NAIC Insurance Summit

Artificial Intelligence in Insurance: Micro Risk Segmentation

June 21, 2018

Birny Birnbaum
Center for Economic Justice

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

About Birny Birnbaum

Birny Birnbaum is the Director of the Center for Economic Justice, a non-profit organization whose mission is to advocate on behalf of low-income consumers on issues of availability, affordability, accessibility of basic goods and services, such as utilities, credit and insurance.

Birny, an economist and former insurance regulator, has authored reports and testimony for numerous public agencies and consumer organizations, covering a wide variety of topics, including analysis of insurance markets, insurers' use of big data, market regulation, force-placed insurance, homeowners and flood insurance, consumer credit insurance, title insurance and insurance credit scoring. He has served for many years as a designated Consumer Representative at the National Association of Insurance Commissioners. He is a member of the Federal Advisory Committee on Insurance, chairing the Subcommittee on Affordability and Availability of Insurance.

Birny served as Associate Commissioner for Policy and Research and the Chief Economist at the Texas Department of Insurance. In that role, Birny was responsible for review and approval of rate filings, the development of data collection programs for market surveillance and the analysis of competition in numerous insurance markets.

Prior to his work at the TDI, Birny served as Chief Economist at the Texas Office of Public Insurance Counsel where he provided expert testimony in rate and rule hearings on behalf of insurance consumers before the TDI. While at OPIC, Birny performed the first auto insurance redlining study in Texas.

Birny was educated at Bowdoin College and the Massachusetts Institute of Technology. He holds the AMCM certification.

Why CEJ Works on Insurance Issues

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

CEJ works to ensure ***fair access*** and ***fair treatment*** for insurance consumers, 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 to ***promote resiliency and sustainability*** of individuals, businesses and communities.

Outline

1. Big Data Defined

2. Insurers' Use of Big Data

3. What's So Big About Big Data

4. Big Data Challenges for Regulators
    a. Transparency
    b. Unfair Discrimination in Pricing
    c. Unfair Competition
    d. Unfair Claims Settlement
    e. AI, Machine Learning and Micro Risk Segmentation

5. Big Data Opportunities for Regulators
    a. Articulate the Values for Insurance
    b. Enhanced Market Monitoring – Regulatory Big Data
    c. Ethical Algorithms and Disparate Impact
    d. Empowering Consumers for More Competitive Markets
    e. Economic Analyses of Markets

Big Data Defined

Insurers' use of Big Data has transformed the way they do marketing, pricing, claims settlement and their approach to risk management. For purposes of my talk, 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 artificial intelligence (AI) or machine learning, the models change automatically. Coupled with the increased volume and granularity of data is the digital technology to generate, access, process, analyze and deploy big data and big data algorithms in real time.

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/Mobile Phone Tracking/Data Harvesting
- Vehicle Registration and Service Records
- Facial Analytics
- Mainstream Credit Files: Loans, Credit Cards
- Alternative Credit Data: Telecom, Utility, Rent Payment

Sources of Data include consumers (via telematics or wearable devices), government, data brokers, online data aggregators and many others.

Examples of Insurer Big Data Algorithms

Pricing/Underwriting:

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

Claims:

- Claim Optimization/Demand Models
- Fraud Scores
- Facial Analytics
- Severity Scores
- Telematics

Big Data Algorithms as Insurance Market Gatekeepers

- Marketing: web searches and web advertising that pre-score and channel consumers to particular products, providers and price-levels.

- Pricing: pre-fill applications and pricing without the consumer providing information, pricing based not just on risk but on price optimization / consumer demand models, real-time competitive options and/or socio-economic characteristics

- Claims: automated, instant claim settlement proposals based on data generated by a vehicle, home telematics or wearable device and utilizing price optimization/consumer demand models to determine amount of claim settlement offer a particular consumer is likely to accept based on his or her personal data.

- Common characteristics – opaque algorithms, little or no disclosure or transparency to consumer, great potential to penalize most vulnerable consumers, limiting loss mitigation role of insurance

Allstate CEO to Investment Analysts, May 2017[1]

The insurer's "universal consumer view" keeps track of information on 125 million households, or 300 million-plus people, Wilson said.

"When you call now they'll know you and know you in some ways that they will surprise you, and give them the ability to provide more value added, so we call it the trusted adviser initiative," said Wilson.

Allstate's Data Analytics Subsidiary[2]

"Arity is a data company — an insight company, really — whether or not it's data from fitness sensors or home sensors," Hallgren says. "But everything out of the gate so far is focused on the connected car." That's because the company is benefiting from the wealth of data its parent company has gathered from its DriveWise programs and other telematics initiatives — 22 billion miles in total, according to Hallgren.

[1] "Allstate CEO: Agents Will Have Access to Data on 125 Million Households," Best's New Service, May 30, 2017

[2] "Allstate's Arity Unit Navigates Rapidly Changing World of Data, " Digital Insurance, June 5, 2017

What’s So Big about Big Data?

1. 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 and creating new tools for loss mitigation.

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

3. The current insurance regulatory framework generally 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.

4. 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.

Current Regulatory Framework Challenged in Era of Big Data 1

Old, Old School Big Data and the Current Regulatory Framework:

- Oversight of Statistical Plans and Data Collection
- Licensing and Oversight of Advisory Organization Providing Pricing Assistance to Insurers
- Filings and Statistical Data Contain and Reference Almost All Information Insurers Use for Pricing and Claims Settlement
- Complete Transparency to Regulators; Mostly Transparent to Consumers
- Market Regulation Based, Generally, on Auditing Model

Current Regulatory Framework Challenged in Era of Big Data 2

Old School Big Data: Credit-Based Insurance Scores

- Limited Consumer Protections for Completeness and Accuracy of Data via the Fair Credit Reporting Act
- Limited Oversight of Modelers and Models; Failure to Enforce or Amend Advisory Organization Statutes
- Limited Transparency to Regulators, Little or None to Consumers
- Consumer Protections in Name Only – "Life Events" and "Neutral Scoring"
- Failure to Address Disparate Impact
- Regulators' – and the Public's – Lack of Data for Evaluation of Scoring Models and their Impact on Affordability and Availability Exposed

Current Regulatory Framework Challenged in Era of Big Data 3

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,
- Little or No Transparency to Regulators, None to Consumers
- Problems That Emerged with Credit Scoring Grow
    - Lack of Data to Monitor Market Outcomes
    - Lack of Oversight of Collective Pricing Activities
    - Lack of Tools to Address Disparate Impact
    - Insurer Opposition to Providing Data
    - Big Data Issues with Anti-Fraud and Claim Settlement

Current Regulatory Framework Challenged in 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 and anti-fraud/claim settlement algorithms can be proxies for protected classes, but with no recognition of disparate impact, risk classifications and algorithms that have the effect of discriminating against protected classes are permitted. Big Data amplifies this problem.

Insurance Is Different from Other Consumer Products

1. **The insurance is required** – by law and by lenders requiring protection of home or vehicle collateralizing the loan.

2. **Contract is a promise for future benefits** if an undesirable event occurs. If the product "fails" – the consumer learns the insurance policy won't cover the loss – she is stuck and can't purchase another policy that would protect her against a known loss.

3. **Consumers have little or no information about the insurers' performance.**

4. **Cost-based pricing is required by actuarial standards of practice and financial solvency.** The requirement for cost-based pricing is to protect insurer financial condition and prevent intentional or unintentional unfair discrimination

5. **There is profound public interest in broad coverage** – failure or inability of consumers and businesses to access insurance has implications not just for individual families and businesses, but for taxpayers, communities and the nation.

Big Data Challenges to Insurance Market Regulation: Transparency

- What personal consumer information are insurers using for what purposes and how are they using it?

- Complexity of models, use of non-insurance data force regulators to rely on insurer and third-party modeler representations. Independent analysis not done and, given current regulatory resources, not possible.

- AI and Machine learning – models change on their own based on operation of the models – raises questions about insurers' ability to supervise models, let alone regulators' ability to do so.

- Insurers happy to explain their models one-on-one with individual insurance regulators – strain on regulatory resources and potential for inconsistent explanations across states.

Big Data Challenges to Insurance Market Regulation: Transparency

To protect alleged trade secrets and to avoid regulatory scrutiny, insurers' filings have become less transparent by:

1. Shifting rating factors to "underwriting" via "tier placement."
2. Utilizing composite factors
3. Utilizing third-party algorithms
4. Filing massive pdf documents
5. Referencing – but not providing – complex models
6. AI and Machine Learning models
7. Utilizing price and claims optimization to depart from actuarial indications.

Root Insurance Private Passenger Auto Rate Filing:

D09 Royce Independence Factor (RIF) The algorithm used to calculate the Royce Independence Factor (RIF) will return an expected factor for each driver based on their rating profile. This factor will be assigned to a decile and rated according to the RIF factors in the rating plan.

Intermap Flyer:

Filing Flood Rates That Can't Be Copied

Why file flood rates?
Because the Protection Gap on Flood is an enormous opportunity.
Because the data and analytics exist to underwrite flood effectively.

One of the reasons carriers have been slow to file flood rates is because they are concerned with publishing a proprietary approach to rating the risk. But what if it was possible to file rating that can't being (sic) copied.

Big Data Challenges to Insurance Market Regulation: Unfair Discrimination in Pricing

- Price Optimization – Consumer Demand, Competitive Options
- Disparate Impact on Protected Classes
- Evaluating the Reliability of Actuarial Justification

Big Data Challenges to Insurance Market Regulation: Unfair Competition

What Do Advisory Organizations Do and Why Are They Regulated?

Advisory organizations engage in collective actions by insurers – actions that, but for regulatory oversight, would be considered collusion or violations of state and federal anti-trust and anti-competition prohibitions.

The classic advisory organization activities are the development of policy forms and pricing assistance information. And the classic pricing assistance tools are the Insurance Service Office (ISO) advisory loss costs – analysis of industry-wide claim costs by detailed rating factors that can be used by insurers in the development of the insurers' rates.

Regulation of Advisory Organizations

Advisory organizations are licensed and the organization and its products are subject to examination. States have examined NCCI and ISO over the years.

Loss costs must be filed and, in most states, are subject to review and approval by the Commissioner. The loss cost filings have historically been based on data collected pursuant to regulatory-approved statistical plans and have included a complete explanation of the analysis from data compilation to loss cost recommendation.

Why are Advisory Organizations Regulated?

Why are these activities regulated? Because the advisory organization is involved in collective action, including policy coverage and pricing, for insurers.

What are the key activities of an advisory organization that require regulatory oversight? The collection of data from multiple insurers, the analysis of those data – perhaps including additional, non-insurance data – and the provision of pricing recommendations to insurers.

These are precisely the activities of third-parties currently providing big-data pricing and claim settlement algorithms to insurers – collecting data from insurers, adding additional data and producing pricing or claim settlement algorithms for insurers – price optimization, telematics, insurance credit scores, claim scores and more.

While some organizations engaged in advisory organization activities are regulated as advisory organization, most are not – presenting a challenge for regulators to conduct the oversight required by state laws regarding unfair discrimination.

Big Data Challenges to Insurance Market Regulation: Unfair Claim Settlement / Anti-Fraud

- Claim Optimization

- Asymmetric Access and Use of Data

- Anti-Fraud

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."

Carpe Data – Mining Social Media and Online Data

Carpe Data's Auto Loss Score utilizes social and online data to segment risk and determine loss potential. Our scoring algorithm is custom-tailored to employ data elements that are highly predictive in determining risk and loss potential as they relate to Personal Auto Insurance.

Claims Activity: Claims Monitoring Simplified

Scaled search and analysis of highly impactful social and public data allows automated monitoring to find a "needle in a haystack" of relevant, actionable claimant activity across bodily injury, workers comp, and disability claims.

Why It Matters

Carpe Data's next generation data allows insights into your claimants that would otherwise be either too difficult or expensive to find, such as:

- – Identification of exaggeration, malingering, abuse
- – Identification of and/or supporting evidence for possible fraud cases

Carpe Data: Life &Disability

Our Lead Qualification Score assesses applicants in real-time based on their online data. Social data characteristics are incorporated in the point of sale process to segment risks, allowing for identification of individuals more likely to qualify for target programs. Applicants are assigned scores that correspond with their likelihood of qualification, allowing Carriers to make informed decisions and segment risk into specific programs

Vital Status Leveraging proprietary data aggregation technology, Carpe Data identifies and mines publicly available online obituary data sourced from across the web. Vital Status provides carriers with an expansive and inexpensive means to update, supplement, and strengthen death master record data with up-to- date mortality data.

The Big Data Challenge & Opportunity for Insurance Market Regulation

The current regulatory paradigm is premised on the following: By monitoring all of the inputs into insurer marketing, pricing and claims settlement practices, regulators can ensure good consumer outcomes.

While it is debatable whether the "good inputs ensure good outcomes" is a reliable approach to consumer protection, it is simply no longer feasible for regulators to monitor the massive increase in sources, volume, uses and complexity of inputs to insurer models in an era of Big Data.

With Regulatory Big Data, the insurance market regulation paradigm can shift from attempting to monitor all inputs for insurer pricing and claim settlement to collecting and robustly analyzing data on actual consumer market outcomes. Stated differently, there is a need and opportunity for massively increased collection of granular insurer transaction data accompanied by regulatory data mining and big data analytics. The regulatory approach can transition from an auditing model to an analytics model.

Articulate What the Future of Insurance Should Look Like.

"Before we choose our tools and techniques, we must first choose our dreams and our values, for some technologies serve them while others make them unobtainable." Tom Bender

Our Dreams and Values for Insurance:

Empowered consumers and businesses partnering with risk management and sustainability companies who also provide insurance.

Greater, not less, transparency in insurance pricing, sales and claims settlements.

Monitor Markets More Comprehensively and Efficiently

<u>What data are insurers using for what purposes?</u> Routine collection – and publication – by regulators of the types, sources and uses of data by insurers for marketing, sales, pricing, claims settlement and loss mitigation.

<u>What consumer outcomes are insurers producing?</u> Routine collection and analysis by regulators of granular consumer insurance market outcomes, including transaction-detail data on quotes, sales and claim settlements.

<u>Public data to empower consumers.</u> Routine publication of insurer-specific anonymized consumer market outcomes.

Regulatory Big Data: Why Transaction Data?

Summary Data: Aggregated by pre-determined categories – e.g. Market Conduct Annual Statements

Transaction Data: Individual sales, policies and claims data – e.g., Principles-based reserving data collection, some state (TX) and advisory organization (ISO and NCCI) data collection.

Big Data applications – data mining and predictive analytics – are based on transaction-level data. Why? Univariate vs. Multivariate analyses

Univariate Limitations: Spurious Correlations, Limited Analysis Options

Transaction Data: Allows for Data Mining, Examination and Identification of Issues Not Previously Identified, Allows for Analysis by Adding Additional Consumer Information, Multivariate Analysis

Ethical Algorithms and Disparate Impact

Many states prohibit insurance discrimination on the basis of race, religion or national origin – for underwriting, pricing or claims settlement regardless of actuarial justification. For other rating factors, at a minimum, actuarial justification is required.

What is actuarial justification? A showing of a statistical relationship (correlation) between a particular characteristic of the consumer, vehicle, property or environment and the designated outcome – e.g., claim frequency, claim severity, pure premium, loss ratio, fraudulent claim, of a claim, likelihood of a fraudulent claim, loss ratio, retention, cross-sales, demand models.

Intentional Discrimination vs. Disparate Impact

Insurers argue that states' unfair discrimination laws prohibit intentional discrimination based on race – the explicit use of in pricing or claims settlement.

Disparate Impact refers to practices that, while not explicit discrimination on the basis of, say, race, have the same effect as if race was the basis for the discrimination.

Insurers also argue that Disparate Impact is not recognized as unfair discrimination in state insurance laws. (Disparate impact in homeowners insurance is specifically recognized as a violation of the federal Fair Housing Act).

Industry Trade Arguments against Disparate Impact in Insurance:

- Insurers don't consider race, religion or national origin, so there can be no unfair discrimination on the basis of these factors.
- Regulators have no authority to consider disparate impact:

Absent discriminatory treatment or failing to match price to the risk, the issue is whether they are even appropriate inquiries to apply to insurance rating. This is especially the case since some states prohibit even asking about the applicant's or policyholder's race or some other protected class status. As a result, the rating for a particular risk is truly color blind.
AIA and NAMIC Comments to NAIC Big Data Working Group, January 26, 2018

Why is Discrimination on the Basis of Race Considered Unfair Discrimination

Why are race, religion and national origin considered suspect classifications by the Supreme Court?

(1) there is a history of discrimination against the group in question;
(2) the characteristics that distinguish the group bear no relationship to the group members' ability to contribute to society;
(3) the distinguishing characteristics are immutable; and
(4) the subject class lacks political power.

Avraham, Ronen; Logue, Kyle D.; and Schwarcz, Daniel Benjamin, "Understanding Insurance Anti-Discrimination Laws" (2013). *Law & Economics Working Papers.*
http://repository.law.umich.edu/law_econ_current/52

Industry Arguments on Disparate Impact Flawed

The industry claim that their algorithms are "color blind" is, of course, nonsense to anyone familiar with algorithms because algorithms can reflect and perpetuate the historical biases of the data and the developers.

Further – if intentional discrimination against protected classes is prohibited, why would we ignore or permit unintentional discrimination that has the same effect be permitted?

Given that states (for auto insurance) and lenders (for auto and property insurance) require the purchase of insurance, and

That states (fines, loss of civil rights, imprisonment) and lenders (force-placed insurance) penalize consumers who fail to maintain required insurance, then

It is reasonable and necessary for insurance regulators to effectively monitor availability, affordability and actual market outcome on, among other reasons, the basis of protected classes.

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.

Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*

America's poor and working-class people have long been subject to invasive surveillance, midnight raids, and punitive public policy that increase the stigma and hardship of poverty. During the nineteenth century, they were quarantined in county poorhouses. During the twentieth century, they were investigated by caseworkers, treated like criminals on trial. Today, we have forged what I call a digital poorhouse from databases, algorithms, and risk models. It promises to eclipse the reach and repercussions of everything that came before.

Why Is Disparate Impact Relevant for Insurance Pricing?

TransUnion Criminal History Score

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

“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.”

What is the likelihood that TU Criminal History Scores have a disparate impact against African-Americans? Consider policing records in Ferguson, Missouri.

US DOJ Investigation of the Ferguson Police Department

Ferguson’s approach to law enforcement both reflects and reinforces racial bias, including stereotyping. The harms of Ferguson’s police and court practices are borne disproportionately by African Americans, and there is evidence that this is due in part to intentional discrimination on the basis of race.

Ferguson’s law enforcement practices overwhelmingly impact African Americans. Data collected by the Ferguson Police Department from 2012 to 2014 shows that African Americans account for 85% of vehicle stops, 90% of citations, and 93% of arrests made by FPD officers, despite comprising only 67% of Ferguson’s population.

US DOJ Investigation of the Ferguson Police Department (2)

FPD appears to bring certain offenses almost exclusively against African Americans. For example, from 2011 to 2013, African Americans accounted for 95% of Manner of Walking in Roadway charges, and 94% of all Failure to Comply charges.

Our investigation indicates that this disproportionate burden on African Americans cannot be explained by any difference in the rate at which people of different races violate the law. Rather, our investigation has revealed that these disparities occur, at least in part, because of unlawful bias against and stereotypes about African Americans

Potential for Disparate Impact of Anti-Fraud Models

Algorithms reflect bias in data and bias in modelers. Consider data used in anti-fraud models. These models tap a variety of data sources to identify characteristics of the consumer and/or the claim correlated with a fraudulent or suspicious claim. But what is the source of the claims identified as fraudulent used in the development of the models? Do the set of claims identified as fraudulent reflect historical bias in anti-fraud and claim settlement practices? Do the modelers have unintentional biases based on cultural backgrounds?

Example: Propensity for Fraud

"Unstructured data has become an opportunity instead of a problem. Many insurers have the ability to change unstructured information into structured data and actively mine this for the opportunities available therein."

"This [propensity] modelling is used to determine the likelihood of a new policy holder to commit a fraudulent act and it can be done in real-time … Fraud detection has changed in its location relative to the insured. Insurers are now able to run predictive and entity analytics during multiple touch points, essentially as each new piece of information is added. This not only improves detection capabilities in the event of fraud, but it also allows an insurer to assess a fraud-risk. Some have begun providing risky policy holders with high-priced policies in order to drive them to other service providers."

"The Role of Data and Analytics in Insurance Fraud Detection,"
www.insurancenexus.com, June 2016 (UK)

Ethical Algorithms:
Minimizing Disparate Impact in Insurance Models

<u>One Tool: Consider Prohibited Risk Classes in Model Development</u>

Step 1: Include race, religion and national origin – or proxies for these characteristics if actual individual characteristic unknown – as independent variables – control variables – in the model.

By using the characteristics as independent variables in the development of the model, the remaining independent variables' contribution (to explaining the dependent variable) is shorn of that part of their contribution that is a function of correlation with the prohibited characteristics. For the independent variables other than race, religion and national origin, what remains is a more accurate picture of the remaining independent variables' contribution to the target outcome.

Step 2: Omit race, religion and national origin when the model is deployed.

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

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.

Illustration of One Technique to Minimize Disparate Impact

Let's create a simple model to predict the likelihood of an auto claim:

$$b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e = y$$

Say that $X_{1,}$ X_2 + X_3 are miles driven, driving record and credit score and we are trying to predict y – the frequency of an auto claim.

Let's assume that all three Xs are statistically significant predictors of the likelihood of a claim and the b values are how much each X contributes to the explanation of claim.

b_0 is the "intercept" – a base amount and e is the error term – the portion of the explanation of the claim not provided by the independent variables.

What Happens When We Explicitly Consider A Variable For Race?

$$b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \mathbf{b_4R_1} + e = y$$

R_1 is a control variable – by including race in the model development, the correlation of the Xs to race is statistically removed and the new b values are now the contribution of the Xs, independent of their correlation to race, to explaining the likelihood of a claim

When the model is deployed, the variable for race is removed – the Xs remain, but the b values now minimize disparate impact.

Why is a Statistical Test for Disparate Impact Consistent with Actuarial Justification Used by Insurers?

Actuarial justification is a statistical test – that a particular characteristic of the consumer, vehicle, property or environment is correlated with a particular outcome, like pure premium (average claim cost). The same statistical test can be used to evaluate and minimize disparate impact. Stated differently – if a particular correlation and statistical significance is used to justify, say, insurance credit scoring, those same standards of correlation and statistical significance are reasonable evidence of disparate impact and unfair discrimination on the basis of prohibited factors.

Ethical Algorithms: Reasonable and Necessary for Insurance Pricing and Claims Settlement Models

1. Minimizes Disparate Impact – Stop the Cycle of Perpetuating Historical Discrimination.
2. Promotes Availability and Affordability for Underserved Groups
3. Improves Cost-Based Insurance Pricing Models
4. Improve Price Signals to Insureds for Loss Mitigation Investments
5. Help Identify Biases in Data and Modelers / Improve Data Insights
6. Improve Consumer Confidence of Fair Treatment by Insurers

Big Data Challenges for Insurance Regulators: AI, Machine Learning and Micro Risk Segmentation

Flawed Premise: Greater Risk Segmentation Means More Refined and More Accurate Pricing, Better Matching of Price to Risk.

First, what is the goal of ever more refined risk segmentation?

a. Insurer Solvency / Financial Condition?
b. Pricing / Price Optimization?
c. Marketing/Cross Selling?
d. Claims Settlement / Claims Optimization?
e. Loss Mitigation / Loss Prevention?
f. Greater Transparency / Accountability to Consumers and Regulators?

How Can AI, Machine Learning and Micro Risk Segments Help?

Second, let's dispense with "more granular segmentation is more accurate." Micro risk segmentation is a function of greater reliance on more complex predictive algorithms. Greater segmentation – more variables, more data sources – introduces new risks – modeling risk, data bias risk, unfair discrimination risk. Greater segmentation does not create greater "accuracy," where "accuracy" purports to be better matching price to risk.

Third, the concept of ever more "accurate" and granular risk segmentation must mean greater and greater disparity between the most and least favored consumers with great implications for availability and affordability of insurance and likely burdens on consumers in already-underserved communities under the guide of "matching price to risk" or "fighting fraud."

How Can AI, Machine Learning and Micro Risk Segments Help?

Let's revisit possible goals:

a. Insurer Solvency / Financial Condition? ***No Benefit, Potential Solvency Risk***

b. Pricing / Price Optimization? ***At Odds With Cost-Based Pricing, Will Inevitable Disadvantage Traditionally Unserved Communities***

c. Marketing/Cross Selling? ***Maybe, If Transparent. No, If Not.***

d. Claims Settlement / Claims Optimization? ***See b. If Transparent, Opportunities to Improve Claim Settlements.***

e. Loss Mitigation / Loss Prevention? ***Yes! If Transparent and Oriented Towards This Goal!***

f. Greater Transparency / Accountability to Consumers and Regulators? ***See e.***

How Can AI, Machine Learning and Micro Risk Segments Help?

"Before we choose our tools and techniques, we must first choose our dreams and our values, for some technologies serve them while others make them unobtainable."

AI and Machine Learning have potential for new insights into and new insurer-policyholder partnerships for loss prevention, resilience and sustainability with resulting improvements in insurance affordability and availability.

The key is to apply these technologies to the relevant goals. More opaque marketing, pricing and claims settlement algorithms based on the false claim of "better matching price to risk" are not the relevant goals.

New Tools to Empower Consumers

What data about me are you collecting and how well are your protecting my personal information? Insurers' and producers' transparency about and use and protection of consumers' personal information;

What is your actual history of treating consumers? Insurers' and intermediaries' performance based on actual market outcomes for consumers; and

What types of tools and assistance do you offer to help me manage my risk and control my premium? Insurers' and intermediaries' tools and partnerships for loss mitigation, loss prevention and consumer empowerment for risk management to control premium costs

Assist, Not Criminalize, Low-Income Consumers'

- Cost-Based Pricing Essential. Don't use insurance pricing to address affordability problems – no subsidies through pricing.
- Prohibit risk classifications that penalize consumers because of economic status.
- Put greater resources into assisting low-income consumers than to tracking, enforcement, penalizing, criminalizing and jailing consumers who cannot afford insurance.
- Create new product and pricing options to assist low-income consumers – low-cost auto product, pay-by-the-mile insurance.
- Federal, state and local government and insurer investments in resilient structures instead of subsidies to achieve affordable premiums.

Develop / Improve / Reinvigorate Capabilities for Economic Analysis of Markets, Competition and Anti-Trust.

Inconsistent and sporadic enforcement of advisory organization oversight – many organizations now providing pricing tools as advisory organizations without oversight as advisory organizations.

Will future success in insurance market be determined by quality of products and services or by amount of consumer data the insurer/intermediary/service organization controls?

The largest insurers – with the most data – have a profound competitive advantage over small- and medium-sized insurers because of far greater data assets.

Regulatory Intervention to align market forces with consumer interest, when needed. Regulatory data and economic analysis skills need to meaningfully monitor structure and competitive nature of insurance markets.