Session 060: Bias, Fairness and Discrimination Issues in the Use of Statistical Modeling

SOA Antitrust Compliance Guidelines
SOA Presentation Disclaimer

2019 ANNUAL MEETING & EXHIBIT

Bias, fairness, and discrimination issues in the use of statistical modeling

Thomas D. Fletcher, PhD

Shane De Zilwa, PhD

Natasha Cupp

October 2019

SOCIETY OF ACTUARIES®

PartnerRe

Bias, fairness, and discrimination issues in the use of statistical modeling

Thomas D. Fletcher, PhD
VP Data Analytics – North America

October 17, 2019

Disclaimers

The thoughts are mine and no one wants to take credit.

The following presentation is for general information, education and discussion purposes only, in connection with the SOA Conference 2019. Any views or opinions expressed are those of the presenters alone. They do not constitute legal or professional advice; and do not necessarily reflect, in whole or in part, any corporate position, opinion or view of PartnerRe or its affiliates, or a corporate endorsement, position or preference with respect to any issue or area covered in the presentation.

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

A brief historical example of tests as models

Cognitive ability predicts well, but leads to fairness issues

- Bias/ fairness issues are not only a perennial topic, but are key part of discussions in recent times
- Drawing on 50+ years of evolution in other domains (e.g., personnel psychology), examples of key concepts are given as well as mechanisms to ameliorate fairness concerns
- Consider as a parallel to UW, testing of potential employees. A decision is being made to hire or not hire
- Cognitive ability is a good predictor of performance but has a history of unfair discrimination

- Ultimately, we have a privilege when we use data, and an implicit assumption that we want to build models that are useful
- As an industry we have a societal obligation to be fair or at least do no harm. Insurance can be viewed as a social good
- However, in insurance, unlike in other disciplines, we often don't challenge ourselves (or allow) to collect the requisite data to ensure fairness
- Further, some common concepts are often confused when delineating our goals (discrimination vs. validation)

SOURCES:
Cascio, W. & Aguinas, H. (2010). *Applied Psychology in Human Resource Management 7th.*
Society for Industrial and Organizational Psychology (2003). *Principles for the Validation and use of Personnel Selection Procedures*.
AERA, APA, NCME (2014). *Standards for Educational and Psychological Testing.*
American Psychological Association (2017) Ethical principles of psychologists and code of conduct.
EEOC (1978) *Uniform Guidelines on Employee Selection Procedures*.

Level setting of concepts

Brief vocabulary lesson – Is your model defensible?

Validity

Model measures what it purports to measure
Statistical or conceptual relation to the outcome
Relevant and fit for purpose

Fairness

Socio-political concern rather than technical
Courts, Society, etc. determine fairness
May differ by jurisdiction and context

Differential Validity

Predictor-Target relationship differs by class
Does not necessarily reflect unfair discrimination
Intercept differences are more common than slope differences (in selection research)

Unfair Discrimination

A goal of models should be to discriminate among relevant differences defensibly
Guion (1966) – "… exists when persons with equal probabilities of success … have unequal probabilities of [being selected]"

Utility

Overall usefulness derived from a model
Can be economic/monetary or efficiency
Includes costs and consequences of model use

Disparate Treatment

Denial of equal opportunity by class membership; Validation in one class does not automatically carry forward to another class

Bias

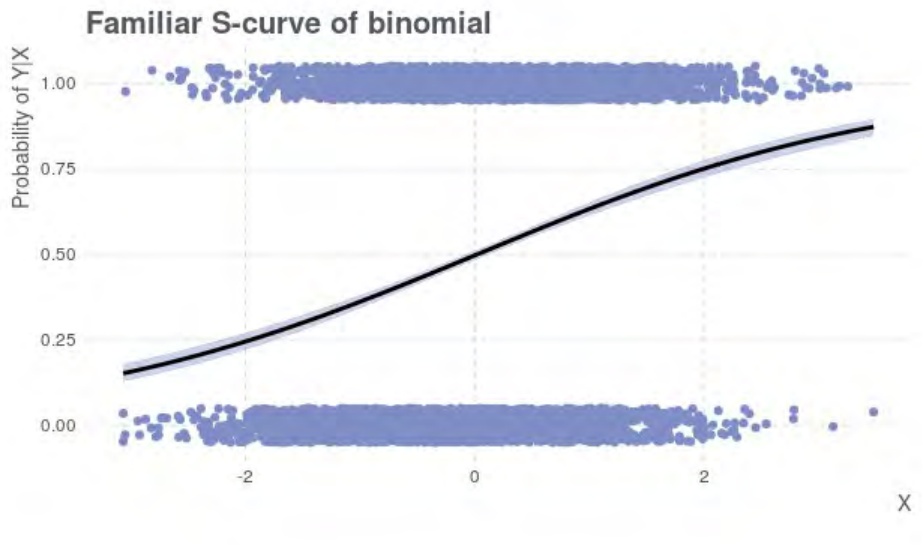
Moderation of slopes; systematic under/over prediction by one of the sub-classes
Cleary (1968) – may connote unfairness in usage unless appropriate measures taken

Adverse Impact

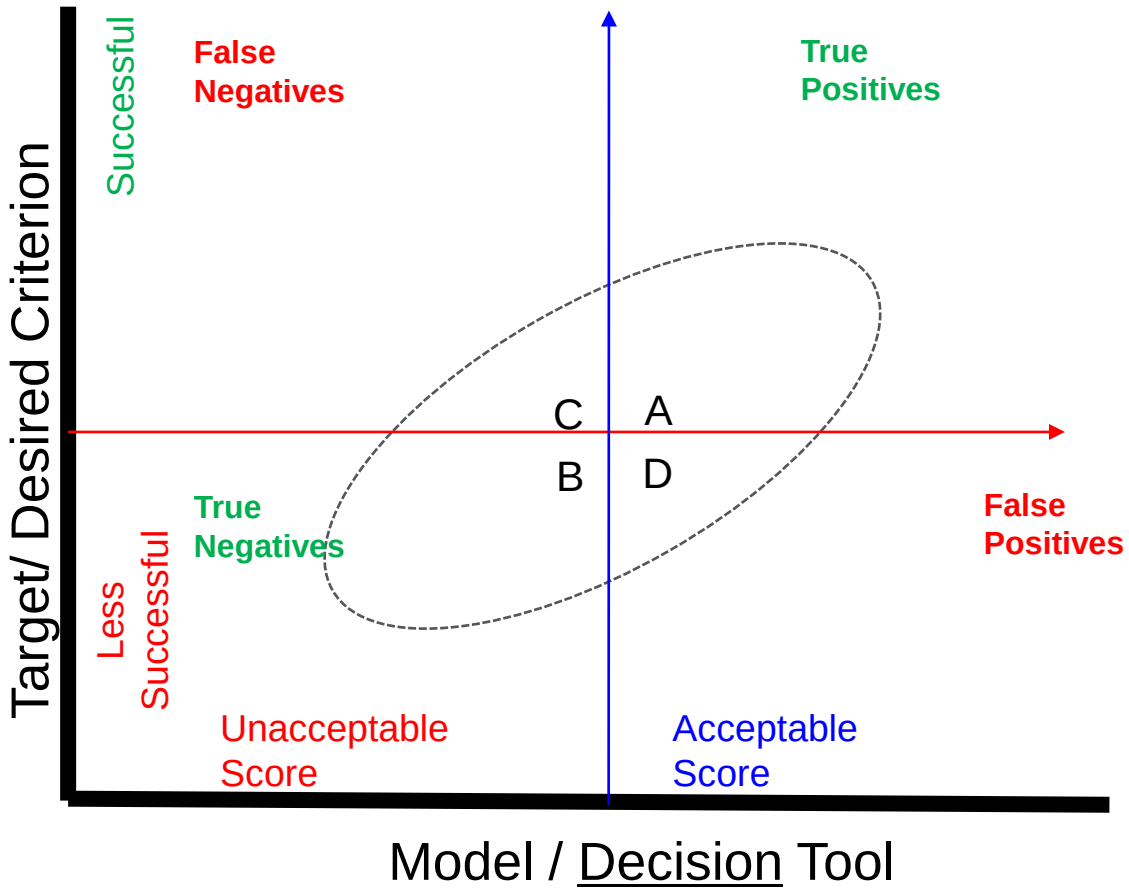
Potential by-product of model use
Disproportionate selection of one class over another regardless of validity of model.
Assessed after model deployment (e.g., via 4/5ths rule)

Leveraging models for decisions

A visual display of putting models into practice



- From the s-curve fit via a GLM, we know that there are errors in models (0,1 becomes a probability 0:1)
- Errors in decisions are balanced based on desired outcomes by success criteria and acceptable score thresholds



Statistical validity can be represented graphically

A homogenous population with high and low validity

Model Validity

A can be re-presented via fitted values (Xs) and Target (Ys)
Regression line represents the direction, magnitude, bias in prediction. Slope can be systematically under/over or on target

Ellipsis and Strength

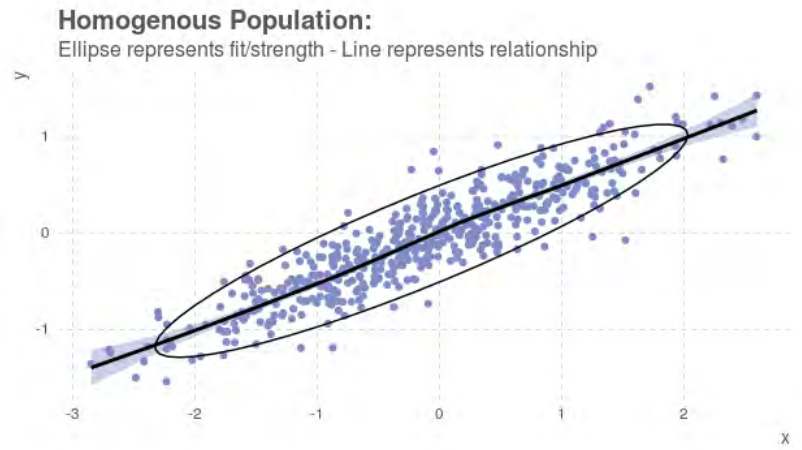
Ellipsis represents model fit (or R^2) in a linear regression
Captures the overall shape of the errors from regression line
Narrow ellipsis is a strong model, circle reflects no validity

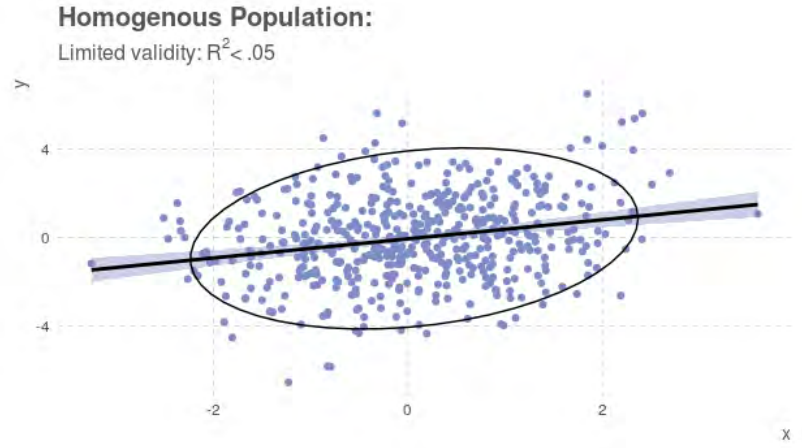
Homogeneity Assumption

As depicted here, the models reflect a single, homogenous population with no assumed sub-classes
A sub-class might reflect errors in an unclosed system (missing info). Classes (if present) should not be systematically different

Misc. Other

Multivariate outliers exist outside the ellipsis or far from line
Error bands can be seen in the tails of the model (far left/right)
We could represent these relationships with just the ellipses



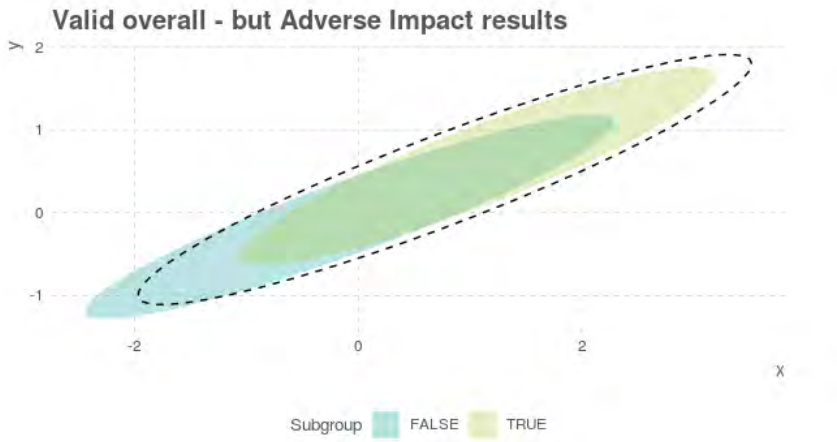


Ignoring sub-groups has consequences

Adverse impact and unfair discrimination

Model Issue

Model has demonstrable validity; one single slope
Yet, model usage will result in adverse impact (majority subgroup [yellow] will be *selected* disproportionately more than non-majority)

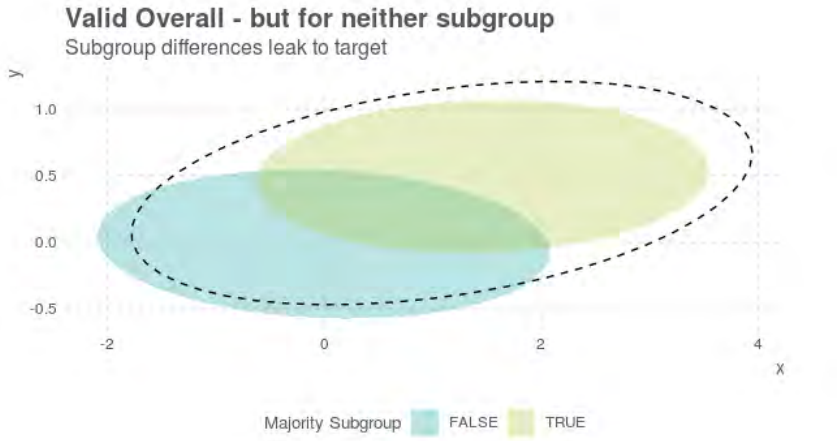


Potential Remedy

Add other predictors or additional model/decision criteria
Defend model usage and consequences (may be justified)

Model Issue

Overall, validity is false; subgroups demonstrate the model is not actually valid – rather is discriminates on subgroup alone



Potential Remedy

Such a model is not defensible and should not be used
Build a better model

Subgroups could differ in model scores or target

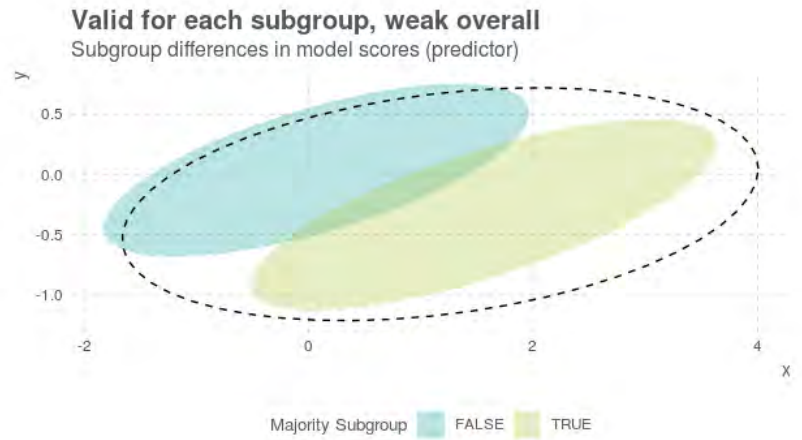
Differences introduce bias and mask validity potential

Model Issue

Model is valid within classes – yet score differences exist
Weak overall (ignoring only hurts validity evidence)
May erroneously conclude model isn't useful
Differential selection would result even though equal proportions would be found above a *success target*

Potential Remedy

Differential Prediction
Set cut scores for each group differently

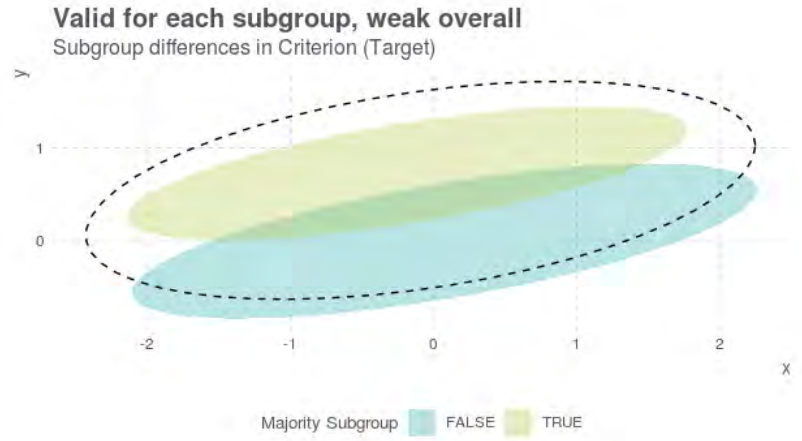


Model Issue

Model is valid within classes – yet target differences exist
Weak overall (ignoring only hurts validity evidence)
May erroneously conclude model isn't useful
Differential performance results (choose equal proportions of classes but unequal success rates) potentially reinforcing prejudices (Barlett & O'Leary, 1969)

Potential Remedy

Differential Prediction
Set cut scores for each group differently
Or, don't use the model as is

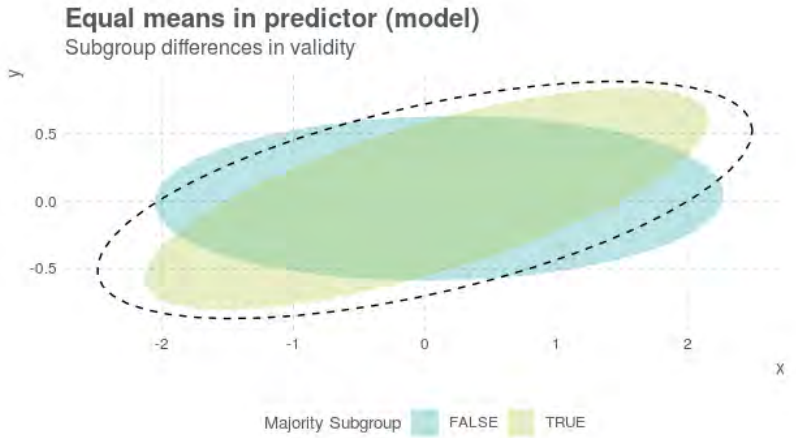


Models may not work well on all classes

Differential validity masks true validity for subgroups

Model Issue

Predictor and Criterion Scores are equal (slope differences)
Limited validity of one class lowers overall utility of the model
Model may not be justified for one class, and is under performing for another

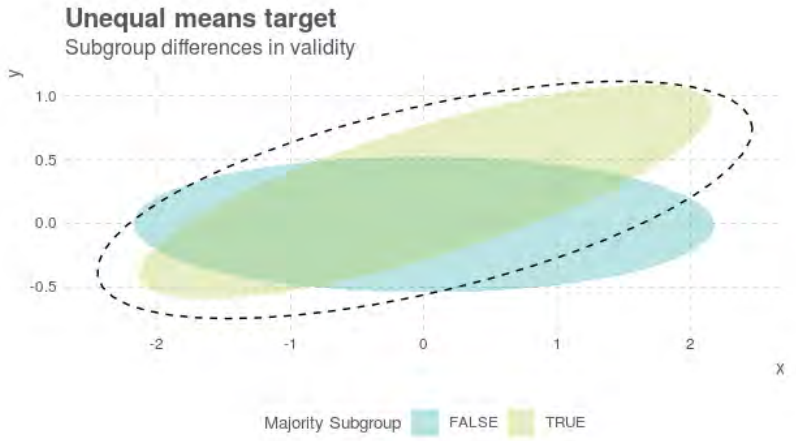


Potential Remedy

If justifiable limit usage of model only to the valid class
If model used on non-valid subclass – equal selection but differential performance would result reinforcing stereotypes
Alternate methods are required for the non-valid group

Model Issue

Differential performance (success on target) exists, and slope differences exist (lack validity in one class)
Limited validity of one class lowers overall utility of the model
Model may not be justified with combined data



Potential Remedy

If justifiable limit usage of model only to the valid class. Cut-score should be based on the valid class (not combined group)
Alternate methods are required for the non-valid group
If cut scores included data from the non-valid class – greater errors in prediction will result (even if only implemented on the valid class)

PartnerRe

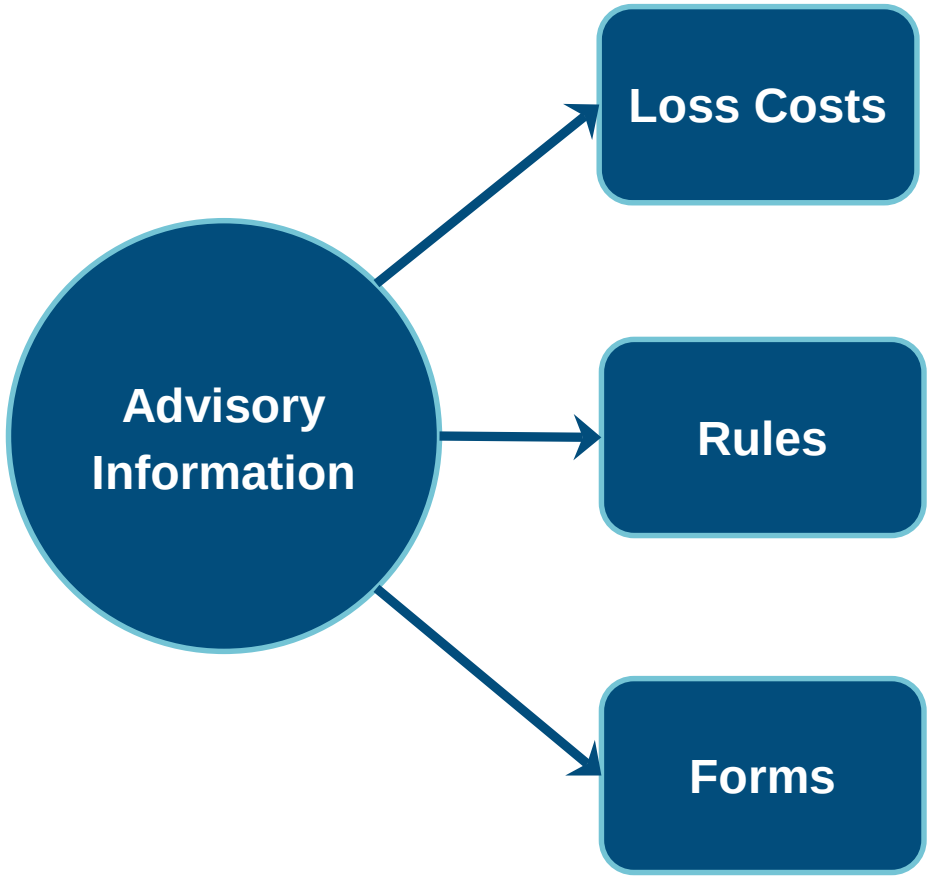
2018 ANNUAL MEETING & EXHIBIT

Bias, fairness, and discrimination issues in the use of statistical modeling: Current issues in model building

Shane De Zilwa
Verisk

SOCIETY OF ACTUARIES®

ISO Role in the Property/Casualty Industry



Advisory rating information based on line, class, or occupancy (created from data analysis)

Classification, policy writing, rating rules (the commercial and personal lines manuals), and other supplementary rating information including rating plans, factors, and relativities)

Policy forms (language) for all lines of business under ISO jurisdiction

Property/Casualty Regulatory Environment

- Forms, rates and rating factors are often regulated at the front end
- Must usually be approved or acknowledged before use
- Regulatory standard is that *“rates not be excessive, inadequate or unfairly discriminatory”*

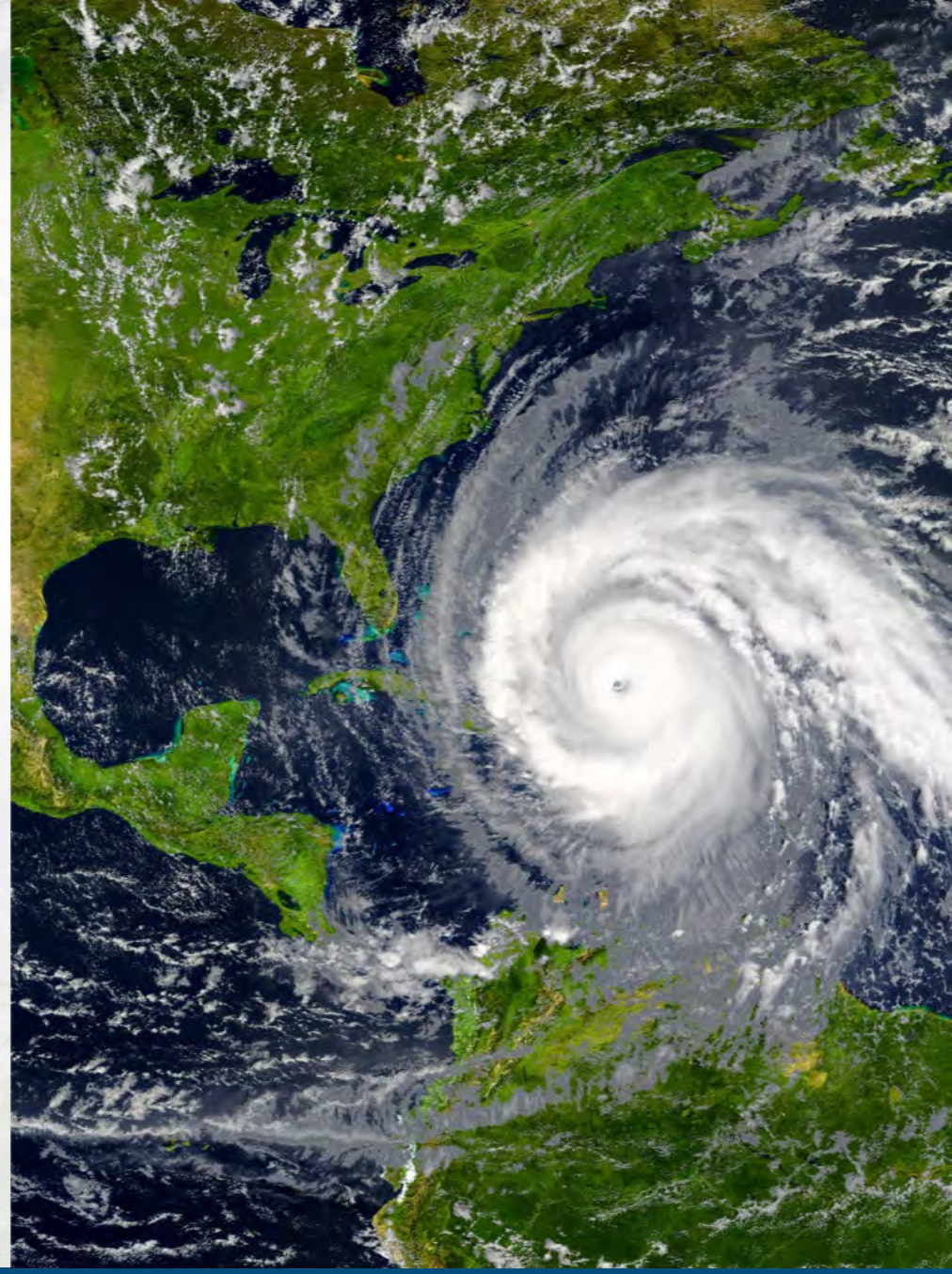
Fair vs Unfair Discrimination

- **In general, fair discrimination based on actuarial analysis is allowed**
- **Must show that classification scheme is supported by data or actuarial judgment**
- **Data typically shows differences by type of vehicle, age, sex, marital status, territories (urban vs rural.)**
- **The current arms race is looking to improve the granularity of pricing in support of the fundamental principle of cost based pricing.**
- **ASOP 12 – Risk Classification**
  - 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.

Catastrophe Models

- Historical method before cat models was to use data from long time periods (30-50 years) to estimate cost of extreme wind events like hurricanes
- *Hurricane Hugo (1989) and Andrew (1992)* showed the inadequacy of the traditional approach since the return period for hurricanes was longer than the experience period.
- Cat models simulate many years (e.g. 100,000 years) of events to get a fuller picture of the loss potential
- Models validated with claims data from actual events
- Initial use of cat models involved internal actuarial review as well as external review from relevant experts (meteorologist, seismologist)
- Regulatory road shows to get regulators comfortable with the models before filing

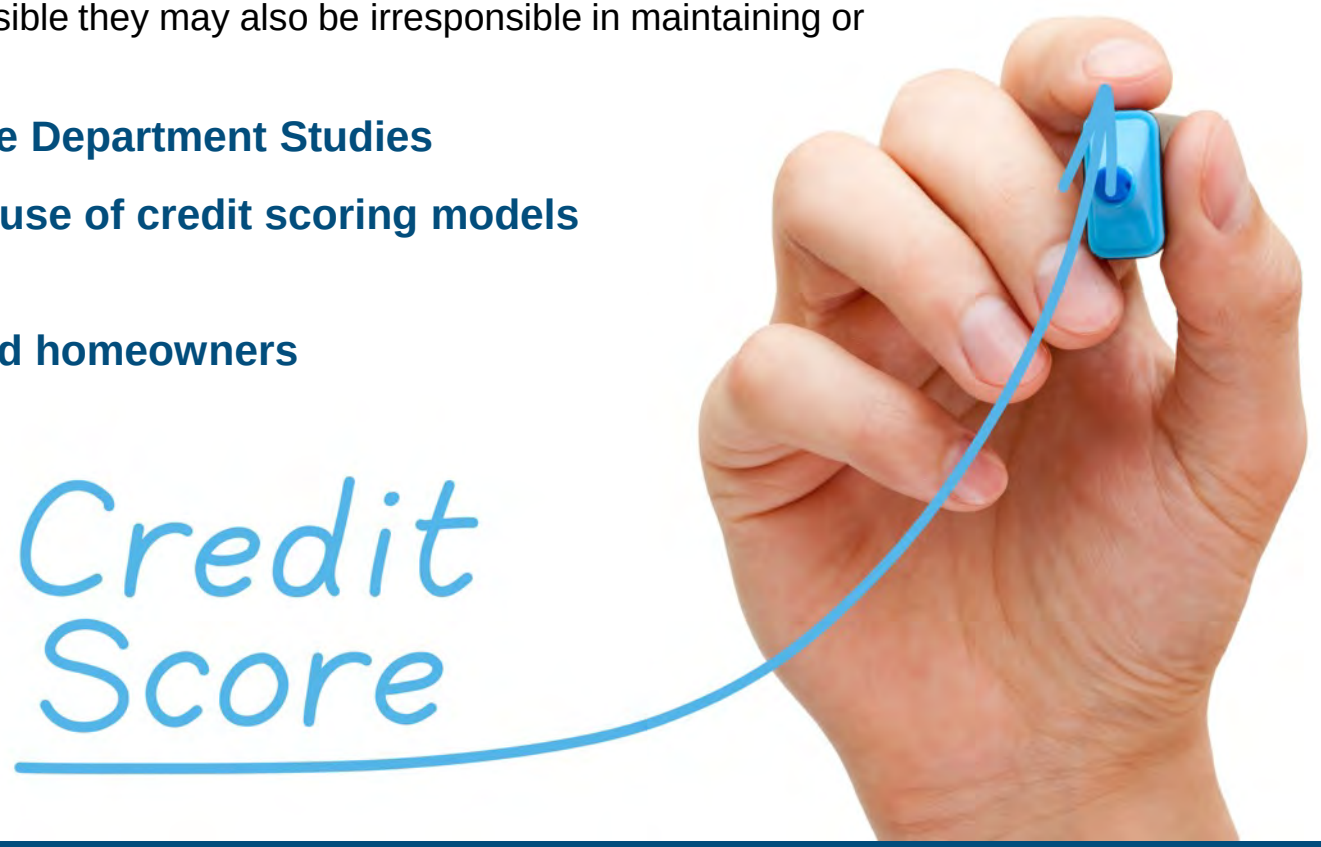
- ASOP 38 - Using Models Outside the Actuary's Area of Expertise



Credit Models

- **Credit scores used in banking**
- **Vendors discovered that credit score was correlated with loss experience for personal auto**
- **Causality -- How does my credit history make me a riskier driver or homeowner?**

  Possible explanation: if someone is financially irresponsible they may also be irresponsible in maintaining or operating a car or maintaining a safe home
- **Federal Trade Commission and Texas Insurance Department Studies**
- **NCOIL developed a model law that allowed the use of credit scoring models**
  - Most states have adopted the law
- **Credit score is widely used in personal auto and homeowners underwriting and rating**

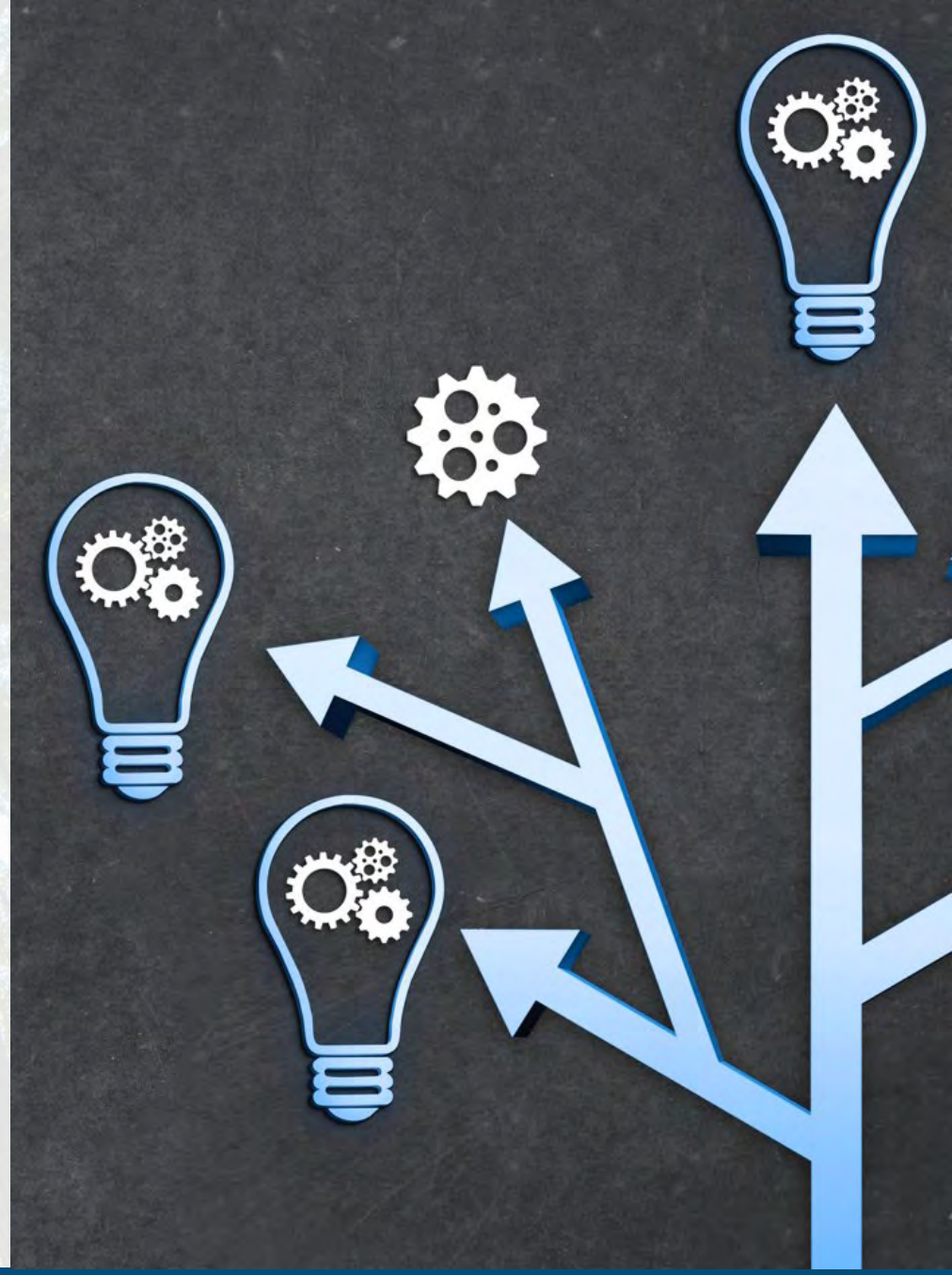


GLMs

- **Next generation of models based on availability of data and the development and proliferation of advanced data analytics techniques**
- **Multivariate analysis which could uncover previously unknown relationships**
- **Best practice to exclude from modeling any variables directly related to protected classes**
- **External expert on disparate impact reviewed initial ISO models**
- **Some states developed questionnaires to help evaluation of models**
- **NAIC CASTF is working a white paper of best practices for review of GLMs in personal auto and homeowners**
  - Current draft has 80 information items about the model development
  - 4 levels of information, about 35 needed with initial filing
  - No explicit info requested on protected classes

Decision Trees and Random Forests

- **Machine learning techniques that help determine conclusions about a target value based on observations of related input values**
  - Multivariate and non-linear, so the relationships between different variables and between the different variables and the target are not fixed throughout the range of possible values all variables can take on
    - Can result in better predictions than traditional linear techniques
  - May be used to model discrete or continuous outcomes; classification trees or regression trees, respectively
- **Decision tree – a single "flow chart" to get from inputs to outputs via Boolean logic**
  - Easy to understand, flexible for different applications, white box model
  - Non-robust, potentially over-fitted/biased to input data
- **Random forest – made of a multitude of decision trees to increase robustness and mitigate overfitting**
  - Individual trees may be constructed differently (different input variables, different node criteria) and their individual outcomes weighted together to determine a single outcome
  - More robust, but results may be less intuitive due to needing to interpret multiple non-linear trees



Who is Looking At These Issues?

- **NAIC Working Groups and Task Forces**
- **Innovation & Technology (EX) Task Force**
- **Big Data Working Group**
- **Artificial Intelligence Working Group**
  - OECD AI Principles exposure
- **Accelerated Underwriting Working Group**
  - Adopted work plan

NY Circular Letter 1

- **On the P/C side state specific restrictions may exist, to varying degrees, related to use of certain variables, such as gender related restrictions for personal auto found in California, Hawaii, Massachusetts, Montana, North Carolina, and Pennsylvania**

- **NY issued circular letter 1 in January concerning life insurance and stating, in part, that insurers "should not use external data sources, algorithms or predictive models in underwriting or rating unless the insurer has determined that the processes do not collect or utilize prohibited criteria and that the use of the external data sources, algorithms or predictive models are not unfairly discriminatory"**

- **Potential challenges in responding to this circular letter include:**
  - ✓ Most insurers do not collect, nor do they want to collect, information on protected classes that may be necessary to evaluate relative to circular letter.
  - ✓ What is the statistical criteria to show that there is no unfair discrimination?
  - ✓ Some insurers may not be questioned on the model until a market conduct exam which would be after the model/variables have been in use

NY Circular Letter – Excerpts

- **First, an insurer should not use an external data source, algorithm or predictive model in underwriting or rating unless the insurer has determined that the external tools or data sources do not collect or utilize prohibited criteria. An insurer may not simply rely on a vendor's claim of non-discrimination or the proprietary nature of a third-party process as a justification for a failure to independently determine compliance with anti-discrimination laws. The burden remains with the insurer at all times.**

- **Second, an insurer should not use an external data source, algorithm or predictive model in underwriting or rating unless the insurer can establish that the underwriting or rating guidelines are not unfairly discriminatory in violation of Articles 26 and 42. In evaluating whether an underwriting or rating guideline derived from external data sources or information is unfairly discriminatory, an insurer should consider the following questions:**

  1) Is the underwriting or rating guideline that is derived, in whole or in part, from external data sources or information supported by generally accepted actuarial principles or actual or reasonably anticipated experience that justifies different results for similarly situated applicants?

  2) Is there a valid explanation or rationale for the differential treatment of similarly situated applicants reflected by the underwriting or rating guideline that is derived, in whole or in part, from external data sources or information?

2018 ANNUAL MEETING & EXHIBIT

Thank you

SOCIETY OF ACTUARIES®

Ladder

Bias, fairness, and discrimination issues in the use of statistical modeling

Tasha Cupp
Ladder Financial Inc.
General Counsel

Disclaimer

The following presentation is for general information, education and discussion purposes only, in connection with the SOA Conference 2019. Any views or opinions expressed are those of the presenters alone. They do not constitute legal or professional advice; and do not necessarily reflect, in whole or in part, any corporate position, opinion or view of Ladder Financial Inc. or its affiliates, or a corporate endorsement, position or preference with respect to any issue or area covered in the presentation.

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

How Ladder Thinks About the Use of Models

Will Use of the Models Further Our Mission?

- Our mission is to make life better for all consumers by expanding access to life insurance.

- Any models we consider adopting must fit with this mission.

Education is Critical

- There is no substitute for reading the academic literature; there's a lot of incredible work being done in this area.

- We also monitor legal and regulatory developments as it relates to issues of discrimination in the use of these models.

- Finally, we try to participate in the conversation by offering ourselves as a resource, and by listening to the concerns of others.

How Can We Improve the Models We Use?

- The assessment of each of our models is continuous; we have a cross-functional team that works to ensure that each model remains well-calibrated.

- With each model, our team is tasked with finding where there is room to improve upon the choices we've made about outcomes, predictors, and training procedures.

- Our team is constantly seeking out opportunities to employ new learnings and to benefit from new research in the field.

What Are Our Biggest Challenges?

People Sometimes Fear the Unfamiliar

- "It is tempting to think that human decision-making is transparent and that algorithms are opaque…. [W]ith respect to discrimination, the opposite is true. The use of algorithms offers far greater clarity and transparency about the ingredients and motivations of decisions, and hence far greater opportunity to ferret out discrimination."

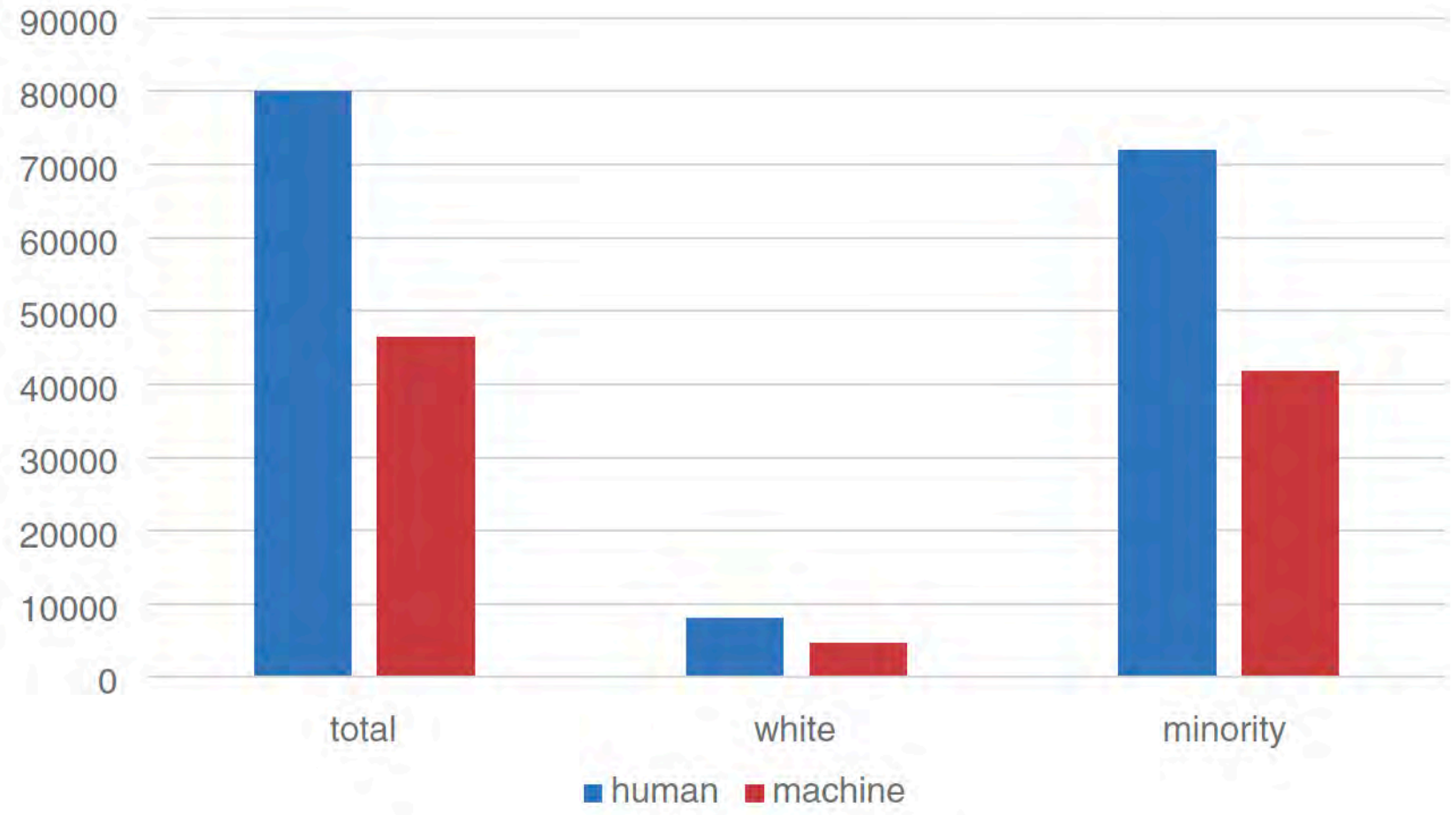
- "Algorithms have the potential to help us excise disparate treatment, to reduce discrimination relative to human decision-making, to limit disparate impacts, and also to predict much more accurately than humans can in ways that disproportionately benefit disadvantaged groups…."

- Yes, there are challenges that need to be navigated when using models. But let's not let a fear of the unfamiliar stop us from tackling these challenges and realizing the benefit of these models.

Source: Kleinberg, Jon, et al., "Discrimination in the Age of Algorithms," Journal of Legal Analysis, Vol. 10, 2018, laz001, Apr. 22, 2019.

An Example: Pre-trial Release Decisions

Figure 3. Pre-trial detention rates in New York City under current human (judge) decisions versus algorithmic release rule that holds failure to appear (FTA) rate constant at current level.



Source: Kleinberg et al., p. 50.

Today's Regulatory Framework is Made for Human Decision-making

- Today’s regulatory framework generally tries to safeguard against the biases of people by prohibiting the use of protected characteristics in decision-making processes.

- This approach doesn’t work with models. With models, we need the ability to collect and use information on protected characteristics in order to test for disparate impact and, in some cases, mitigate the discriminatory effects of biases in the historical data.

- Let’s engage in thoughtful dialogue about the appropriate regulatory framework for models.

Questions?


#ladderlove

SOCIETY OF ACTUARIES

Session Presented By:

Predictive Analytics and Futurism Section

Provides opportunities for actuaries to deepen their understanding of predictive analytics and emerging technologies relevant to the future of the actuarial profession and insurance industry.

Section Developed Content & Benefits

Predictive Analytics and Futurism Newsletter

Discusses futurism and the latest predictive analytics trends. Published three times a year. Digital editions now available.

SOA Meetings and Seminars

Section developed content presented during meeting sessions and seminars.

Podcasts

Expert led technical podcasts exploring the latest predictive analytics concepts and techniques.

Webcasts

Discounts on section developed webcasts. Free access to section created webcasts over one-year old.

Join the PAF Section Today! ***SOA.org/PAF***

CONNECT WITH SECTION MEMBERS