An Actuarial View of Data Bias

Definitions, Impacts and Considerations

Academy **Webinar**

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

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.

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Today's Presenters – Members of the Data Science and Analytics Committee

Shawna Ackerman, MAAA, FCAS

Liaw Huang, MAAA, FSA, FCA, EA

Reese Mularz, MAAA, FCAS

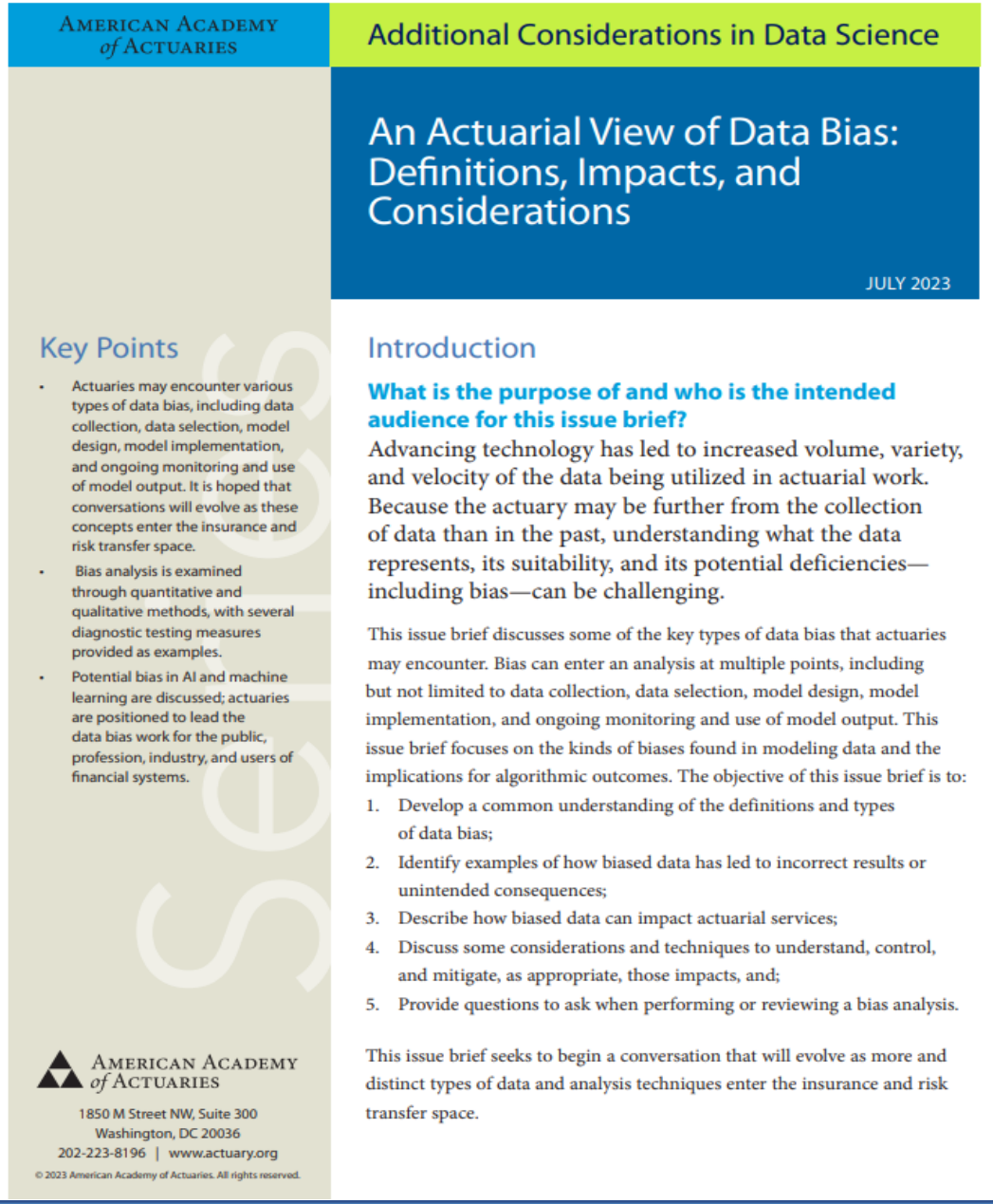
Dorothy Andrews, MAAA, ASA - Chairperson

Data Science and Analytics Committee

To further the actuarial profession's involvement in Big Data and machine learning technologies and to inform public policy decision making related to the use of Big Data, predictive models, and other advanced analytics in unbiased and objective terms. The committee monitors federal legislation and regulatory activities and is charged with developing papers intended to educate stakeholders.

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

[risk_brief_data_bias.pdf (actuary.org)]



American Academy of Actuaries

Additional Considerations in Data Science

An Actuarial View of Data Bias: Definitions, Impacts, and Considerations

JULY 2023

Key Points

- Actuaries may encounter various types of data bias, including data collection, data selection, model design, model implementation, and ongoing monitoring and use of model output. It is hoped that conversations will evolve as these concepts enter the insurance and risk transfer space.
- Bias analysis is examined through quantitative and qualitative methods, with several diagnostic testing measures provided as examples.
- Potential bias in AI and machine learning are discussed; actuaries are positioned to lead the data bias work for the public, profession, industry, and users of financial systems.

American Academy of Actuaries
1850 M Street NW, Suite 300
Washington, DC 20036
202-223-8196 | www.actuary.org
© 2023 American Academy of Actuaries. All rights reserved.

Introduction

What is the purpose of and who is the intended audience for this issue brief?

Advancing technology has led to increased volume, variety, and velocity of the data being utilized in actuarial work. Because the actuary may be further from the collection of data than in the past, understanding what the data represents, its suitability, and its potential deficiencies—including bias—can be challenging.

This issue brief discusses some of the key types of data bias that actuaries may encounter. Bias can enter an analysis at multiple points, including but not limited to data collection, data selection, model design, model implementation, and ongoing monitoring and use of model output. This issue brief focuses on the kinds of biases found in modeling data and the implications for algorithmic outcomes. The objective of this issue brief is to:

1. Develop a common understanding of the definitions and types of data bias;
2. Identify examples of how biased data has led to incorrect results or unintended consequences;
3. Describe how biased data can impact actuarial services;
4. Discuss some considerations and techniques to understand, control, and mitigate, as appropriate, those impacts, and;
5. Provide questions to ask when performing or reviewing a bias analysis.

This issue brief seeks to begin a conversation that will evolve as more and distinct types of data and analysis techniques enter the insurance and risk transfer space.

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Agenda

- Introduction
- General problems associated with data bias
- How can data bias impact actuarial services
- Considerations in performing a bias analysis
- Considerations in reviewing a bias analysis
- Continued work on data usage and bias issues

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Introduction

- Proliferation of data types and sources used in actuarial work
- More models and more complex models
- Exponential increase in awareness of and concern for model risk and bias
- Bias can enter at each point of the model development pipeline
  - Data selection, processing, use
  - Assumptions
  - Modeling

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Introduction

The National Institute of Standards and Technology (NIST) groups bias into three categories: **statistical bias**, **cognitive (human) bias**, and **systemic bias**.

- **Statistical bias**: the degree to which the estimate differs from the truth
- **Cognitive (human) bias**:
  - Who are the decisionmakers?
  - How is the information perceived?
- **Systemic bias**: data reflects the people, institution, society that created, captured and curated it

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Introduction

Two general conditions that lead to data bias:

1. The dataset is not representative of the underlying population for which the prediction or algorithm will be used, and/or
2. The method used to collect, process, use and interpret the data is flawed.

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Representation Bias

The dataset is not representative of the underlying population. This can arise due to:

- Inadequate sampling – for example, dataset collected from smartphone apps may under-represent lower-income and older groups; data collected from voluntary responses (response bias or self-selection bias); lack of geographical diversity; non-random sampling (sampling bias).
- The population of interest has changed since the data collection – for example, data collected in one time frame used for another (temporal bias).
- Historical - Existing biases in society can persist in the data generation process even with a perfect sampling and feature generation.

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Cognitive Bias

There are over 180 identified cognitive biases describing, for example, how we

- Generalize information
- Favor simple solutions
- Project our current mindset and assumptions on to the past and future
- Are drawn to details that confirm our existing beliefs
- Find patterns

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Key points

- Bias exists in the data, in collection, processing, use and interpretation
- Challenging to eliminate the bias entirely
- Awareness of the limitations
- Multiple strategies are needed:
  - Diverse and representative datasets
  - Broad range of perspectives
  - Human reviewers following guidelines to review and rate the model
  - Feedback

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

General problems associated with data biases

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

General Problems Associated with Data Biases

Incorrect conclusions

I may end up with wrong predictions

My data may not reflect reality

I may perpetuate existing biases without knowing

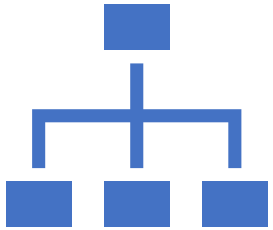
Unwanted consequences

I may be surprised by the spurious correlations in the outcome

My results “did not make sense” or “missed the point”

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

General Problems Associated with Data Biases



Inadequate system performance

My system exhibits uneven performance among subgroups

I cannot trust the system in all situations

Misinformed policy decisions

I need to be reminded that other people may be impacted by the decisions made by the system

I may under-correct or over-correct mistakes, or overlook signals that indicate serious problems

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

How can data bias impact actuarial services?

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

How can data bias impact actuarial services?

- Actuarial services are data-driven
- Just as in the general case, data bias can impact actuarial services
  - Incorrect conclusions
  - Unwanted consequences
  - Inadequate system performance
  - Misinformed policy decisions
- It is important to be aware of the various types of data bias to assess potential impacts

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

How can data bias impact actuarial services?

Guiding Actuarial Standards, Practices, and Considerations

- ASOP 56, Modeling: Assists actuaries in designing, developing, selecting, modifying, using, reviewing, or evaluating models
  - Sampling Bias
    - A subset of data is gathered which is intended to reflect the population as a whole
    - If this subset does not properly reflect the entire population, the model may be inappropriate for its intended use
- ASOP 23, Data Quality: Provides guidance when performing actuarial services involving data
  - Requires actuary to disclose the potential existence of bias if, after adjustments and assumptions are applied, the results may be highly uncertain or contain significant bias

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

How can data bias impact actuarial services?

Guiding Actuarial Standards, Practices, and Considerations

- Qualification Standards for Actuaries Issuing Statement of Actuarial Opinion in the United States
  - One hour must be on bias topics
  - FAQ #51, 53

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Examples

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Risk Classification

- Actuarial Service: Determine expected value of future costs associated with individual transfer of risk
- Historical Bias
  - Homeownership rates have historically differed by race
  - Using homeownership in a personal auto rating plan could introduce bias
  - Using the true drivers of expected costs, such as experience and driving record, would help limit this bias
- Availability Bias
  - Recent large increase in shark attacks on the East Coast
  - Life insurers may restrict underwriting guidelines on surfers

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Reserving

- Actuarial Service: Estimate value of future claims and expenses
- Aggregation Bias
  - Long-tailed liability and short-tailed property data does not develop similarly
  - If aggregated, the selected development pattern may hold in the aggregate but not if applied to the individual lines
- Confirmation Bias
  - Management may prefer favorable development on prior reserves
  - The actuary, being aware of this, may tend towards a priori loss ratio and development assumptions which tilt the results towards management's preference

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Modeling

- Actuarial Service: Utilize advanced analytical techniques to enhance analysis and decision making
- Omitted Variable Bias
  - Leaving out important variables can cause signal to be lost or other variables attempt to account for the lost signal
  - For homeowners insurance, leaving out claim history could cause other variables such as roof age and construction type to account for the claim history signal
  - This may cause roof age and construction type to give nonsensical results
- Confirmation Bias
  - The actuary may expect a certain result for some variables in the model
  - If the variables do not follow the expected result, the actuary may tweak the model until the result aligns with this initial expectation

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Considerations in performing a bias analysis

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Understanding Bias Analyses

Basic Approaches

Quantitative

- Statistical analysis of modeling or training data
  - Sufficiency, balance, credibility, and representativeness
  - May capture many types of biases, such as sampling, measurement, selection bias, etc.
- Scenario and sensitivity testing of modeling parameters

Qualitative

- Social science research
  - Better suited to identifying biases stemming from historical policies and practices
- Diverse modeling team and SMEs

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Components of Bias Analysis

Define the approach – where you are going to chase potential bias

- Data
- Algorithms
- Outcomes

Define the measure of fairness

- Different definitions may exist depending on the modeling intent
- Fairness Tree might help drive the discussion with stakeholders

Define the metrics

- Metrics based on group fairness, i.e., the focus is to have equality of metrics across groups
- Metrics based on individual fairness, i.e., require a similar classification for similar individuals

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Bias Metrics in Regulatory Context

Four-Fifth Rule (80% rule) - HR

- Goal: avoid discrimination in hiring
- The Rule of Thumb established by federal guidelines on hiring
  - Proportion of positive outcomes in protected class should be no less than 80% of proportion of positive outcome in reference class.
  - Not a legal definition, but practical measure of potential discrepancy in hiring, promotion, etc.

Ratio Percentage Test – Defined Benefit Pension

- Goal: avoid discrimination against non-highly compensated employees
- % of non-highly compensated employees benefiting from the plan must be at least 70% of the % of highly compensated employees

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Analysis of Bias in Data

- Exploratory data analysis
- Feature importance
- Wealth of open-source toolkits
  - **IBM's AI Fairness 360:** comprehensive toolkit to analyze and mitigate bias at every step of the pipeline
  - **Aequitas:** bias audit toolkit from the authors of the Fairness Tree
  - **FairML:** auditing tool for analyzing the relative effects of various inputs on a model's predictions
  - **Fairness Measures:** toolkit including several fairness metrics, such as difference of means, disparate impact, and odds ratio.
- Most if not all ML pipeline providers (e.g., Amazon Sagemaker, Microsoft's Azure ML, DataRobot, etc.) include some sort of bias and fairness analysis metrics

Challenges:

- Some of the different metrics have inherent trade-offs (fairness vs. accuracy)
- Thresholds for metrics need to be developed

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Mitigating Bias in Machine Learning

- Pre-processing, or input data
  - Reweighting
  - Re- and oversampling
  - Data Transformation (in a smart way)
- In-processing
  - Introducing constraints and penalties into algorithms (e.g., regularization)
  - Adversarial debiasing, where an adversarial model aims to predicts the protected characteristics based on the predictor model
- Post-processing
  - Adjusting outcomes to achieve parity in false positive and/or false negative rates

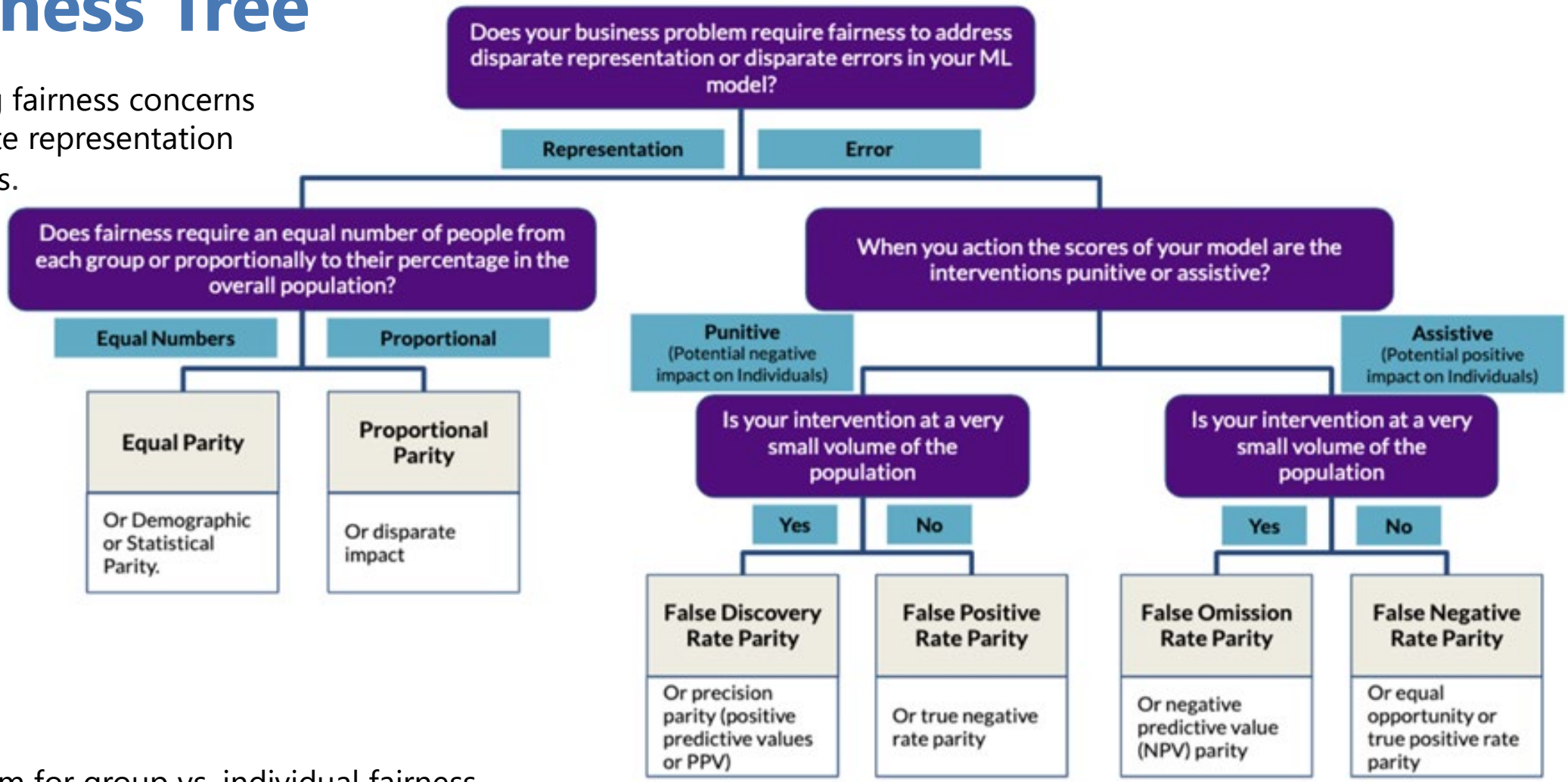
© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Considerations in reviewing a bias analysis

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

The Fairness Tree

Aids in navigating fairness concerns based on disparate representation or disparate errors.



Some methods aim for group vs. individual fairness.

Source: Bias and Algorithmic Fairness. The modern business leader's new... | by Jan Teichmann | Towards Data Science

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

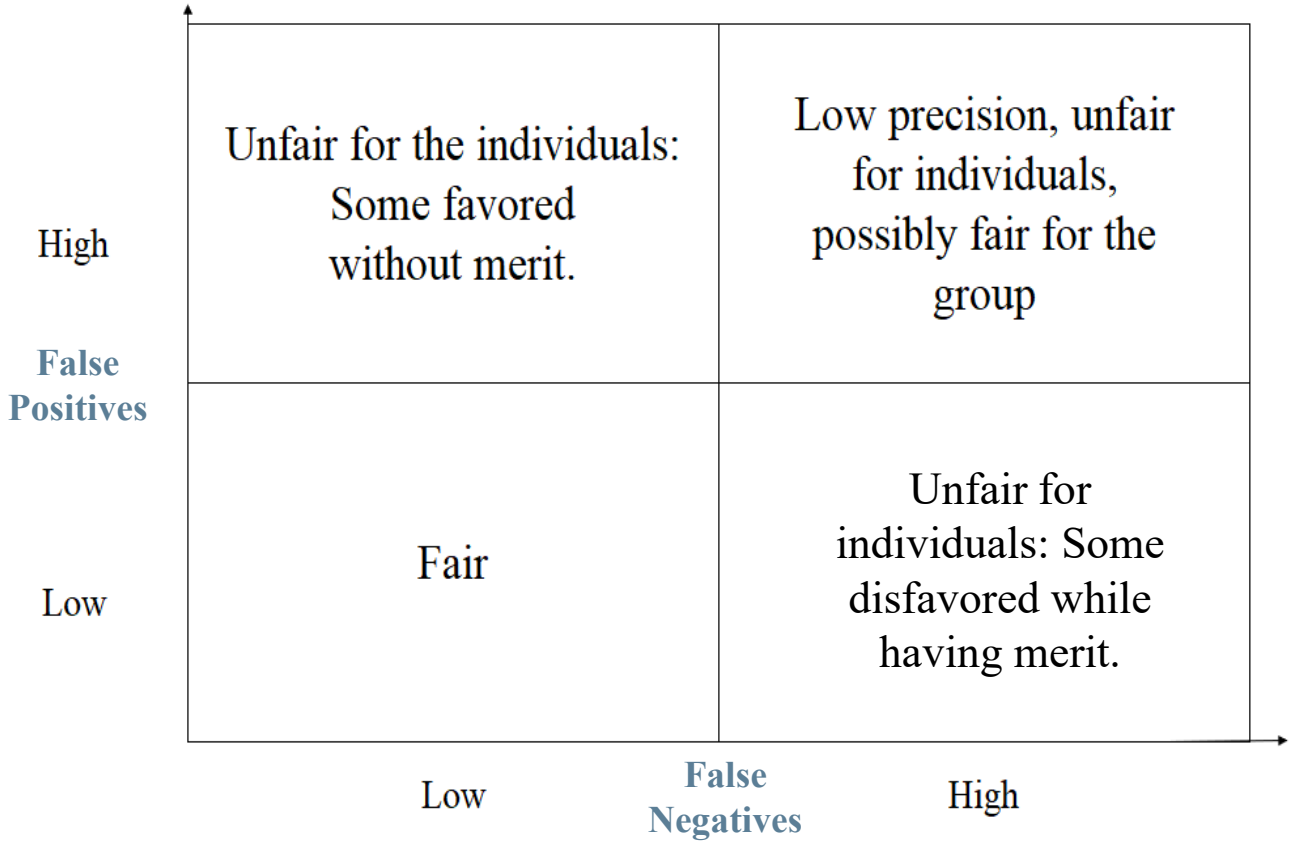
Group vs. Individual Fairness

Group Fairness Concepts

- Egalitarianism
- Anti-Discrimination
- Equity
- More Costly

Individual Fairness Concepts

- Consistency
- Individual Justice
- Equality
- Less Costly



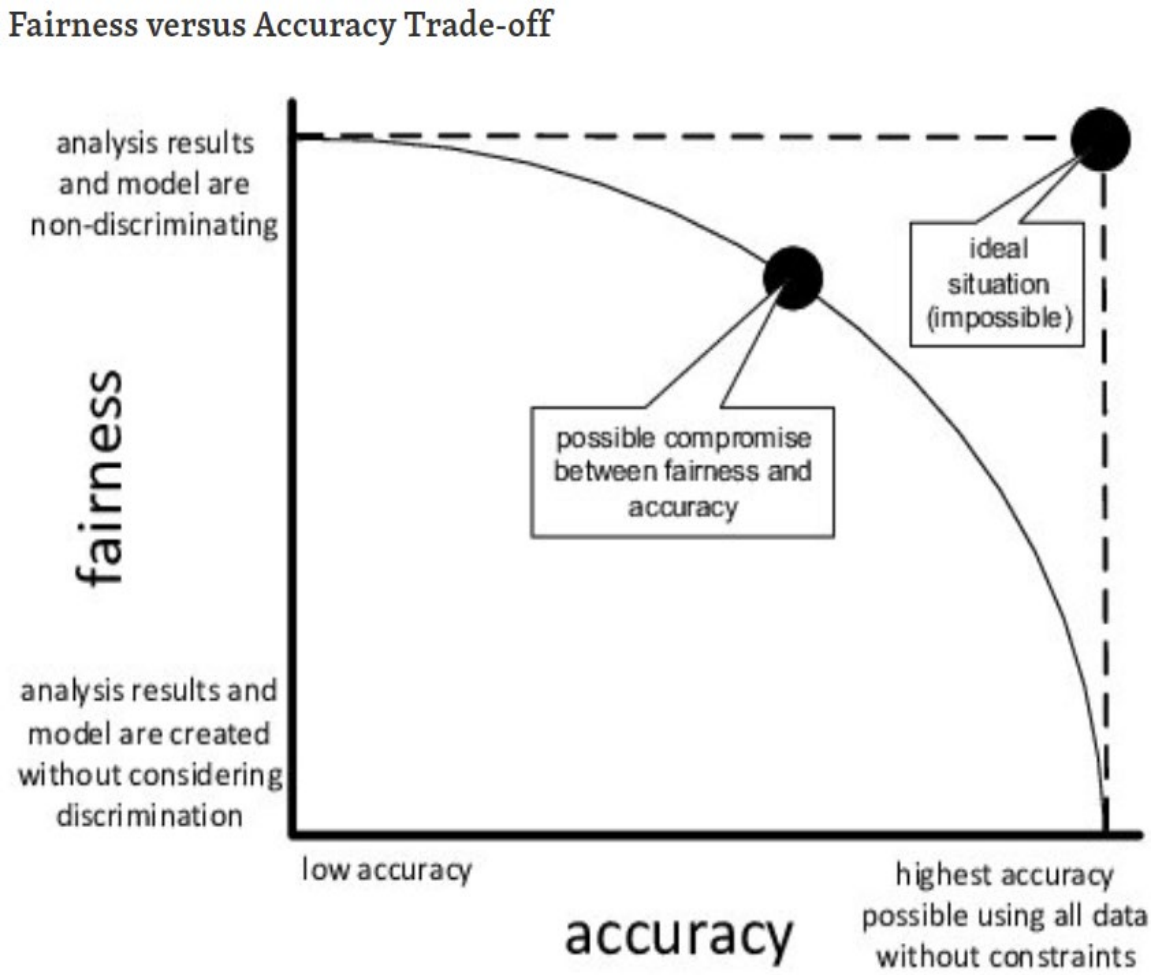
Impossible to satisfy Group and Individual Fairness simultaneously!

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Fairness v. Accuracy Trade-Off

Basic Tenet:
Fairness Comes at a Cost

- Fairness Quantification is Difficult
- Increasing Fairness, ↓ Accuracy
- ML Models Optimize for Accuracy
- ML Models Don't Optimized for People
- Intersectionality May Be Sacrificed
- Fairness is a Continuous Process
- There is NO Single Fairness Checkpoint!
- Strong Internal Governance is Needed



© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Interrogating Data and Models for Bias

Four Types of Diagnostic Questioning

- General
- Data Related
- Model Related
- Social Systemic Related

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Evaluating A Bias Analysis

General	Data	Model	Social Systemic
Purpose of Algorithm	Data Biases Tested	Parameter Assignment	Systemic & Social Links
Measures of Fairness	Links to Discrimination	Treatment of Offsets	Socioeconomic
Objective of Analysis	Demographic Diversity	Interrater Reliability	Behavioral
Bias Thresholds	Demographic Balance	Sensitivity Analysis	Telematics
Bias Sources Identified	Attribute Rate Analysis	Aggregation Bias Check	Crime & Census Data
Diversity of Reviewers	Historical Bias Presence	Model Success Definition	Consumer Data
Reference Groups	Historical Bias Tweaks	Outcome Harm Analysis	Price Optimization
Fairness Standards	Weight Assignment	False Positives & Negatives	Social Science Links
Error Analysis	Age of Data	Human Oversight Needs	Insurer Action Deltas

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

General Related Diagnostic Questions

- What was the original purpose of the algorithm targeted in the bias analysis?
- What are the measures of fairness used in the bias analysis?
- What are the objectives of the bias analysis and were they achieved?
- What is the threshold for measuring and correcting bias in the algorithm?
- Were multiple sources of bias detected, e.g., from data collection, data processing, etc.?
- How diverse is the group of people that conducted or reviewed the bias analysis?
- What was the reference group(s) in the analysis against which other groups were compared?
- What was the fairness standard? How were unequal outcomes assessed against the fairness standard? How was an unequal outcome judged to be fair or unfair?
- Were there more errors for some groups versus the reference group?

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Data Related Diagnostic Questions

- What were the types of biases in the data that were tested?
- Are any data elements linked to a history of discriminatory policies and practices?
- How diverse is the sample based on demographic factors like age, sex, race, ZIP code, and credit score?
- What is the composition within the groupings in the dataset, e.g., the racial and/or ethnic makeup in the ZIP codes in the data?
- What is the balance of demographic factors across variable factor levels?
- What is the demographic profile of consumers that get the highest rates?
- How was the data evaluated for historical bias? How were model results adjusted for historical bias?
- Which variables were assigned the greatest weights?

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Model Related Diagnostic Questions

- How were model parameters assigned? How were they assessed for bias?
- Were independent experts employed to review the results? What was the interrater reliability?
- How sensitive are the analysis outcomes to small changes in a data point or different samples?
- What tests were performed to determine whether data groupings did not result in aggregation bias?
- How was model success defined? Was the metric being used to measure success biased?
- Where does the model get the classification wrong and which demographic is most affected? Why does the model get the classification wrong?
- Was a demographic analysis of the false-positive rates provided?
- How much human oversight is required to implement the model?

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Social Systemic Related Diagnostic Questions

- Can any of the model variables be linked to a history of systemic discrimination?
- Do any of the variables fall into one or more of the following categories:
  (a) Socioeconomic
  (b) Behavioral
  (c) Telematics
  (d) Crime
- Is there social science research supporting the adversely discriminatory effects of model variables?
- Do company actions differ across groups when the scores are similar across those groups?
- Can the variable be traced to historical practices and policies that are adversely discriminatory?

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Possible Approaches for Mitigating Bias

- Remove proxies for protected attributes from modeling data
- Ensure modeling data is representative of all applicable risks
- Expand modeling team for diversity and Interdisciplinarity
- Apply post-model development bias correcting adjustments
- Regulator Determined **Solidarity Tax and Rebate (STR)**
- Only use variables relates to the risk being insured.
- Apply debiasing methodologies - AI Fairness 360 library
- Keep a Human in the Loop!

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Innovation Cybersecurity and Technology (H) Committee

NAIC Model Bulletin

Sets forth expectations as to how Insurers will govern the development and acquisition and use of AI systems.

Finalized December 2023

Treasury RFI on use of AI in financial services – June 2024

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Biden Executive Order on AI

Are there implications for insurance companies?

Eight Principles of the Order:

1. Must be Safe & Secure
2. Promote Responsible Innovation
3. Commit to Supporting American Workers
4. Advance Equity and Civil Rights
5. Protects Interests of Americans
6. Protect Privacy and Civil Liberties
7. Increase Capacity to Regulate, Govern and Support Responsible AI
8. Engage with International Partners to develop an AI Governance Framework

National Institute of Standards & Technology (NIST) is tasked with a leading role implementing the EO.

The New York Times

Biden Issues Executive Order to Create A.I. Safeguards

The sweeping order is a first step as the Biden administration seeks to put guardrails on a global technology that offers great promise but also carries significant dangers.

The order is an effort by President Biden to show that the United States, considered the leading power in fast-moving artificial intelligence technology, will also take the lead in its regulation. Doug Mills/The New York Times

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

U.S. ARTIFICIAL INTELLIGENCE SAFETY INSTITUTE

Artificial Intelligence Safety Institute Consortium (AISIC)

In support of efforts to create safe and trustworthy artificial intelligence (AI), NIST has established the U.S. Artificial Intelligence Safety Institute (USAISI). To support this Institute, NIST has created the U.S. AI Safety Institute Consortium. The Consortium brings together more than 280 organizations to develop science-based and empirically backed guidelines and standards for AI measurement and policy, laying the foundation for AI safety across the world. This will help ready the U.S. to address the capabilities of the next generation of AI models or systems, from frontier models to new applications and approaches, with appropriate risk management strategies.

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Continued work on data usage and bias issues

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.

Other AAA Publications to BOLO:

- Defining Big Data
- Defining Data as an Asset
- Natural Experiments
- Auditing Algorithms for Bias
- Using Entropy to Identify Bias in Data

American Academy of Actuaries

Additional Considerations in Data Science

Discrimination: Considerations for Machine Learning, AI Models, and Underlying Data

February 2024

Key Points

- Unfair discrimination takes place when insurers consider factors that are unrelated to actuarial risk while determining whether to provide insurance to particular individuals or groups, and if so, at what price and with what terms.
- Insurance legislation has put in place measures to prevent unfair discrimination while still permitting actuarially justified risk selection. However, within insurance companies, various functions like marketing, rating, and underwriting have become more reliant on big data, algorithms, and machine learning. These processes might utilize variables that appear neutral on the surface but can lead to unequal impacts on different groups of people.
- Discrimination can originate from multiple sources, including the data, the algorithm, and the overall models used in these practices.

This issue brief explores the topic of discrimination in machine learning algorithms and artificial intelligence (AI) algorithms, and the underlying data of these models. It will define discrimination (including distinguishing between discrimination, unfair discrimination, and unjust discrimination); present practical methods for testing and monitoring algorithms; provide a regulatory overview of the issue; and identify considerations for actuaries, algorithm creators, and regulators.

The following topics are discussed in the issue brief:

I. Defining discrimination
Includes a high-level discussion of issues around unlawful, unfair, and discrimination, and several case studies highlighting the challenges with AI/machine learning models and their potential to discriminate.

II. Identifying discrimination through disparate impacts in models
Includes qualitative and quantitative testing options, a discussion of protected groups and proxy variables, monitoring activities, and suggestions for a company's framework around model governance.

III. Regulatory landscape and additional considerations
Includes an overview of the regulatory landscape surrounding this issue, resources for actuaries, and considerations for insurers.

American Academy of Actuaries
1850 M Street NW, Suite 300
Washington, DC 20036
202-223-8196 | www.actuary.org

Updated © February 2024.
American Academy of Actuaries. All rights reserved.

risk-brief-discrimination.pdf (actuary.org)

© 2023 American Academy of Actuaries. All rights reserved.
May not be reproduced without express permission.