SOA Research INSTITUTE
SOCIETAL PURPOSE

Mitigating Potential Unwanted Bias in Life and Annuity Insurance Products and Processes:

A Collection of Essays

October | 2025

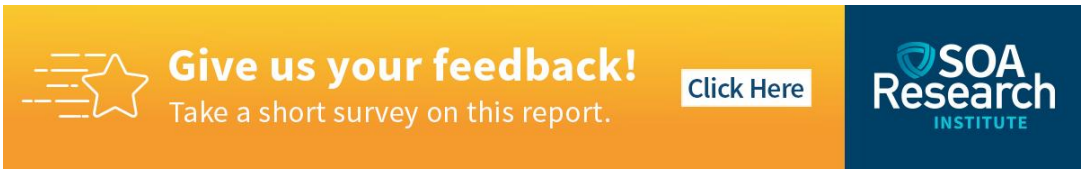
Mitigating Potential Unwanted Bias in Life and Annuity Insurance Products and Processes

A Collection of Essays

Contents

Caveat and Disclaimer
The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute, the Society of Actuaries or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

Copyright © 2025 by the Society of Actuaries Research Institute. All rights reserved.



Copyright © 2023 Society of Actuaries Research Institute

Mitigating Potential Unwanted Bias in Life and Annuity Insurance Products and Processes

A Collection of Essays

Introduction and Acknowledgments

INTRODUCTION

In early 2025, the Society of Actuaries (SOA) Research Institute issued a call for essays to collect thoughts and perspectives on specific opportunities for avoiding or mitigating potential unwanted bias in the products and processes associated with life and annuity insurance. Essays were intended to meet one of two objectives with respect to any specific process, part of a process, or method that is used at any point in the development, distribution, or administration of annuity or life insurance products:

(1) Increase awareness of the potential for unwanted bias that may exist in a specific process, part of process, or method that may currently be used in any point along the entire product value chain for life or annuity insurance products; and
(2) Offer methods, techniques, procedures, or approaches for eliminating or reducing the likelihood that unwanted bias could exist in that specific process, part of a process, or method.

A project oversight group (POG) reviewed blinded versions of the essays, and judged them for publication and awards. Judging criteria included creativity, originality, and the extent to which an idea might help promote the elimination or reduction of the likelihood of potential unwanted bias in specific life or annuity insurance processes and products. The POG selected four essays for publication and awarded each a prize.

AWARD WINNERS

Eliminating Potential Historical Data Biases in Life and Annuity Insurance Pricing: A Framework for Fairness and Transparency
Siyu Chen, FSA

Beyond Redlining: Addressing Potential Zip Code Bias in Life Insurance Pricing
Joshua Owusu

Equity Underwritten: Mitigating Bias in Risk Assessment and Pricing Processes
Marco Pirra, AFFI, CAS

Reimagining Underwriting in Life and Annuity Insurance
Niranjan Rajendran, B.Sc. (Hons)

THE CALL FOR ESSAYS

At the Society of Actuaries Research Institute, calls for essays are substantively different from calls for short research papers. Research Institute research papers are required to be fact-based and objective and to

Copyright © 2025 Society of Actuaries Research Institute

avoid advocacy, especially with respect to public policy. Research papers published by the Research Institute may inform readers about public policy topics but must refrain from taking a position on or advocating for a public policy issue.

Essays that the Research Institute published may be fact-based, short research papers. Alternatively, they may be more experiential in nature as a means of highlighting issues or calling for change, although they must refrain from advocating for or taking a position on a specific legislative or regulatory initiative. Both types of essays were invited in this call for essays, and both types of essays are included in this collection.

For context, the two sections of the call for essays that outline the subject matter request are replicated below.

OVERVIEW

The Society of Actuaries Research Institute (SOA) is interested in collecting thoughts and perspectives on specific opportunities for avoiding or mitigating potential unwanted bias in the products and processes associated with life and annuity insurance processes. Results of this call for essays are intended to meet two objectives: (1) Increase awareness of the potential for unwanted bias that may exist in a specific process, part of process, or method that may currently be used in any point along the entire product value chain for life or annuity insurance products; and (2) Offer methods, techniques, procedures, or approaches for eliminating or reducing the likelihood that unwanted bias could exist in that specific process, part of a process, or method.

SAMPLE TOPICS OF INTEREST FOR ESSAYS

This invitation for essays allows essay authors to choose as their subject any specific process, part of a process, or method that is used at any point in the development, distribution, or administration of annuity or life insurance products. Life and annuity insurance product value chains typically include:

- Product design and development processes, which include but are not limited to:
    - Product conceptualization
    - Market research
    - Data analysis
    - Risk assessment
    - Pricing
    - Product design and structuring
    - Financial modeling and projections
    - Regulatory compliance review
- Marketing, sales, and distribution processes, which include but are not limited to:
    - Market analysis, target market determination, and market segmentation
    - Analyzing competitor pricing
    - Analyzing market trends
    - Analyzing customer price sensitivity
    - Modeling to forecast sales volumes and revenue
    - Assessing the cost-effectiveness and reach of distribution channels (e.g., agents, brokers, online platforms)
    - Evaluating the effectiveness of customer retention strategies and loyalty programs
    - Analyzing key performance indicators (KPIs) such as conversion rates, customer acquisition costs, and return on investment (ROI)
    - Regulatory compliance review
- Underwriting- and pricing-related processes, which include but are not limited to:
    - Developing criteria for accepting or rejecting applications
    - Defining risk classes

Copyright © 2025 Society of Actuaries Research Institute

  - Determining the terms and pricing of insurance policies
  - Assessing applicants' risk factors
- Operations and technology processes, which include but are not limited to:
  - Policy administration
  - Customer service
  - Use of technology to streamline processes
- Claims management processes, which include but are not limited to:
  - Assessing claims
  - Reviewing/processing claims
  - Settling claims

ACKNOWLEDGMENTS

The SOA Research Institute thanks the Project Oversight Group (POG) for their careful review and judging of the submitted essays. Any views and ideas expressed in the essays are the authors' alone and may not reflect the POG's views and ideas nor those of their employers, the authors' employers, the Society of Actuaries, the Society of Actuaries Research Institute, nor Society of Actuaries members.

Dorothy Andrews, ASA, MAAA

Kaitlin Creighton, FSA, MAAA, CERA

Mohammed Amine Elmeghni, FSA, MAAA

Patricia Fay, FSA, MAAA

Robert Gomez, FSA, MAAA, CERA

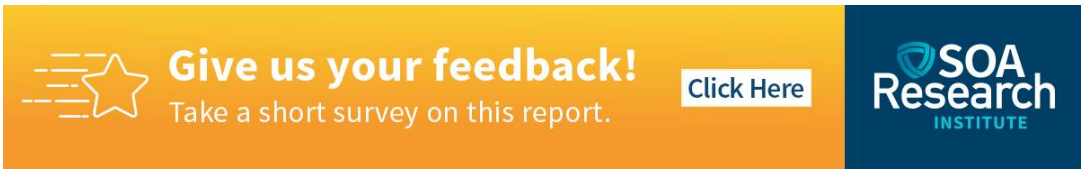
Damion Gooden, FSA, EA, MAAA

Shisheng (Rose) Qian, FSA, CERA

John Robinson, FSA, MAAA

David Schraub, FSA, MAAA, CERA, ACA

Jonah, von der Embse, FSA, MAAA, CERA



Copyright © 2025 Society of Actuaries Research Institute

Award Winner

Eliminating Potential Historical Data Biases in Life and Annuity Insurance Pricing: A Framework for Fairness and Transparency

Siyu Chen, FSA

October 2025

The views and ideas expressed in this essay are the author's alone and do not represent the views or ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, or the author's employer.

INTRODUCTION

This article delves into the issue of using historical, potentially biased data in life and annuity insurance pricing models. It analyzes how such potential biases, rooted in demographic and geographical factors, may be perpetuated. A comprehensive three-dimensional solution framework is proposed, focusing on data reconstruction, model innovation, and product design integration to mitigate these biases while ensuring actuarial integrity.

PROBLEM BACKGROUND AND CURRENT INDUSTRY PRACTICES

Actuarial pricing for life and annuity insurance has long been anchored in historical data, including mortality, morbidity, and lapse rates. However, these datasets may not be neutral; they may mirror past societal inequalities. For instance, in annuity products in China, women are often charged higher premiums due to their longer average life expectancies according to China Life Insurance Mortality Table. This practice may fail to account for the narrowing gender gap in health outcomes brought about by modern medical advancements.

As regulatory bodies around the world start to prohibit discriminatory pricing,[1] and as consumers become more aware and demanding of transparency, insurers are under increasing pressure. Pricing models that continue to replicate potential historical biases not only risk legal consequences in relevant markets but also damage the company's reputation.

TYPES AND IMPACTS OF BIAS IN PRICING MODELS

DATA COLLECTION BIASES

Sample Selection Bias

Sample selection in historical data collection may lead to bias. For example, data may overrepresent certain groups, such as urban, high-income populations. This means that when insurers use such data for pricing,

Copyright © 2025 Society of Actuaries Research Institute

the needs and risks of other groups, like rural or low-income individuals, are misjudged. As a result, premiums for these underrepresented groups may be either overestimated or underestimated.

Measurement Bias

Measurement bias can also arise when health metrics indirectly reflect socioeconomic conditions rather than direct risk factors. For example, historical data may show higher cancer mortality rates in rural areas, but this correlation often reflects delayed diagnosis due to limited access to healthcare—not inherent biological risk. When insurers use these historical incidence rates directly in pricing, rural populations may be unfairly charged higher premiums, penalizing them for systemic gaps in medical infrastructure.

MODEL REINFORCEMENT EFFECTS

Generalized Linear Regression Models

Generalized linear regression models are widely used in experience studies.[2][3] These experience data serve as critical assumptions and foundational elements for actuarial pricing, but generalized linear regression models can exacerbate historical trends. If historical data shows that a particular region has a higher mortality rate, the model may simply assume that this trend will continue and set premiums accordingly, without considering changing factors or the root causes of the historical trend.

Machine Learning Models

Machine learning models, although powerful, can also uncover and amplify hidden biases.[4] These models may find correlations between certain demographic factors, like race, and risk factors, such as disease prevalence, without establishing a causal relationship. This can lead to discriminatory pricing based on these spurious correlations.

A THREE-DIMENSIONAL SOLUTION FRAMEWORK

In the proposed framework, the initial dimension is dedicated to data reconstruction. By harnessing advanced data-engineering techniques, it aims to tackle potential historical biases while upholding actuarial precision.

Dynamic Adjustment Factors

One approach to data reconstruction is the development of socioeconomic compensation coefficients.[5] These coefficients can be used to adjust raw historical data based on factors like healthcare access, education levels, or income distribution used in actuarial pricing. For example, a formula could be developed where the adjusted data is calculated by multiplying the raw data by a factor that takes into account the difference between the local healthcare access index and the national average.

The adjusted data calculation incorporates a multiplicative factor derived from normalized indices of systemic inequity:

$$\text{Adjusted data} = \text{Raw data} \times \left(1 + \frac{Local\ Index - National\ Benchmark}{National\ Benchmark}\right)$$

Where $Local\ Index$ represents a standardized measure of the relevant systemic factor (e.g., healthcare access, education attainment) for a specific demographic group or geographic area.

$National\ Benchmark$ represents the median or mean value of the same index across the entire population.

The proposed approach offers several advantages over traditional methods. Dynamic responsiveness is a key strength, as the coefficients adapt to real-time changes in systemic conditions, unlike static

Copyright © 2025 Society of Actuaries Research Institute

adjustments such as flat gender-based discounts (e.g., improvements in rural healthcare infrastructure). Moreover, transparency is enhanced because the formula explicitly links data adjustments to measurable societal factors, ensuring regulatory compliance and fostering public trust. Additionally, the framework demonstrates generalizability, as it can be extended to address multiple equity dimensions such as education and income by incorporating additional indices.

Synthetic Data & Counterfactual Modeling

To combat potential historical data biases in the insurance industry, two advanced techniques, synthetic data generation and counterfactual analysis,[6] offer promising solutions. These methodologies operate synergistically within the data reconstruction to eliminate potential inherent biases and validate the fairness of reconstructed datasets.

Synthetic data generation, particularly through Generative Adversarial Networks (GANs), is a cutting-edge approach to creating unbiased datasets. GANs consist of two neural networks: a generator and a discriminator. The generator's role is to generate synthetic data that mimics real-world patterns related to mortality, morbidity, and other relevant insurance factors. Meanwhile, the discriminator assesses whether the generated data is statistically similar to the original real-world data. During the training process, sensitive attributes such as gender, race, and geographical location can be either excluded from the input data or adjusted so that they do not influence risk assessment. This way, the resulting synthetic data can be free from historical biases. For example, instead of reflecting historical gender-based differences in life expectancy, the synthetic mortality data can assume equal health outcomes for all genders.

Counterfactual analysis is another tool that provides a critical evaluation framework to quantify bias reduction in reconstructed datasets. It involves constructing a causal model that identifies the relationships between various factors, such as healthcare access, lifestyle choices, and risk levels. Once the causal model is established, insurers can simulate scenarios where potential historical biases are eliminated. For instance, they can assume that all regions have equal healthcare access regardless of their actual geographical and socioeconomic differences. By comparing the original pricing based on historical data with the counterfactual pricing, insurers can measure the extent of bias, if any, in the current pricing system. The counterfactual premium can be calculated using a formula like:

$$\mathrm{P}^{*} = f(X_i^{adjusted}, \theta)$$

Where $X_i^{adjusted}$ represents the adjusted set of attributes with biases removed, and θ represents the model parameters.

By integrating synthetic data generation and counterfactual analysis, insurers can design more equitable pricing models. Synthetic data provides a clean starting point for model training, while counterfactual analysis quantifies bias reduction and validates fairness. This combination not only helps in ensuring fairness in pricing but also enables insurers to meet regulatory requirements and build trust with customers. It transforms the way insurers use historical data, turning it from a source of potential bias into a tool for innovation and fairness in the insurance industry.

The second dimension of the proposed framework focuses on model innovation, leveraging advanced machine learning techniques to address potential bias while maintaining predictive accuracy.

Fairness—Constrained Modeling

Insurers can integrate fairness-enhancing algorithms into their pricing workflows to explicitly mitigate potentially biased outcomes. Tools like Fairlearn and AI Fairness 360[7] enable the enforcement of fairness constraints during model training, ensuring that predictions do not systematically favor or penalize specific

Copyright © 2025 Society of Actuaries Research Institute

groups (e.g., gender or race). For example, the ExponentiatedGradient algorithm in Fairlearn minimizes demographic parity disparities by adjusting model weights to balance prediction accuracy across subgroups. This is achieved through a constrained optimization process:

$$min_\theta E_{(X,Y)}\left[\ell\left(Y, f(X,\theta)\right)\right] \; s.t. \; DP_{group} \leq \varepsilon$$

where DP_{group} measures demographic parity, defined as equal true positive rates across genders, and ε is a tolerance threshold. By embedding such constraints, insurers can prevent models from replicating potential historical biases while preserving actuarial soundness.

Dynamic Risk Calibration

To mitigate bias and enhance fairness, insurers can adopt dynamic risk calibration, which replaces static demographic proxies with real-time behavioral data and advanced analytics. This approach integrates granular inputs such as telemedicine usage, fitness tracker metrics, and claim patterns to create personalized risk profiles. For example, wearable devices can monitor heart rate variability and physical activity levels, enabling insurers to adjust premiums based on actual health trends rather than historically assumed demographic stereotypes. Machine learning algorithms, such as recurrent neural networks (RNNs) for time-series analysis, can detect subtle patterns in this data to predict mortality or morbidity risks with greater precision.[8] Concurrently, explainability techniques like SHapley Additive exPlanations (SHAP)[9] decompose model decisions, ensuring transparency by quantifying the contribution of each feature. By prioritizing actionable behavioral signals over immutable attributes, dynamic calibration aligns pricing with individual risk while minimizing reliance on historical, possibly biased factors, fostering a potentially more equitable underwriting process.

In the proposed framework, the third dimension is about product design integration, which plays a pivotal role in translating the efforts to reduce potential bias from data reconstruction and model innovation into tangible, fair insurance products.

Hybrid Pricing Models

Hybrid pricing models offer a strategic approach to balance fairness and risk-based pricing. These models are crafted by integrating bias-adjusted base premiums with adaptable discount mechanisms. The base premium is first computed using a bias-free model, such as one trained on synthetic data or incorporating fairness-constrained algorithms. Once the base premium is determined, discounts can be introduced based on an individual's proactive engagement in risk-reducing activities. For instance, participation in health management programs, which may include regular exercise, preventive health checkups, or smoking cessation initiatives, can lead to premium discounts. By rewarding positive behaviors, hybrid pricing models not only encourage policyholders to take better care of their health but also ensure that premiums are more closely aligned with an individual's actual risk, rather than being influenced by potential biases that may be embedded in historical data.

Transparency Mechanisms

Transparency mechanisms are essential for building trust between insurers and customers. Transparency would be increased if insurers make a concerted effort to disclose the weights assigned to different socioeconomic factors in the premium calculation. This could involve providing a detailed breakdown of how factors like education level, geographical location, or income contribute to the final premium. Additionally, interactive rate simulators can be developed to allow customers to input their own data, such as lifestyle choices, health conditions, and demographic information, and instantly see how these factors impact their premiums. This hands-on approach empowers customers, as they can gain a deeper understanding of the pricing process and make more informed decisions about their insurance coverage.

Copyright © 2025 Society of Actuaries Research Institute

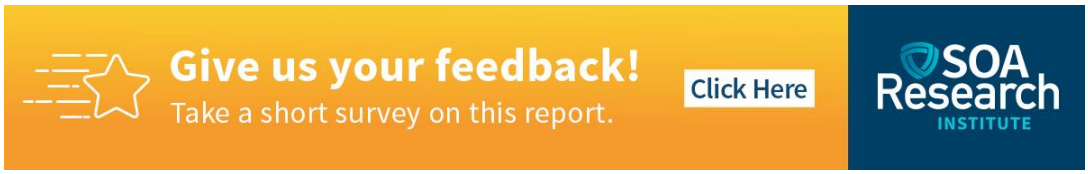
Moreover, transparency helps to hold insurers accountable and ensure that the pricing is based on objective and fair criteria.

The three dimensions of the framework operate synergistically in a loop to ensure holistic bias mitigation. Data Reconstruction serves as the foundational layer, rectifying potential historical biases through dynamic adjustment factors and synthetic data generation to provide unbiased, representative datasets. This cleaned data then becomes the input for Model Innovation, where fairness-constrained algorithms such as Fairlearn and dynamic risk calibration techniques like RNNs with SHAP explainability train models that avoid reinforcing potential historical inequities while maintaining predictive accuracy. Finally, Product Design Integration translates the outputs of these fair models into tangible solutions—such as hybrid pricing models combining bias-adjusted bases with behavior-based discounts and transparency tools like rate simulators—that operationalize fairness for customers. This interplay creates a feedback loop: unbiased data enhances model fairness, fair models inform ethical product design, and transparent products build consumer trust, collectively upholding actuarial integrity while addressing regulatory, ethical, and market demands for sustainability.

CONCLUSION

Addressing potential historical data biases in life and annuity insurance pricing is essential for the industry's ethical and sustainable development. The proposed three-dimensional framework offers a comprehensive solution that combines technical, regulatory, and customer-centric approaches. By implementing these strategies, insurers can not only reduce potential historical data-based biases but also enhance their reputation and tap into new market segments.

* * * * *



REFERENCES

[1] Xin, X., Huang, F. “Anti-Discrimination Insurance Pricing: Regulations, Fairness Criteria, and Models.” *SSRN Electronic Journal*, 2021.

[2] Cerchiara, Rocco Roberto, Watson,et al. “Generalized linear models in life insurance: decrements and risk factor analysis under Solvency II.” 2008.

[3] Haberman, S, Renshaw, A. E. “Generalized Linear Models and Actuarial Science.” *Journal of the Royal Statistical Society Series*, 1996.

[4] Barry, L., Charpentier, A. “Melting contestation: insurance fairness and machine learning.” *Ethics and Information Technology*, 2023.

[5] Lux, G., Theresa Hüer, Buchner, F., et al. “Implementation of socio-economic variables in risk adjustment systems: A quantitative analysis using the example of Germany.” *Health Policy*, 2025.

[6] Abroshan, M., Elliott, A. “Counterfactual Fairness in Synthetic Data Generation”. 2022.

Copyright © 2025 Society of Actuaries Research Institute

[7] Pandey, H. “Comparison of the usage of Fairness Toolkits amongst practitioners: AIF360 and Fairlearn.” *TUDelft*, 2022.

[8] Dattangire, R., Veerakumar, S., Funde, N., et al. “A New Hybrid RNN-LSTM Model for Predicting the Health Insurance Prices.” 2024.

[9] Bhongade A, Dubey Y, Palsodkar P, et al. “Explainable AI Model for Medical Insurance Premium Prediction Using Ensemble Learning with SHAP Analysis.” 2024.

Copyright © 2025 Society of Actuaries Research Institute

Award Winner

Beyond Redlining: Addressing Potential Zip Code Bias in Life Insurance Pricing

Joshua Owusu

October 2025

The views and ideas expressed in this essay are the author's alone and do not represent the views or ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, or the author's employer.

INTRODUCTION

Insurance models, whether for pricing, risk assessment, or customer engagement, commonly rely on variables such as age, gender, and location. While these factors can be useful in making predictions about outcomes like mortality and morbidity, they may not be free from ethical concerns. In particular, geographic rating variables like ZIP codes, though appearing neutral, can reflect socioeconomic and racial disparities due to historical redlining.[1]

Although redlining was outlawed in 1968 by the Fair Housing Act, its legacy could continue to influence the datasets used in insurance, potentially introducing biases that impact decision-making.[2]

This essay discusses how ZIP codes may introduce bias in insurance datasets, focusing on their indirect influence on data used for life and annuity products. It then explores adversarial debiasing as a promising machine learning approach to mitigate these effects.

UNDERSTANDING ZIP CODE BIAS

The history of ZIP code bias dates back to the 1930s, when the Home Owners' Loan Corporation created maps that labeled Black-majority neighborhoods as "hazardous" for investment.[2] Even when race may not be directly used, today's models may still learn biased patterns from historical data. By linking certain locations to higher risk, these models can unintentionally repeat past discrimination.

For life and annuity insurance products, the connection between ZIP code and bias in datasets, while often indirect, is still significant. Studies have found that redlined areas have comparatively fewer healthcare

[1] Redlining refers to systematic denial of services (for example, loans or mortgages) based on location without considering the qualifications of the individual applicant.

[2] Aaronson, Hartley, and Mazumder, (2021 November), "The Effects of the 1930s HOLC 'Redlining' Maps," American Economic Journal: Economic Policy 13 (4): 355–92. https://doi.org/10.1257/pol.20190414.

Copyright © 2025 Society of Actuaries Research Institute

facilities and higher mortality rates, following decades of disinvestment.[3, 4] Consequently, an algorithm trained on datasets from such areas could attribute higher risk to all individuals from those locations. Moreover, reliance on ZIP codes or their proxies can lead to missed opportunities for insurers by inaccurately tagging valuable clients from certain locations as high risk.

However, just like the age variable, (where a healthy 40-year-old might be lower risk than an unhealthy 30-year-old), ZIP codes can be misleading indicators of personal risk, as they may reflect systemic disadvantages more than individual health or behavior.

ADVERSARIAL DEBIASING: AN ALGORITHMIC ETHICS COACH

Adversarial debiasing offers an approach to mitigating bias within datasets. The process begins by training a primary model to predict a target outcome (Y) using inputs (X_i) while simultaneously training an adversarial model to predict a sensitive variable (Z) like ZIP code from the primary model's output.[5] Through multiple training cycles, the system penalizes the primary model whenever the adversary successfully predicts Z, gradually forcing it to develop fair representations.

The optimization process balances the two competing objectives through a total loss function $L = L_y - \alpha L_d$, where L_y represents the primary loss and L_d represents the adversary's loss[6] (α is a hyper parameter that controls the trade-off between the two objectives). A higher α prioritizes fairness and tries to prevent the adversary from predicting the sensitive variable correctly. A lower α favors prediction accuracy even if some bias remains.

Optimal hyper parameters (such as the alpha value in this case) can be selected by testing a range of values using cross-validation.[7] For each alpha, the model is trained and evaluated using performance metrics (like accuracy or Root Mean Squared Error). The final alpha is carefully selected to ensure that the model has the right balance between the insurer's need for accuracy and fairness across different demographic groups.

This technique has been applied to reduce inequality in several non-insurance contexts. For example, it helped address bias in COMPAS, a tool used in U.S. courts to predict reoffending, which often labeled Black defendants as high risk.[6] In another study using COVID-19 diagnosis predictions, adversarial debiasing helped reduce unfair differences in hospital data across different locations.[5] These cases demonstrate its potential to prevent variables like ZIP codes from acting as proxies for race in insurance datasets.

---

[3] Lynch et al., (2021 June) "The Legacy of Structural Racism: Associations between Historic Redlining, Current Mortgage Lending, and Health," *SSM – Population Health*, Vol. 14, https://doi.org/10.1016/j.ssmph.2021.100793.

[4] Krieger et al.,(2020 July) "Structural Racism, Historical Redlining, and Risk of Preterm Birth in New York City, 2013-2017," *American Journal of Public Health* 110, 1046–1053, https://doi.org/10.2105/AJPH.2020.305656.

[5] Yang et al., (2023) "An Adversarial Training Framework for Mitigating Algorithmic Biases in Clinical Machine Learning," *npj Digital Medicine* 6, 55, https://doi.org/10.1038/s41746-023-00805-y.

[6] Wadsworth, Vera, and Piech, (2018) "Achieving Fairness through Adversarial Learning: An Application to Recidivism Prediction," arXiv:1807.00199, https://doi.org/10.48550/arXiv.1807.00199.

[7] James et al., (2013), "An Introduction to Statistical Learning: with Applications in R," Springer, DOI 10.1007/978-1-4614-7138-7, https://www.statlearning.com/.

Copyright © 2025 Society of Actuaries Research Institute

Other debiasing methods, such as reweighting and preprocessing, aim to tackle bias by adjusting the dataset prior to training. However, they can sometimes fall short of reducing the influence of proxy variables in the data.[8]

By applying adversarial debiasing, insurers can create models that maintain actuarial integrity while reducing unfair geographic biases.

THE CASE FOR ADVERSARIAL DEBIASING

REGULATORY COMPLIANCE

Concerns over potential discrimination have triggered discussions about stricter oversight in the United States, similar to measures adopted in Europe.[9] While it does not explicitly mention ZIP codes, the new EU AI Act emphasizes the need for unbiased data in AI applications. Article 10 requires that datasets used for training, validation, and testing be representative, and examined for biases that could lead to discrimination.[10]

Adversarial debiasing offers a proactive solution, helping ensure that insurance datasets and the models built upon them align with emerging fairness regulations.

REPUTATION MANAGEMENT

Insurers can demonstrate their commitment to equitable practices, increasing consumer trust and improving brand image. In an era of public awareness around algorithmic bias, consumers are increasingly drawn to companies that prioritize ethical decision-making[11]. By adopting adversarial debiasing, insurers show a commitment to fair data practices. This sets them apart from competitors that are slower to make such changes.

MODEL TRANSPARENCY

The process of adversarial training can make it easier to audit and interpret the role of various variables, including proxies like ZIP codes. By explicitly identifying and minimizing the model's reliance on them during training, adversarial debiasing can help provide a structured approach for uncovering hidden sources of bias. This clarity supports regulatory compliance efforts and facilitates internal model validation, enabling actuaries and data scientists to better justify geographic differences in datasets.

THE LIMITS OF ADVERSARIAL DEBIASING

Adversarial debiasing has some drawbacks. First, it requires more computing power and training time because the system must balance two competing goals: accuracy and fairness. Additionally, designing an effective adversary and selecting hyper parameters require careful consideration, as overly aggressive

---

[8] Wongvorachan et al., (2024) "A Comparison of Bias Mitigation Techniques for Educational Classification Tasks Using Supervised Machine Learning," *MDPI*, 10.3390/info15060326.
[9] Frees and Huang, (2021) "The Discriminating (Pricing) Actuary,"SSRN, http://dx.doi.org/10.2139/ssrn.3592475
[10] European Parliament & Council of the European Union, (2024 June 13), *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (the "Artificial Intelligence Act"). Official Journal of the European Union, L 1689*, 12 July 2024, https://data.europa.eu/eli/reg/2024/1689/oj.
[11] Sepideh Ebrahimi et al., "Reducing the Incidence of Biased Algorithmic Decisions through Feature Importance Transparency: An Empirical Study," *European Journal of Information Systems* 34, no. 4 (2025): 636–64, https://doi.org/10.1080/0960085X.2024.2395531.

Copyright © 2025 Society of Actuaries Research Institute

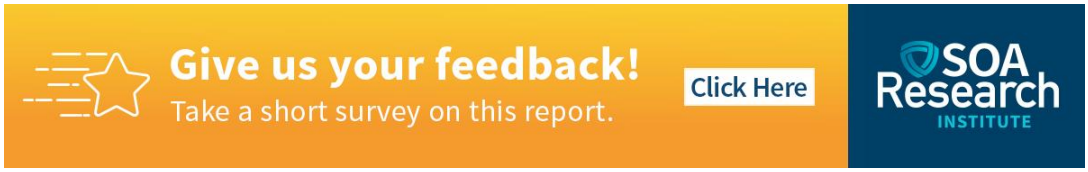
models can remove useful signals in geographic trends. This can reduce the overall performance and accuracy of the model.[12] Third, the method works best with large, balanced datasets, which smaller insurers may not have.

Despite these obstacles, adversarial debiasing remains one of the best ways to increase fairness in insurance datasets without compromising actuarial rigor.

CONCLUSION

Adversarial debiasing offers a forward-looking solution to address potential unfairness within insurance datasets. Since using ZIP codes can reflect past discrimination, insurers need better tools that are both accurate and fair. This method helps reduce potential ZIP code bias by teaching models to learn without relying on sensitive geographic information.

* * * * *



[12] Members of the CAS Race and Insurance Pricing Research Task Force (2025) *Practical Application of Bias Measurement and Mitigation Techniques in Insurance Pricing: Part 2 - Advanced Fairness Tests, Bias Mitigation, and Non-Modeling Considerations*, Casualty Actuarial Society, https://www.casact.org/sites/default/files/2025-01/Practical_Application_of_Bias_Measurement_Part_2.pdf.

Copyright © 2025 Society of Actuaries Research Institute

Award Winner

Equity Underwritten: Mitigating Bias in Risk Assessment and Pricing Processes

Marco Pirra, AFFI, CAS October 2025

The views and ideas expressed in this essay are the author's alone and do not represent the views or ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, or the author's employer.

INTRODUCTION

Fairness in insurance has always been a critical concern. Insurance inherently engages with inequalities, aiming to compensate for unpredictable financial losses while distributing risk across populations. However, some argue that insurance systems have at times contributed to social disparities through practices such as redlining or gender-based pricing.[1] . These practices highlight the need to re-examine fairness not only as a regulatory or ethical obligation but as a foundational principle for trust and access to financial services.

The evolution of insurance has introduced increasingly complex methods for evaluating and pricing risk, particularly through data-driven systems and artificial intelligence (AI). While these tools offer precision and efficiency, they also pose new challenges: statistical and algorithmic approaches introduce their own fairness frameworks, which may be disconnected from historical and legal understandings of equity. As insurers shift toward machine learning-based underwriting and pricing, it becomes essential to interrogate how *bias*—defined here as systematic unfairness in outcomes across social groups—may arise, how it can be measured, and how it might be mitigated. Similarly, this essay approaches *fairness* as a pluralistic concept, encompassing group parity, individual equity, and procedural transparency, depending on the context.

This essay focuses specifically on underwriting and pricing in life insurance, exploring how bias manifests in AI-driven systems and what solutions are emerging at the intersection of actuarial science, statistics, and computer science. It draws on conceptual distinctions between intended and unintended bias, discusses practical fairness techniques, and places these in the context of broader legal, societal, and regulatory developments. These broader regulatory and societal forces help explain why fairness in underwriting has

[1] Mosley, Roosevelt, and Radost Wenman, "Methods for quantifying discriminatory effects on protected classes in insurance," *CAS research paper series on race and insurance pricing* 26 (2021), https://www.casact.org/sites/default/files/2022-03/Research-Paper_Methods-for-Quantifying-Discriminatory-Effects.pdf; and Squires, Gregory D., "Racial profiling, insurance style: Insurance redlining and the uneven development of metropolitan areas," *Journal of Urban Affairs* 25.4 (2003): 391-410, Print.

Copyright © 2025 Society of Actuaries Research Institute

become a central concern and provide essential context for the technical and organizational strategies discussed in the sections that follow.

MOTIVATIONS FOR FAIRNESS: TRANSPARENCY, REGULATION, AND ACCOUNTABILITY

The push to mitigate bias in underwriting is not only an ethical initiative—it is also being driven by growing demands for transparency, legal accountability, and regulatory compliance. As insurers adopt algorithmic tools to evaluate risk and set prices, regulators have begun to intervene. Under the EU's proposed AI Act, insurance underwriting and pricing fall under the "high-risk" category, requiring explainable systems, human oversight, and formal documentation of fairness practices. U.S. state regulators and consumer protection advocates are also intensifying scrutiny of algorithmic decision-making in insurance.

These developments are transforming fairness from an ethical aspiration into a compliance requirement. Insurers are expected to demonstrate proactive assessment of discrimination risks, not just statistical accuracy. Internal governance mechanisms such as fairness review boards and bias audit systems are emerging as essential tools for accountability.

Importantly, transparency also matters to policyholders. Consumers increasingly expect to understand the factors behind their premiums and to have access to clear dispute processes. Addressing these expectations reinforces trust in insurance institutions and helps align the industry with evolving public values.

Fairness, therefore, must be seen not just as a mathematical property but as a legal, social, and reputational mandate. Its definition varies across historical, cultural, and legal contexts. As a result, efforts to ensure fairness must be interdisciplinary, involving actuarial science, data ethics, law, and public policy.

BIAS IN UNDERWRITING: INTENDED AND UNINTENDED

Bias can enter underwriting systems in multiple ways. Intended bias occurs when known disparities in data representation or feature selection are consciously accepted, typically for predictive performance. For instance, if a model is trained predominantly on male applicants and uses gender in rating, the resulting premium recommendations may disadvantage women—even if their risk is equal or lower. While technically rational, such design choices can be ethically and legally problematic, potentially constituting discrimination.

Unintended bias, by contrast, can arise from proxy variables or unrepresentative data. A model trained on urban policyholder data may fail to generalize to rural populations, leading to poor predictions and irrelevant recommendations. Similarly, socioeconomic variables like credit score or ZIP code, while predictive, often correlate with race and income, which may embed systemic inequities into pricing.

Understanding these forms of bias is key. Fairness cannot be reduced to excluding sensitive attributes from models. Indeed, the literature distinguishes between "fairness through unawareness," where protected variables are omitted, and "fairness through awareness," where these variables are included explicitly to monitor and mitigate disparities. The latter approach enables more transparent and equitable model behavior.

Risk classification is foundational to underwriting. Insurers categorize applicants based on health status, lifestyle, and other factors to assign premiums. However, classifications such as BMI thresholds or geographic location can disproportionately affect certain groups.

Copyright © 2025 Society of Actuaries Research Institute

For example, BMI cutoffs may not account for ethnic variations in body composition. Similarly, while ZIP code is more commonly used in property insurance, it has been incorporated indirectly in some life insurance contexts—such as through credit-based scoring or third-party data enrichment—and may serve as a proxy for racial segregation or environmental inequality. These issues call for regular audits of classification systems, testing for disparate impact across demographics and allowing flexibility for individual improvements, such as wellness participation or lifestyle changes.

A pluralistic view of fairness is needed here: different conceptions of fairness (group parity, individual justice, causal attribution) may conflict, and no single metric captures equity in all contexts. This complexity should not deter action, but it requires transparency in choosing and justifying fairness criteria.

FEATURE SELECTION AND PROXY DISCRIMINATION

One major source of bias lies in feature selection. Variables like education level, employment type, or housing status may serve as stand-ins for sensitive attributes, leading to indirect discrimination. Such proxy discrimination may not violate formal model constraints but can still result in unfair outcomes.

To identify such effects, insurers can use *sensitivity analyses* and *counterfactual testing*, in which protected attributes—such as ZIP code or gender—are altered in synthetic test cases to observe if the model's outputs change significantly. While this process involves hypothetical modifications, it does not necessarily introduce personal bias if implemented in a controlled and systematic way. Rather than evaluating real individuals, these methods use matched or simulated records to assess how much sensitive attributes alone influence predictions.

Admittedly, counterfactual fairness testing relies on assumptions about which variables can be changed independently and what constitutes a "fair world." These assumptions are not free of normative judgment. However, they provide a practical lens to uncover structural dependencies in models. Additional tools like adversarial debiasing—where an auxiliary model tries to predict protected attributes from the main model's output—can further expose hidden correlations, offering a more data-driven and less subjective means of identifying bias.

The analytical tool Shapley Additive Explanations (SHAP) values and causal graphs enable users to visualize how the model functions during prediction tasks. These tools do not eliminate bias but enable stakeholders to understand and challenge model decisions. The most promising approach involves using causal inference frameworks because these methods determine whether observed relationships between variables and outcomes represent real effects or random associations.

MODELING TECHNIQUES AND TECHNICAL SOLUTIONS

Machine learning introduces powerful tools for bias mitigation. The counterfactual fairness framework tests hypothetical situations by evaluating how an applicant would be treated if their sensitive attribute were modified. The model shows unfairness when it produces different results for applicants based on their sensitive characteristics. The method of adversarial debiasing requires repeated model modifications to identify and eliminate bias in the predictions.

Explainable AI (XAI) introduces transparency as a fundamental enhancement to artificial intelligence systems. Through feature decision explanations XAI enables developers along with auditors to detect biased patterns and make necessary adjustments for bias remediation. SHAP values provide specific measurements of sensitive proxy variable impact on model predictions thus aiding model improvement as well as compliance verification.

Copyright © 2025 Society of Actuaries Research Institute

Data preprocessing techniques provide insurers with a proactive way to decrease algorithmic bias that may surface during modeling. The process includes de-biasing, cleaning, class distribution balancing, and dataset reweighting to prevent systematic underrepresentation of any demographic group. A high-quality, diverse training dataset is essential because models developed with narrow or unbalanced data tend to replicate any embedded social inequalities. However, how can one know if a dataset meets this standard? Several diagnostics can help: distributional analysis across protected attributes (such as age, race, gender, income), missing data rates by subgroup, and coverage comparisons with population-level statistics (e.g., census or public health data). These assessments can identify whether some groups are over- or underrepresented, or if key variables are biased in how they are recorded or collected. Additionally, fairness-aware data audits can flag latent disparities in data that may not be immediately visible. Though no dataset is perfect, transparent evaluation of representativeness is a crucial first step toward fair modeling.

When fairness constraints are directly incorporated into model training processes, they help algorithms produce balanced outcomes. The constraints function as restrictions that discourage discriminatory patterns while promoting equitable prediction distributions between subgroups. Loss functions should integrate these constraints to enable simultaneous optimization of performance and fairness.

Periodic bias audits are essential. Model performance evaluation examines different demographic groups to measure output variations, including approval rates, false positives, and pricing differentials. When model parameters reveal discrepancies through bias identification, the model parameters should be modified and feature weights adjusted to minimize inequities.

Multiple methods stacked together starting from data collection through preprocessing and modeling constraints and ending with auditing and explainability create a resilient framework for AI-driven underwriting which eliminates bias as much as possible.

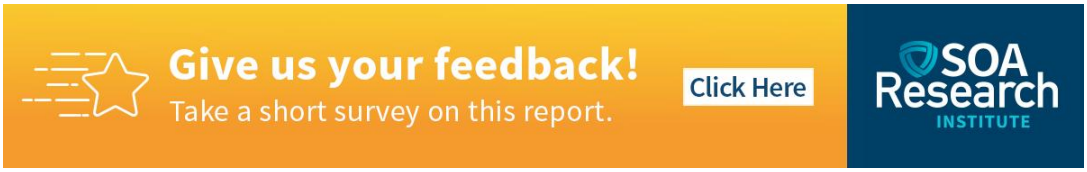
CONCLUSION AND FUTURE DIRECTIONS

The potential for underwriting bias has existed since the inception of underwriting practices, but machine learning technologies have reshaped both its nature and extent. Technical solutions present possibilities, but they are not sufficient on their own. The concept of fairness needs implementation across data sources, models, organizational structures, and how society views insurance products.

Underwriting's future development requires multiple disciplines to work together. The definition of fairness criteria needs collaboration between actuarial science, statistics, computer science, law, and social science to create both technically valid and socially acceptable standards. Insurers need to work with regulators and the public through transparent engagement to embed fairness as a concrete objective within their organizational mission.

Such measures will not only ensure compliance and reduce reputational risk but also reaffirm the role of insurance as a tool for solidarity and protection, one designed to unite rather than divide.

* * * * *



Copyright © 2025 Society of Actuaries Research Institute

REFERENCES

Charpentier, A. (2024). “Insurance, biases, discrimination and fairness.” Berlin: *Springer*. https://www.springerprofessional.de/en/insurance-biases-discrimination-and-fairness/27085462

De Giovanni, Domenico, Marco Pirra, and Fabio Viviano. (2025). “Joint mortality models based on subordinated linear hypercubes.” *ASTIN Bulletin: The Journal of the IAA*: 1-20. https://doi.org/10.1017/asb.2025.8

Fahrenwaldt, Matthias, et al. (2024). “Fairness: plurality, causality, and insurability.” *European Actuarial Journal* 14.2: 317-328. https://link.springer.com/article/10.1007/s13385-024-00387-3

Leong, Jessica, Richard Moncher, and Kate Jordan. (2024). “A practical guide to navigating fairness in insurance pricing.” CAS RESEARCH PAPER. https://www.casact.org/sites/default/files/2024-08/A_Practical_Guide_to_Navigating_Fairness_in_Insurance_Pricing.pdf

Giudici, P., Pirra M., Zieni R. (2025). “Cyber Risk Management with time varying artificial intelligence models.” Accepted XAI-2025: The 3rd World Conference on eXplainable Artificial Intelligence https://xaiworldconference.com/2025/

Copyright © 2025 Society of Actuaries Research Institute

Award Winner

Reimagining Underwriting in Life and Annuity Insurance

Niranjan Rajendran, B.Sc. (Hons) October 2025

The views and ideas expressed in this essay are the author's alone and do not represent the views or ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, or the author's employer.

ABSTRACT

While life and annuity insurance sectors have embraced technological modernization, underwriting practices often remain grounded in legacy systems that reflect potential historical and systemic bias. In this context, social equity refers to the fair and just distribution of resources, protections, and opportunities, ensuring that all individuals, regardless of background, have access to insurance benefits without unfair barriers. This essay explores how bias defined as unjust or prejudicial treatment stemming from data, systemic structures, or algorithms may manifest in underwriting and proposes strategies for moving from exclusionary models to inclusive ones. This essay suggests that underwriting can evolve into a tool for improving social equity and fostering public trust by integrating diverse data sources, transparent governance of artificial intelligence (AI) applications, and participatory design.

INTRODUCTION

Insurance is designed to provide financial protection against unforeseen risks. However, if the mechanisms for assessing that risk, especially underwriting, reflect systemic bias, the promise of protection is unequally fulfilled. In this essay, bias is defined as the unjust distortion of decision-making outcomes caused by historical, structural, or algorithmic influences. In this essay, terms such as equity and fairness are used to distinguish between principles of justice: equity addresses structural differences and unequal starting points, while fairness concerns consistent and impartial treatment. While these concepts are interrelated, they are not interchangeable in all contexts.

Life and annuity insurance policies, meant to deliver financial security, may reinforce inequities when using outdated frameworks. To provide inclusive coverage in today's complex social and economic environments, underwriting models must be reimagined. This essay discusses how industry actors could adopt strategies to align underwriting practices with the ethical and operational goals of fairness, innovation, and accountability.

UNDERSTANDING THE ROOTS OF BIAS IN UNDERWRITING

Bias in underwriting does not necessarily result from malicious intent. It may emerge from the ways systems are constructed and the data they rely on. Key contributors include:

1. Historical Data Inequities: Traditional underwriting relies on historical data to estimate future risk. However, these datasets may be embedded with past inequities. For example, communities

Copyright © 2025 Society of Actuaries Research Institute

historically denied access to healthcare, education, or stable employment may appear as higher-risk due to factors unrelated to actual individual behavior. Without adjustments, these models may inadvertently perpetuate systemic disadvantages. (Solon Barocas, 2019).

2. Algorithmic Bias and Automation Risks: As insurers turn to automated underwriting systems, machine learning models trained on biased data replicate and reinforce those biases. For example, while "ZIP code of birth" is not commonly used, location-based variables like current residential ZIP codes can serve as proxies for race or income, leading to discriminatory patterns. These models, left unchecked, may continue to produce exclusionary outcomes on a larger scale. (Prince, 2020).

3. Proxy Variables and Redlining: Neutral-seeming inputs like education level or employment history may correlate with demographic factors such as race, gender, or socioeconomic status. These inputs can unintentionally result in digital redlining, a term used to describe algorithmic exclusion of disadvantaged populations, even when explicit demographic data is not used. Identifying and recalibrating these inputs is essential for promoting equitable outcomes. (Binns, 2018)

4. Opacity and Lack of Accountability: Many underwriting systems provide limited transparency. Applicants often receive vague denials or high premiums without clear explanations. This lack of clarity prevents consumers from understanding decisions or contesting unfair outcomes, thereby reducing accountability and trust in the process.

Addressing bias in underwriting is both an ethical and strategic imperative. From my perspective on a moral standpoint, insurance should serve as a social safety net, not a barrier. Excluding vulnerable populations not only contradicts this mission but may also erode trust in the industry.

From a business perspective, inclusion can be a source of competitive advantage. A 2021 McKinsey & Company report, "Diversity Wins: How Inclusion Matters," found that organizations prioritizing diversity and inclusion were more innovative and financially successful than their peers. Inclusive underwriting could open access to millions of underserved customers, expanding the market while enhancing an insurer's reputation for social responsibility. (Diversity Wins: How Inclusion Matters., 2021).

STRATEGIES FOR EQUITABLE UNDERWRITING

The following strategies are proposed as a framework to build underwriting systems that support fairness while addressing real-world market constraints:

1. Incorporating Expansive Data
   Supplement socioeconomic datasets with data that better mirrors the realities of diverse communities. These supplemental datasets include:
   - Alternative credit datasets (e.g., utility and rent payment histories)
   - Community health indicators
   - Non-linear employment records (freelance and gig work)

   Collaborations with public agencies and community-based organizations could help insurers gather more representative and context-rich data. (Raji, 2019).

2. Algorithmic Fairness Audits
   As a part of continuous governance, periodically engage independent third parties to audit AI models for potential algorithmic bias, including:

Copyright © 2025 Society of Actuaries Research Institute

- Assessing differential treatment of various demographic groups using group-specific metrics
- Incorporating Explainable Artificial Intelligence (XAI) to illuminate the pathways that lead to particular decisions

Employing fairness measures to ascertain and mitigate bias, such as equal opportunity or disparate impact ratios. This approach emphasizes continuous improvement over reactive compliance. (Binns, 2018).

3. Ongoing Bias Training for Underwriting Teams
   Despite increasing automation, human underwriters still influence key decisions. Regular training on implicit bias, cultural awareness, and inclusive judgment can help underwriting teams make more thoughtful and equitable assessments. This training should be viewed as an evolving process rather than a static obligation.

4. Transparent, Applicant-Centric Communication
   Building consumer trust requires transparency. Insurers might consider:

   - Communicating underwriting criteria
   - Providing detailed explanations for application decisions
   - Offering applicants an opportunity to appeal or supply additional information

   Such transparency empowers applicants to understand, challenge, and learn from underwriting decisions.

5. Feedback Loops and Participatory Design
   A robust feedback system could help insurers identify and address unintended consequences of their policies. This may include:

   - Collecting and analyzing applicant experiences and concerns
   - Involving diverse stakeholders, including community representatives, in the underwriting model design

   This participatory approach helps ensure that the systems reflect the needs and values of the populations they aim to serve.

A VISION FOR THE FUTURE: EQUITABLE UNDERWRITING IN ACTION

A more inclusive underwriting framework might consider:

- Financial behavior over traditional employment status
- Current health outcomes instead of generalized mortality tables
- Decision letters that are explanatory and educational, rather than opaque and discouraging

This vision is not speculative. With responsible governance and a commitment to equity, insurers can evolve underwriting into a tool that supports financial inclusion rather than exclusion.

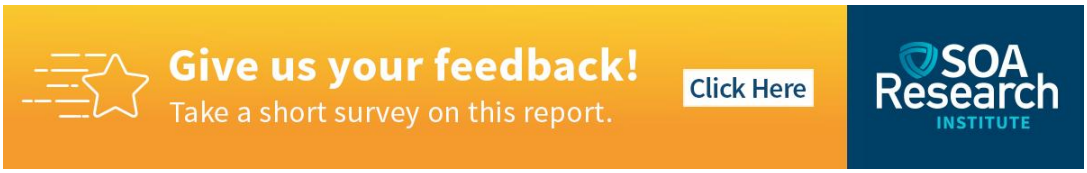
Copyright © 2025 Society of Actuaries Research Institute

CONCLUSION

Potential bias in underwriting is shaped by structural, data-driven, and algorithmic factors that can restrict access to insurance coverage for many individuals. Addressing this challenge involves the application of existing tools and approaches. These include the use of more representative and contextually appropriate data, the implementation of fairness audits for AI systems, the provision of ongoing training for human decision-makers, increased transparency in communications with applicants, and greater involvement of diverse communities in system design.

As the insurance industry continues to evolve, it faces a clear opportunity to reassess long-standing practices. By moving toward underwriting models that prioritize inclusion, fairness, and accountability, insurers may enhance both the effectiveness and the social value of their services. The decision to modernize underwriting in line with contemporary ethical and technological standards may play a significant role in shaping the future of equitable risk protection.

* * * * *



REFERENCES

Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy.

Company, M. &. (2021). Diversity Wins: How Inclusion Matters.

Prince, A. E. (2020). Proxy Discrimination in the Age of Artificial Intelligence and Big Data.

Raji, I. D. (2019). "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products."

Solon Barocas, M. H. (2019). "Fairness and Machine Learning."

Copyright © 2025 Society of Actuaries Research Institute

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
8770 W Bryn Mawr Ave, Suite 1000
Chicago, IL 60631
www.SOA.org

Copyright © 2025 Society of Actuaries Research Institute