Century Challenges

Oversight of Inputs Based on Proposition That Preventing Bad Inputs Will Prevent Bad Consumer Outcomes

Regulatory Oversight of

- Insurer Data – Licensing/Oversight of Advisory Organizations and Statistical Agents; Promulgation of Statistical Plans
- Cost-Based Pricing – Rates Not Excessive, Not Inadequate, Not Unfairly Discriminatory; Oversight of Rate Manuals/Rating Factors; Prohibition on Explicit Use of Certain Characteristics of the Consumer.
- Unfair Discrimination – Treating Individuals of the Same Class and Same Hazard Differently.

Big Data and Modeling of Rates/Prices/Claims/Customer Value

Old Old School Big Data: Advisory Organization Loss Costs. Oversight of Data, Advisory Organization, Analytic Techniques, Filings, Complete Transparency

Old School Big Data: Credit-Based Insurance Scores. Limited Consumer Protections for Completeness and Accuracy of Data via the FCRA, Limited Oversight of Modelers and Models, Limited Transparency

New School Big Data: Predictive Modeling of Any Database of Personal Consumer Information. No Consumer Protections for Completeness and Accuracy of Data, No Oversight of Modelers and Models, No Transparency to Consumers

The Regulatory Framework Breaks Down in an Era of Big Data

- Insurers now using data not subject to regulatory oversight or the consumer protections of the FCRA. Regulators have no ability to ensure the accuracy or completeness of these new data sets.
- Concept of unfair discrimination – consumers of similar class and hazard treated differently – becomes meaningless when insurers submit rating plans with millions of rate classes.
- New risk classifications can be proxies for protected classes, but with no recognition of disparate impact, risk classifications that have the effect of discriminating against protected classes are permitted. Big Data amplifies this problem.

Big Data Algorithms Can Reflect and Perpetuate Historical Inequities

Barocas and Selbst: *Big Data's Disparate Impact*

Advocates of algorithmic techniques like data mining argue that they eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data mining can inherit the prejudices of prior decision-makers or reflect the widespread biases that persist in society at large. Often, the "patterns" it discovers are simply preexisting societal patterns of inequality and exclusion. Unthinking reliance on data mining can deny members of vulnerable groups full participation in society.

New York Times, August 10, 2015: Algorithms and Bias: Q. and A. With Cynthia Dwork

Algorithms have become one of the most powerful arbiters in our lives. They make decisions about the news we read, the jobs we get, the people we meet, the schools we attend and the ads we see. Yet there is growing evidence that algorithms and other types of software can discriminate. The people who write them incorporate their biases, and algorithms often learn from human behavior, so they reflect the biases we hold.

Q: Some people have argued that algorithms eliminate discrimination because they make decisions based on data, free of human bias. Others say algorithms reflect and perpetuate human biases. What do you think?

A: Algorithms do not automatically eliminate bias. . . .Historical biases in the . . .data will be learned by the algorithm, and past discrimination will lead to future discrimination.

Fairness means that similar people are treated similarly. ***A true understanding of who should be considered similar for a particular classification task requires knowledge of sensitive attributes, and removing those attributes from consideration can introduce unfairness and harm utility.***

Q: Should computer science education include lessons on how to be aware of these issues and the various approaches to addressing them?
A: Absolutely! First, students should learn that design choices in algorithms embody value judgments and therefore bias the way systems operate. They should also learn that these things are subtle: For example, designing an algorithm for targeted advertising that is gender neutral is more complicated than simply ensuring that gender is ignored. They need to understand that classification rules obtained by machine learning are not immune from bias, especially when historical data incorporates bias.

White House Report on Big Data

For all of these reasons, the civil rights community is concerned that such algorithmic decisions raise the specter of “redlining” in the digital economy—the potential to discriminate against the most vulnerable classes of our society under the guise of neutral algorithms. . . . .But the ability to segment the population and to stratify consumer experiences so seamlessly as to be almost undetectable demands greater review, especially when it comes to the practice of differential pricing and other potentially discriminatory practices. It will also be important to examine how algorithmically-driven decisions might exacerbate existing socio-economic disparities beyond the pricing of goods and services, including in education and workforce settings

Insurer Use of Big Data Scoring Models Lack Fundamental Consumer Protections

- Accuracy and Completeness of Data
- Oversight of Data Bases
- Disclosures to Consumer About Data Used, How Used and Privacy Protections
- Consumer Ability to Challenge False Information
- Regulators' Knowledge Of and Capability to Provide meaningful Oversight
- Prevent discrimination Against Low-Income and Minority Consumers and other protected classes
- Asymmetric Use of Data
- Greater Cybersecurity Danger for Consumers and Insurers

21st Century Insurance Regulation

1. Identify What Insurers are Doing
2. Monitor Market Outcomes
3. Create an NAIC Resource for the States for Big Data Analytics
4. Develop a 21st Century Approach to Oversight of Risk Classifications, including Oversight of Modelers Acting as Advisory Organizations

Identify What Insurers Are Doing

To a great extent, regulators – and, of course, consumers and policy makers – do not know what types of information insurers are using and what they are using the information for and how they are using it.

A logical first step is to develop a template for states to use, with assistance from the NAIC for collection of requested information, to request from insurers the sources and uses of data for various insurance functions. For each source of data, the insurer would provide a name/description of the data, the source of the data and the use or uses of the data -- pricing (including underwriting), marketing, claims settlement, antifraud and other.

This periodic survey will provide regulators with the basic overview of what types of information are being used by insurers and what the information is being used for. This information is essential for regulators to respond to policy makers and to foster public discussion over potentially controversial types of data.

Monitor Market Outcomes

The old regulatory model of monitoring everything that goes into insurer marketing, pricing and claims settlement practices and models is not feasible in the era of Big Data. Big Data has massively increased the market power of insurers versus regulators and versus consumers

Regulatory Big Data is needed. The data needed for a robust market analysis – one that includes the ability to monitor the affordability and accessibility of insurance in underserved areas as well the ability to perform enhanced market analysis to focus regulatory resources on problem companies and problem markets – is transaction data on premium quotes, policies issued and claims. Insurance regulators lag other financial regulators in the ability to monitor market outcomes and the collection of granular data for market monitoring.

Industry’s “Muscular” Resistance to Regulatory Data Collection[1]

“The number of data calls has now reached a level where it is intrusive and affects insurers’ ability to effectively service their customers,”

“It is very questionable, especially when you consider the added cost to insurers, what value some of these data calls have to the consumer; some are also duplicated across multiple states.”

“Our members see these things as examples of regulatory overreach which has no connection to what regulators should be concerned with,” Sampson said. “On top of that, these requests are uncoordinated and often duplicated state by state.”

“Our members are concerned and have directed us to be more muscular in our approach to pushing back against that trend.”

---

[1] “PCI will be muscular as we combat regulatory overreach, says CEO Sampson," *Intelligent Insurer*, October 23, 2016

Why Would Regulators Be Seeking More Data from Insurers?

The current regulatory framework is based on oversight of inputs into in insurer pricing and practices as the way to ensure fair and good consumer outcomes.

With insurers now using many types and sources of non-insurance personal consumer information for marketing, pricing and claims settlement – data sources over which insurance regulators have no oversight – **regulators must turn to data collection of market outcomes to monitor markets and insurer performance.**

It only makes sense that insurers introduce new marketing, pricing and claim settlement practices more frequently and quickly than ever before, that regulators need to ask for data to monitor the action.

Improved and Expanded Regulatory Data Collection Can be a Win for Insurers, Regulators and Consumers.

Consider the following scenario to look at limitations of current insurance regulatory data collection for market monitoring and market analysis of insurer and producer market performance and of consumer outcomes.

From the Life Market Conduct Annual Statement (MCAS), the market analyst sees that an insurer has a **higher than average ratio of replacements to policies issued for Individual Cash Value Products**, which include:

- Variable Life
- Universal Life
- Variable Universal Life
- Term Life with Cash Value
- Whole Life
- Equity Index Life

What does the market analyst do?

Look at other MCAS data?

Ratios for Policies Surrendered Under 2 Years from Issuance, Between 2 and 5 Years or Between 6 and 10 Years?

Problem: No Way to Relate Issue Age Data to Policy Replacement Timing Data.

Limited data for market analysis means contacting insurer to find some explanation – an explanation that more often than not shows no problem or a problem with reporting of highly-summarized data.

Let’s Change this Scenario – What if the insurers were reporting, and regulators were analyzing, more detailed data?

The Market Analyst sees the higher-than-average ratio for replacements for Cash Value Life Insurance Products.

The Analyst then examines the replacement ratios by product and finds that this company’s sales are almost all one type of product and the company’s replacement ratio for this product is consistent with the replacement ratio for this product across the industry. Result: No need to contact insurer.

Now, let's say the analyst looks at the data by product line, finds that the insurer's sales are predominantly in indexed universal life and finds that the replacement ratio is much higher than the industry average?

The analyst looks to see the average age of policies replaced by the insurer for this product and finds a high percentage of replacements for recently issued policies.

Now the analyst looks to see the average issue age for the replacement policies and finds a very high percentage of replacements at issue age 85 and above.

The analyst continues to work with the data before contacting the insurer.

Now the analyst looks to see which producers were selling the replacement policies to the 85 and over consumers and finds that many agents were doing this.

Or the analyst finds that only one agent was responsible for the high number of replacements for older consumers.

Now we can compare the difference between market analysis of limited data and market analysis of granular data:

Instead of having to go to the insurer with a single metric – you have a high replacement rate for cash value products – the regulator can either answer the question without having to contact the insurer or, perform further analysis to narrow the focus – and inquiry to the insurer – to a particular product, process and/or producer.

More Granular Data Reporting by Insurers and Analysis by Regulators Produces Big Improvements in Efficiency and Effectiveness of Market Analysis.

More granular data reporting allows more refined market analysis.

More refined market analysis means fewer contacts with insurers to explain non-problems.

More granular data reporting means a huge reduction in special data calls because regulators already have the data in almost all cases.

More granular data reporting means more focused regulatory investigations and inquiries.

More granular data reporting means identifying insurers with good consumer market outcomes and leaving them alone.

More granular data reporting means regulatory involvement in company management policies and procedures if there is a problem, not as a routine practice.

This is Not State-of-the-Art Technology –

It is Technology from the 1990's

State-of-the-Art Technology would be Data Visualizations and Predictive Analytics of Market Outcome Data by Regulators.

There Are Opportunities to Build On Existing Data Reporting

Examples include:

- Transaction reporting for Principles-Based Reserving – eventually for life insurance, annuities and long-term care.
- Transaction reporting already in place for workers' compensation and significant portions of the industry in personal lines property and casualty insurance.

Muscular Cooperation, Not Opposition, Makes Sense For Industry

On the one hand, we routinely hear from industry about the costs and inefficiencies of market regulation – the urban legends about endless, multiple and duplicative market conduct examinations and special data calls.

But, on the other hand, every time the logical solution to these alleged problems is proposed – more granular data collection for more efficient and effective market analysis – the trades say, "Hell, No!"

CEJ urges industry to embrace regulatory big data and work with regulators and consumers for quick implementation. We urge regulators to move expeditiously on regulatory big data – it wouldn't be the first time regulators act in the industry's best interest despite industry opposition.

Benefits of Regulatory Big Data

- Calibrate inputs/processes to outcomes
- Assist financial regulation
- More efficient/effective market regulation
- Information to evaluate insurer and producer practices
- Information to evaluate existing and proposed public policies
- Information to meaningfully evaluate affordability and availability and competition issues
- Information to empower consumers and promote competitive markets

On the last point, consider how little information consumers have regarding the market performance of insurers and how publication of such information – frequency of claim denials, frequency of attorney involvement, average time to settle different types of claims and more – would empower consumers in the same way that information on car safety empowers consumers and motivates automakers.

Create an NAIC Resource for the States for Big Data Analytics

Just as the NAIC provides resources to assist the states in other areas – technical/actuarial capability at the NAIC to assist the states with PBR; collection and compilation of massive amounts of financial statement data to assist with financial analysis, collection and compilation of Market Conduct Annual Statement data to assist states with market analysis, to name just a few examples – the NAIC should develop resources to assist states in analyzing new big data models used by insurers.

The NAIC resource would not be a regulator and would not provide regulatory opinions. Rather, the NAIC resource would provide states with technical – actuarial and statistical – expertise to answer states' questions about a big data pricing or claims model. The NAIC resource would also assist states in accepting and processing large data sets as part of analysis of a pricing or claims model.

Modernize Regulatory Oversight of Risk Classifications

Risk classification represents insurers' – and society's – decisions about how to group consumers for the purpose of assigning premium. Risk classification determines the affordability and availability of essential financial security tools – insurance – for consumers. In the vast majority of states, the only justification needed for a risk classification is a correlation. But, in an era of Big Data and the modeling of rates and risk classifications, such a test is arbitrary and opaque. We see this in the huge differences in rate impact of the same risk classification across insurers and even for an insurer across states. There is a need for a 21st century approach to oversight of risk classifications – more transparency and more accountability

21st Century Regulation of Insurance Risk Classification

Require More Than A Correlation:

- Risk-Based Pricing / Avoid Adverse Selection
- Promote Loss Mitigation / Degree of Consumer Control
- Transparency to Consumers/ Accuracy & Completeness of Data
- Honor the Risk-Spreading Purpose and Function of Insurance
- Avoid/Minimize Disparate Impact

Risk Classifications – whether called underwriting guidelines, tier placement rule or rating factors – should all be approved prior to use with opportunity for stakeholder input.

Regulatory Oversight of New Data – update statutes to treat organizations providing advisory organization services as advisory organizations. The key activity is collective pricing, not insurer ownership of the advisory organization.

Consumer Data Ownership and Protections – Ensure consumer-generated data (telematics and wearable devices) is owned, controlled and transportable by the consumer. Ensure basic consumer protections for new data sources – disclosure, consent, access, completeness, accuracy, adverse action notice, ability to correct.