Big Data and Algorithms in Actuarial Modeling and Consumer Impacts—

Six Questions and Answers for Regulators

Presentation by Members of the
Data Science and Analytics Committee (DSAC)
American Academy of Actuaries

American Academy *of* Actuaries
Objective. Independent. Effective.™

actuary.org

January 25, 2022

Disclaimer

Please note: The presenters' statements and opinions are their own and do not necessarily represent the official statements or opinions of the ABCD, ASB, any boards or committees of the American Academy of Actuaries, or any other actuarial organization, nor do they express the opinions of their employers.

This presentation is derived from the *Big Data and Algorithms in Actuarial Modeling and Consumer Impacts* issue paper.

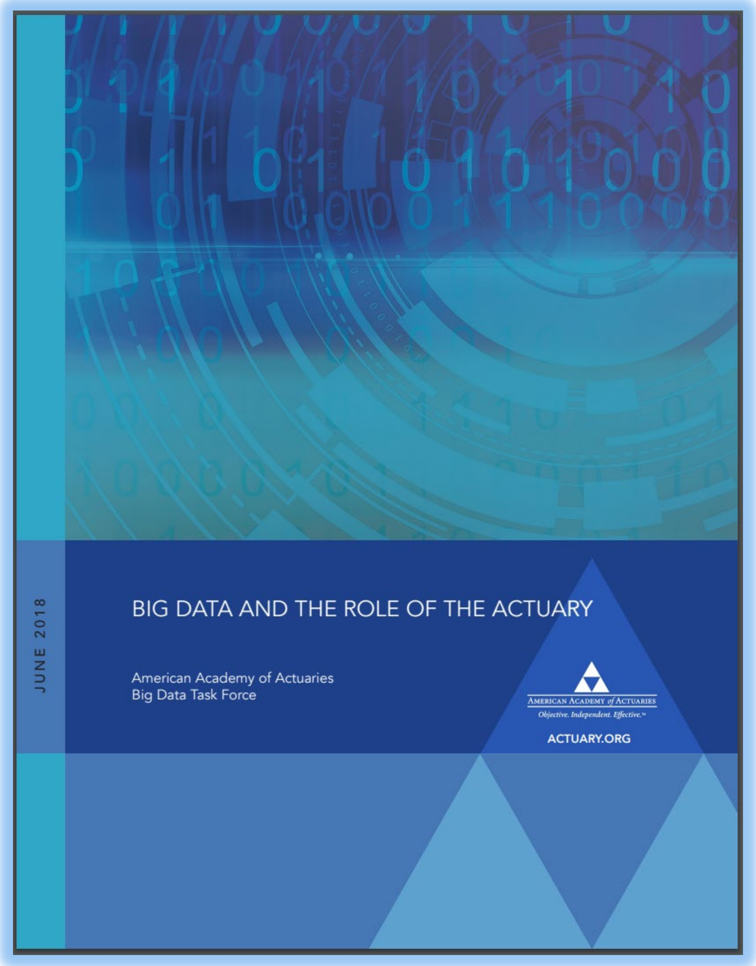
Presenters

- Dorothy L. Andrews—Chairperson, DSAC
- Dave Sandberg—Vice Chairperson, DSAC
- Seong-min Eom—Vice President, Risk Management and Financial Reporting Council (RMFRC)
- Mary Bahna-Nolan—Member, DSAC
- Liaw Huang—Member, DSAC
- Ross Zilber—Member, DSAC

Agenda

- The Genesis and Charge of the DSAC
- Six Question Topics
  - Importance of risk classification mechanism
  - Dangers inherent in modeling data
  - Perspectives on measuring systemic inequalities
  - Importance of professional standards in deciphering the black box
  - Important actuarial standards of practice (ASOPs) in modeling with big data and algorithms
  - Navigating the positive transformation of insurance utilizing big data and algorithms

Data Science and Analytics Committee (DSAC) Genesis



The creation of the Data Science and Analytics Committee resulted from the work of the Academy's Big Data Task Force, which was charged to

- Understand the impact of big data and algorithms on the role of the actuary
- Examine the framework of professional standards to provide guidance for working with these new tools
- Work with policymakers and regulators to address issues related to their use

The task force produced a monograph, *Big Data and the Role of the Actuary*

AMERICAN ACADEMY of ACTUARIES
Objective. Independent. Effective.™

Our Charge

"To further the actuarial profession's involvement in the use of data science, big data, predictive models, and other advanced analytics and modeling capabilities as it relates to actuarial practice.

To monitor federal legislation and regulatory activities, and develop comments and papers intended to educate stakeholders and provide guidance to actuaries."

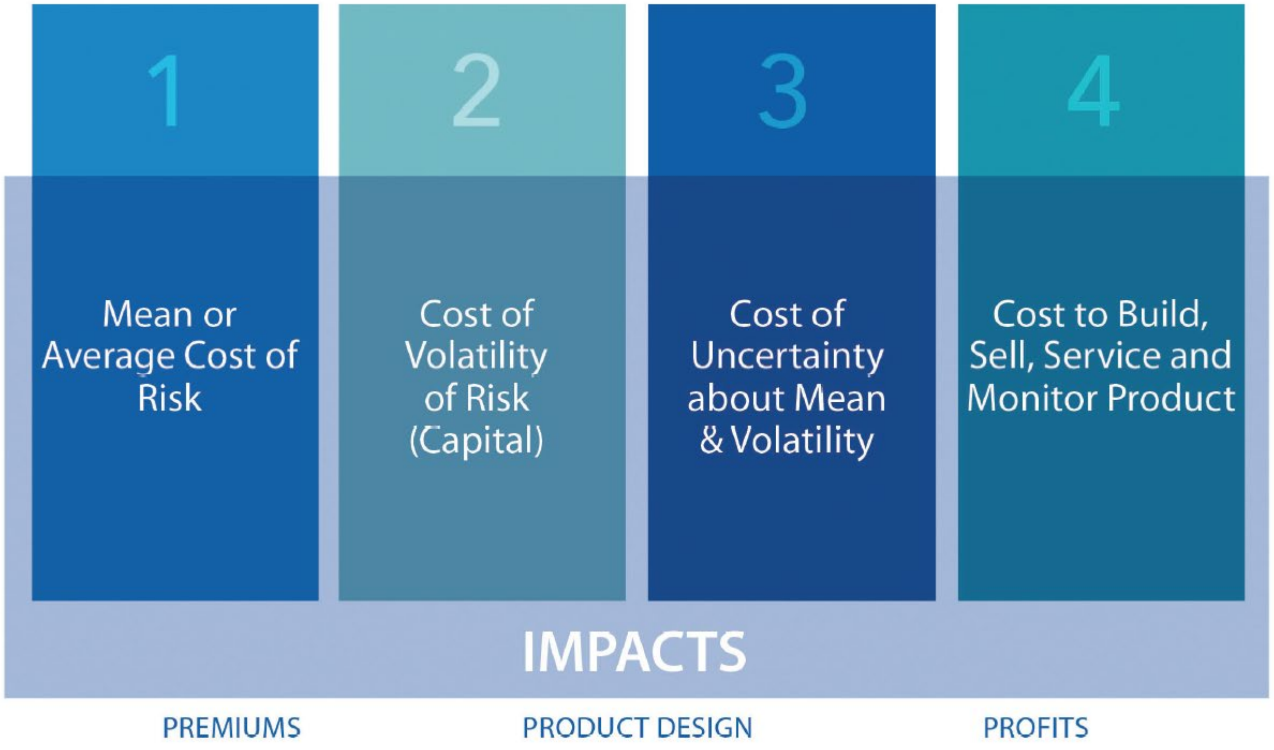
Question 1

Why is it important to preserve the risk classification mechanism in insurance?

—*Mary Bahna-Nolan*

Drivers of Value for Insurance Products

- Insurance covers varying exposures to loss, which can vary by:
  - Line of business
  - Target market and distribution
  - Ability to experience rate
  - Cost of capital
  - Level of uncertainty
    - Around mean
    - Volatility



What Is Risk Classification?

- The selection of risk classes (or assignment to a risk pool) resulting in equitable and fair rates
- Application of rules and algorithms to:
  - Project or estimate level of risk
  - Assign to a “risk pool” and assess a “risk charge”
  - Fully automate or use human judgment
- Can be lengthy, costly process for data gathering and risk assessment

Use of AI in Risk Classification

AI models emerged to segment risks and narrow dispersion around the mean from misclassification, fraud

Prediction errors are often visualized and measured through a confusion matrix

The lower the prediction error, the greater the confidence in the model to predict actual outcomes

Actual Risk Class		Risk Class based on Model Prediction			
		A	B	C	D
	A	Correct	Error	Error	Error
	B	Error	Correct	Error	Error
	C	Error	Error	Correct	Error
	D	Error	Error	Error	Correct

AMERICAN ACADEMY *of* ACTUARIES

Objective. Independent. Effective.™

Impact of Prediction Errors

- The greater the misclassification, the greater the costs associated with:
  - Risk charges
  - Premiums
  - Expected claims
- Fairness measures and constraints may be used to mitigate model bias and constrain misclassification due to systemic influences

Model or Data Identification of Risk Factor Presence

Actual Risk Factor Presence		Risk Not Present	Risk Present
	Risk Not Present	Correct	Type I Error
	Risk Present	Type II Error	Correct

Question 2

What are some of the dangers inherent in data used in risk classifications? What are some emerging practices to address them?

—*Dorothy L. Andrews*

Hidden Dangers in Data

Two Sources:

- Internal Data
- External Data

Hidden Dangers in Data

- Internal Data
  - Tends to be easier to audit if structured to identify errors
  - Unstructured data is often inconsistently conveyed and may be difficult to extract meaning from
  - Data quality issues (e.g., missing, null, etc.) results in imputed values which may be biased
  - Subject to selection bias
  - Unbalanced, lack diversity, overrepresentation, (e.g., CA, TX, often dominate training data in P&C models), outliers

Hidden Dangers in Data

- External Data
  - No access to audit the data, no transparency
  - Subject to biased collection, e.g., voluntary collection
  - Based on limited exposures, lacks diversity
  - Designed for a purpose not fit for the application
  - Can be difficult to correct by the consumer
  - May be collected in a period different from the model period
  - Problems arise in joining it to internal data
  - Loaded with proxy variables correlated with protected characteristics
  - Overly complex feature engineering

Hidden Dangers in Data

- Detecting Problematic Data
  - Look for variables in the following categories
    - Socioeconomic
    - Behavioral
    - Demographic, such as ZIP code
    - Consumer-related data
    - Price optimization related such as retention
    - Nonintuitive relationship with risk
  - Look for highly correlated variables ($\rho>0.5$) with protected attributes

Hidden Dangers in Data

- Detecting Problematic Data
  - Look for spurious correlations
    - Check the directionality of correlated pairs
    - Ask for research validating the relationship
    - Examine statistical significance in the presence of other variables
    - Check for dependency among variables:
      - If A, then B. If not A, then not B
    - Holdout testing
  - Examine variable rationales for intuitive relationship to risk, much harder than it sounds.

Question 3

What are the different perspectives that have been used to measure systemic inequalities on the conduct of insurance? Insurance is a social system, but it cannot solve all social problems.

—*Liaw Huang*

Systemic Influences and Socioeconomics

- *Case Study: I have been asked to develop a model to classify applicants of a new insurance product to a high-risk group using a set of modeling data. The model should not bias against members of protected classes, such as race and gender. How do I review the model results for systemic biases?*

- First, I develop the model without direct discrimination
  - Variables representing protected classes are not used in the model development
  - Data fields indicating race and gender are removed
  - The end result is race- / gender-neutral model

- What about indirect discrimination?

Systemic Influences and Socioeconomics

- First check: When I add the race and gender variable to the model, they do not improve the predictive performance of the model. Should I be alarmed?
  - Is it because race and gender are truly irreverent to the predictive model, or
  - The model uses power proxy variables that adding race and gender does not improve the statistical fit of the model?

- Second check: I look at the correlation of race and gender to the variables used in the model. I find correlations and the dataset is not balanced. I decide to adjust the modeling data. How should I do that?
  - Matching to the society's characteristics or the characteristics of a hypothetically fair society?
  - Matching to the characteristics of the people for which this product is marketed?
  - Matching to the characteristics to the people who are expected to purchase this product?

AMERICAN ACADEMY *of* ACTUARIES

Objective. Independent. Effective.™

Systemic Influences and Socioeconomics

- Third check: I look at the true positive rate by gender and find that the model captures 75% true positives for males but only 65% true positives for females. Is this a cause for concern?
  - What happens if the true positive rate is conditioned on modeling variables? Could this outcome be a consequence of the composition of the data?

- Fourth check: A colleague comments that what is really important is not whom your model gets right, but whom your model gets wrong. So, I look at the probability that a normal applicant is misclassified as high-risk, split by gender, should I expect different results?
  - Yes, the results can look quite different.

Systemic Influences and Socioeconomics

- Checking for and removing of systemic biases is difficult.
- Systemic biases can creep in at every step of the modeling process: data, algorithms, and validation of results.
- Different perspectives on systemic inequality give different measures of biases and inequality. It is possible that different perspectives can give different pictures.
- Not all goals can be achieved simultaneously, so all stakeholders should be involved.

AMERICAN ACADEMY of ACTUARIES

Objective. Independent. Effective.™

Question 4

How do I decipher the black box—and should I? What is important to document?

—Seong-min Eom

Reviewing the Model

- Common items that are included when models are reviewed:
  - Model risk and governance policy
  - Key stakeholders
  - Model documentation
  - Assumptions and data
  - Model testing
  - Model validation report
  - Governance and controls

Data Input and Assumptions

- Quality of Data
- Bias in Dataset
- Appropriateness of Data and Assumptions
  - Intended use of the data
  - Credibility
  - Judgment to select relevant variables
  - Reasonable assumptions

Model Documentation*

- Model owner as of a date
- Intended purpose and uses
- Version/last change date
- Summary of last validation and result
- Assumptions made in model construction
- Developer notes associated with any codes or calculation engine underlying the model
- Data sources and formats

* Model Risk Management Practice Note

Model Documentation*

- Parameter assumptions
- Dependencies on other models and processes
- Key outputs
- Applicable regulations and guidelines
- Limitations and future research areas
- Detailed step-by-step user instruction

* Model Risk Management Practice Note

Understand the Model

- Intended Use of the Model
- Scope and Process of Model Validation
- Model Validation Date
- Quality of Model Documentation
- Testing Performed
- Issues and Limits of the Model
- Reliance

Governance and Controls

- Relevant Governance and Policies
  - Model risk management
  - Model development & approval procedures
  - Key stakeholders
  - Controls in model processes
  - Third-party risk management

AMERICAN ACADEMY *of* ACTUARIES

Objective. Independent. Effective.™

Question 5

Which ASOPs are important in risk classification models?

—Ross Zilber

ASOPs are a pillar of actuarial governance

- The framework of the three pillars—(1) The Code of Professional Conduct (the “Code”), (2) U.S. Qualification Standards, and (3) ASOPs
- The Code identifies responsibility to the public and principals and sets forth what it means for an actuary to act as a professional
- Qualification Standards—perform service when qualified, must be prepared to document their qualifications
- ASOP—Prepare to justify deviations

Relevant ASOPs

- ASOP No. 9—P&C scope on Ratemaking, Loss Reserving, and Valuation
- ASOP No. 12—Risk Classification—All Areas
  - Risk characteristics that are related to expected outcomes—causality, objectivity, documentation
    - Smokers and lung cancers and comorbidity, or gender-distinct pricing
    - Question causality of other variables, e.g., ZIP code

Relevant ASOPs—Continued

- ASOP No. 23—Data Quality
  - Appropriate data, audit, comprehensive
  - Data selection, review, reliance
    - “garbage in, garbage out”
- ASOP No. 41—Actuarial Communication
  - Form and content, clarity of actuarial report
  - Reliance on other sources for data and other information

Relevant ASOPs—Continued

- ASOP No. 56—Modeling
  - Intended user, cost projection, pricing, predictive, reserving, planning
    - Overfitting, intended use of the model, assumptions,
    - Understanding the model
    - Reliance

Question 6

What are the opportunities of using big data and algorithms to positively impact the transformation of insurance, improve the customer experience and navigate the future of insurance? What is left to be done to solve some of the problems highlighted previously?

—Dave Sandberg

Adapting to & Addressing the New Normal

What Ideas Changed Insurance 30 Years Ago?

1. Financial Economics
2. Modeling Principles
3. Enterprise Risk Management(ERM)/Asset and Liability Management (ALM)

These ideas led to: Own Risk and Solvency Assessment (ORSA), Econ. Capital, Three Pillars & Cat Models

Actuaries lead and are navigating this new world for

- Boards of Directors
- NAIC & IAIS
- SEC & FINRA
- FASB & IASB
- FED & EU & Bank of England
- Wall Street Journal & NY Times

Most Current Innovation is Outside the Box

The NEW Frontier

What are the new analytics needed by regulators and actuaries?

1. Discerning potential unicorns vs. innovations vs. expense saving
2. Rating the quality of data assets (a la S&P)
3. Rating algorithms (a la NASA technology readiness levels)
4. Assessing the skill/competence of actuaries to use and or audit data and algorithms

AMERICAN ACADEMY of ACTUARIES

Objective. Independent. Effective.™

The FRONTIER IS Growing

Types of Data

- Structured
  - Internal company data
  - Publicly data sources “mined” by external vendors
  - Multiple-choice surveys

- Unstructured (written—freeform text, images, video, audio)
  - Underwriting files
  - Claim files
  - Suitability reviews

Academy Resources

Link to Paper: https://www.actuary.org/sites/default/files/2021-11/BigData_and_Algorithms_in_Actuarial_Modeling_and_Consumer_Impacts.pdf

Questions

AMERICAN ACADEMY of ACTUARIES

Objective. Independent. Effective.™