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, 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:

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

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 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 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:

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

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. 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.

[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.