NAMIC® ADVOCACY

BIG DATA, ARTIFICIAL INTELLIGENCE, AND RISK-BASED PRICING: DISPELLING FIVE COMMON MYTHS

LINDSEY KLARKOWSKI
Policy Vice President, Data Science, AI/ML, and Cybersecurity
NAMIC

NAMIC®
ISSUE ANALYSIS

LINDSEY KLARKOWSKI

Lindsey Klarkowski, J.D., serves as NAMIC's policy vice president – data science, AI/ML, and cybersecurity. She is responsible for providing NAMIC and its membership with policy expertise and analysis on legislative or regulatory issues related to artificial intelligence, machine learning, and data governance at the state and federal levels.

Prior to joining NAMIC, Klarkowski worked as in-house counsel for a Fortune 500 insurance company, where she provided subject matter expertise and legal counsel on digital commerce, emerging technology, and AI/ML governance. She also worked closely with the company's government affairs team to develop policy strategy on emerging technology issues impacting the insurance industry and analyze related draft legislation and regulation.

Klarkowski earned her law degree from the University of Wisconsin Law School and her bachelor's degree from the University of Minnesota. She is an active member of the Wisconsin State Bar.

For more information about this NAMIC Issue Analysis please visit namic.org/media-publications/publications/public-policy-papers/ or contact:

LINDSEY KLARKOWSKI
lklarkowski@namic.org
317.876.4212

The National Association of Mutual Insurance Companies consists of more than 1,300 member companies, including six of the top 10 property/casualty insurers in the United States. The association supports local and regional mutual insurance companies on main streets across America as well as many of the country's largest national insurers. NAMIC member companies write $383 billion in annual premiums and represent 61 percent of homeowners, 48 percent of automobile, and 25 percent of the business insurance markets.

TABLE OF CONTENTS

NAMIC ISSUE ANALYSIS

INTRODUCTION

A myth is consistently defined as an unfounded or false notion. Myths have captured humanity and been prevalent in societies since the dawn of time – ranging from harmless stories about ancient Greek figures and old wives' tales to the harmful pieces of "fake news" that run rampant across modern social media. Often due to a lack of understanding, or refusal to acknowledge what is factual, myths can be particularly dangerous when they serve as the foundation for or drive the creation of new policy.

Nowhere is there more need to separate fact from fiction than in the insurance policy sphere, specifically relative to the use of big data, artificial intelligence, and risk-based pricing. Particularly prevalent in insurance policy are myths about the effects of insurers using big data and artificial intelligence and unfounded notions about how insurance pricing should operate considering the argued effects. The danger in these myths is that there are growing numbers of advocates and regulators relying on them to advance policy that would ultimately harm the very population they are looking to protect: policyholders.

This paper breaks down big data, artificial intelligence, and risk-based pricing. It also outlines how these concepts interact and dispels dangerous misconceptions about them.

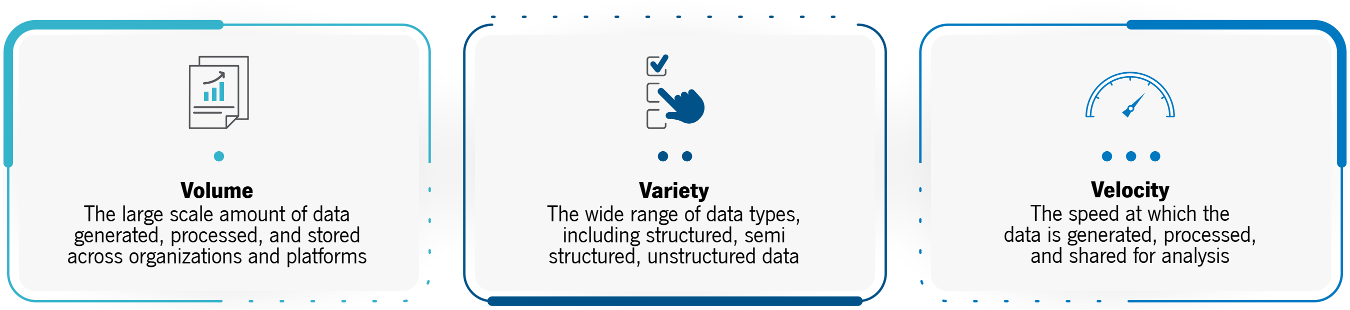
WHAT IS BIG DATA?

Modern technological advancements have brought many changes to society and life as we know it. One particularly relevant to this analysis is the dawn of "big data." In many ways, data collection and analysis have long been a part of our collective history, but the advent of the internet and the accelerating rates at which computers share information have allowed for vast amounts of digital storage and use – in other words, "big" data.

Big data is characterized by its volume - the large amount of the data; its variety - the wide range of data types; and its velocity - the speed at which the data is generated and shared. Storage, processing, and distribution of big data have only improved and grown with computer advancements since the world entered the 21st century.

While the history of big data processing and analysis predates the 2000s, the insurance industry's use of and focus on big data primarily took hold in the 2000s and 2010s. Because the insurance industry is a data-driven and data-focused industry, the availability of big data continues to hold great prospects, particularly in risk-based pricing, which increases precision in underwriting and rating. Not only does expanded availability of data elements - the variety of big data - help increase precision in this regard, applying predictive models or algorithms on top of big data sets can identify patterns to help better inform risk rating and underwriting decisions.

The 3 V's of Big Data:



WHAT IS ARTIFICIAL INTELLIGENCE?

Artificial intelligence (AI) is, surprisingly to many, not a new field of study. The term and concept were coined in the 1950s, but the recent large-scale attention has been due to the rapid advancement of computer processing power and the onset of generative AI. This rapid advancement and widespread availability have sent many racing to understand AI, and its various applications and left society, advocates, lawmakers, and regulators susceptible to believing unfounded claims about the technology.

Generative AI is a specific type of AI model that has garnered much attention in the wake of Chat GPT's public release in November 2022. Using deep learning, generative AI learns patterns in large data sets and then generates new content based on those learnings. The hallmark of generative AI models, like Chat GPT, are those two traits - being trained on vast amounts of data from the internet, and the ability to generate novel content.

In the simplest sense, AI is applied math. It is a field of advanced data analytics and computer science aimed at developing data processing systems that perform functions normally associated with human intelligence - i.e., reasoning, learning, self-improvement.

The field of AI creates systems and models that provide insights into data and are capable of providing or generating new content based on those insights and learnings from large data sets. There are distinct applications and subsets of AI depending on the type of insight or output one is looking for. Machine learning applications, deep learning applications, and generative AI currently are the most prevalent and popular.

No matter the type of model or application, the general process of how AI models work is the same: a model is developed, operates by receiving large data inputs, gets trained on the data input, and generates a data output. Advanced data processing like this is beneficial in the insurance industry in that it brings the prospect of increased risk insights and increased precision in risk-based pricing, which strengthen competition and benefit consumers.

WHAT IS RISK-BASED PRICING?

To understand risk-based pricing, one must first understand the function of insurance. "Put simply, insurance companies buy the risk of specified losses from individuals and businesses, then pay if those losses happen."[1] Because no loss has happened at the time of an insurance policy purchase, insurers must prospectively assess the risk of loss occurring and price the insurance product accordingly based on the potential likelihood and magnitude of loss by the policyholder.

To support this risk-based pricing, insurers use two primary processes: underwriting and ratemaking. Underwriting is the process through which an insurer assesses the risk presented by a policyholder. As part of this process, insurers ask:

1. How likely is it that the policyholder will experience a covered loss;
2. How should the risk be grouped with others like it (i.e., placing risks into groups based on common characteristics to discern expected loss probability); and
3. Do we want to take on this risk?

[1] Cotto, Tony. Why Your Insurance Costs What It Does: A Risk-Based Pricing Primer, p. 2. (2021).

After the insurer has determined the risk of experiencing a loss, placed the risk into a risk classification, and decided to purchase the risk, the insurer undergoes a process known as ratemaking.

Risk-based pricing is a data-driven exercise, which benefits from increased precision through leveraging big data and AI. Because the process of risk-based pricing and the insurance product itself are unique in both legal and consumer contexts, it is highly susceptible to any regulation that restricts an insurer's ability to accurately predict policyholder loss. Policy themes in legislative and regulatory discussions surrounding AI and big data to date, unfortunately, have been based on inconsistent notions of how big data, AI, and insurance operate, resulting in unsubstantiated allegations about negative impacts to the insurance marketplace and policyholders alike.

BIG DATA/AI ADVANCEMENT AND CORRESPONDING POLICY THEMES

Not surprisingly, policy themes and concerns have evolved with the advancement and availability of big data and AI use in the insurance industry. Big data jumped onto the scene first and held everyone's attention in for this century's first couple of decades. With this increased use came increased scrutiny from regulators and advocates. Of particular concern from some on the regulatory and policy side was whether the use of alternative or nontraditional data elements in risk rating and underwriting would lead to proxy discrimination – the idea that a particular data element, while strongly correlated to risk, may also be strongly correlated with a protected class.

Advancements in AI capability and application amplified existing policy attention from regulators and advocates. AI advancements began allowing for greater insights and use of big data, and the combination of big data and AI use in the insurance industry gave rise to a novel and amorphous concern of "algorithmic bias;" and related insurance availability and affordability concerns for vulnerable populations. Largely stemming from a lack of understanding of how the technology functions and is used in risk-based pricing, regulators and advocates have raised concerns over transparency, explainability, and the misconception that, because AI models are complex, their use in risk rating and underwriting must be disadvantaging vulnerable or disadvantaged policyholders.

Not only are these policy themes based on unfounded arguments about the effects of big data and AI use in the context of insurance, but these themes are also detrimental to policyholders. Contrary to what may be perceived as well-intentioned social efforts by regulators, policyholders will be harmed by growing efforts to elevate concepts of "fairness" divorced from actuarial science. Despite facing challenging environments in recent years, insurers continue their efforts to match rate to risk in the most accurate way possible to the benefit of policyholders, and the use of big data and artificial intelligence has been, and can continue to be, beneficial in that regard. When one separates facts from myths in the misconceptions that exist in the insurance policy space (particularly relative to bias and discrimination) around AI and big data, this benefit becomes clear.

MYTH # 1: INSURANCE IS JUST LIKE ANY OTHER INDUSTRY AND CAN BE REGULATED AS SUCH.

Fact: Insurance products are uniquely priced prospectively based on risk and function differently than many other consumer products. Regulation specific to this foundation is therefore necessary.

With amplified policy attention to AI and big data for all industries, insurance is often grouped with other industries in policy conversations and legislative and regulatory proposals, particularly related to algorithmic bias. Because some advocates and policymakers believe algorithmic bias is a universal issue and that every industry must address it, there is often not a distinction for the insurance industry. Nor is there a conversation about how or whether algorithmic bias occurs in the function or pricing of insurance.

Insurance is distinct in function and price from most consumer products.

Compared with products that consumers pick up off the shelf, where price is determined based on known costs, expenses, and profits, an insurance contract "does not and cannot work the same way because the actual cost to the seller is unknown at the time of sale."[2] To price an insurance product, insurers must make educated judgments about a policyholder's expected risk of a covered loss, predicting how likely it is that something will go wrong for the policyholder, thereby requiring the insurer to pay.

Insurers estimate this risk of loss through the processes of underwriting and ratemaking. These underwriting and rate-making processes are tightly regulated by state laws and enforced by state insurance departments. Insurance codes in each state are designed specific to these processes, establish that rates cannot be "inadequate, excessive, or unfairly discriminatory,"[3] and, in some cases, provide parameters for what data elements may be used. For instance, no risk classification may be "based upon race, creed, national origin, or the religion of the insured."[4] The prohibitions mentioned above serve to protect insurance companies and policyholders and are unique to the insurance product. Any law, regulation, or policy proposal that impacts or conflicts with these standards is likely to have a large effect on the way an insurer prices the product, and ultimately, the availability and affordability of the product for consumers. Therefore, any regulation, especially relative to AI and big data use in insurance risk-based pricing, must be unique to the industry and account for the unique nature of the insurance product.

[2] Id., p. 2.
[3] See, National Association of Insurance Commissioners Property and Casualty Model Rating Law (#1780).
[4] National Association of Insurance Commissioners Property and Casualty Model Rating Law (#1780).

MYTH # 2: INCREASING RISK-RATING PRECISION WITH AI AND BIG DATA WILL CREATE A 'RISK POOL OF ONE,' WHICH IS DETRIMENTAL TO CONSUMERS.

Fact: The more accurate risk rating is, the more insurers can take on riskier policyholders, thereby increasing availability.

When determining an offer for insurance, an insurer will examine the applicant's risks of loss and classify those risks to charge proper premium for each classification. The classifying exercise is one in which an insurer's selected risks are pooled with other insureds who pose similar risks. To accurately price insurance products, the insurance industry relies on group estimates of expected loss rather than individual estimates and conducts statistical analyses of those groups to determine average loss probability for each of the group members.[5] Without such group probabilities, "it would be impossible to set a price for insurance coverage at all."[6]

Competition among insurers drives refinement of risk pools, in that a more accurate classification will more closely align premiums with the level of risk an insured presents, thereby driving lower-risk individuals to insurers that have more refined risk classifications and lower corresponding premiums. The advancing availability of data and technology such as AI to derive insights from that data increases an insurer's ability to refine risk classifications. These advancements have led some advocates to embrace the theory of a "risk pool of one" – meaning that advancements in data insights will lead to such precision in risk rating that instead of a pool or classification of risk, insurers will price based on an individual insured, or risk pool of one.

The arguments against increased precision and the concerns about a risk pool of one generally stem from the thinking that increased precision will create an accessibility issue for higher-risk insureds. For instance, if use of big data and AI leads to increased precision in risk rating and pricing, such increased precision could lead to finer segmentation of policyholders into small risk pool groups, resulting in the potential for even smaller, higher-risk pools with resulting high premiums to match. Notwithstanding the fact that this argument ignores what constitutes fair pricing in insurance and incentives for risk reduction, this argument importantly ignores market drivers and competition as beneficial to access and affordability for consumers. Much like other goods, services, and industries, a competitive market for insurance "is the most effective guarantor of low prices [and] widespread availability."[7] Increased precision in risk rating, particularly through leveraging insights from data and advancements in AI, allow for such a competitive market and availability.

Insurer appetite for risk is only possible when insurers are able to grow the pool of potential policyholders, and that is only possible when insurers can spread enough risk. From a market perspective, if an insurer wishes to market its products effectively, it "must utilize a risk classification system that will allow it to offer insurance to as many potential customers as possible."[8]

[5] Detlefsen, Robert, Ph.D. The Case for Underwriting Freedom: How Competitive Risk Analysis Promotes Fairness and Efficiency in Property/Casualty Insurance Markets, p. 4 (2005).
[6] Id., p. 4.
[7] Id., p. 3.
[8] Id., p. 2.

While competition for lower-risk insureds is generally more intense in the marketplace, insurers seeking to improve market penetration and economic advantage will necessarily compete for high-risk insureds.[9] If an insurer can more accurately assess a risk, it can more accurately discern whether it can absorb a higher-risk insured. This fact, coupled with the economic drivers of market penetration, results in increased availability of insurance, even for high-risk individuals.

Lastly, but no less importantly, the theory and argument about a risk pool of one ignores statistical credibility and benefits of risk spreading. While insurers continue to refine the ability to predict risk of loss with advancements in technology and data availability, it remains true that a n individual's previous loss experience is simply not credible enough to statistically warrant an estimate of the likelihood of loss. For insurance to function, it must rely on group estimates of expected loss because "no single individual can truly be said to have an expected loss probability; instead, underwriters use statistical analyses of groups to determine the average loss probability for each group member."[10] Further, a key feature of insurance is risk spreading, which is only accomplished with the help of the law of large numbers[11] and segmenting risk pools as accurately as possible.

The use of big data and AI to increase precision in risk rating will increase availability for consumers through increased competition, market penetration, and the ability to more accurately discern whether a risk can be absorbed.

MYTH # 3: 'FAIRNESS' IS ONE EQUAL PRICE FOR ALL CONSUMERS.

Fact: Socialization of insurance will exacerbate issues of affordability that policymakers are concerned with today and not do anything to solve the underlying problems.

Availability and affordability of property/casualty insurance in certain communities have received considerable policy attention and have been the topic of debates for decades. Certain advocacy groups have suggested that property/casualty insurers engage in unfair discrimination in their pricing and underwriting activities when such pricing and underwriting result in disproportionately negative effects on certain groups of consumers – namely protected classes. In the same vein, consumer groups contend that disproportionately negative effects on certain groups worsen an availability and affordability concern relative to obtaining insurance. Therefore, they often advocate for public policy that results in lesser differentiation in rates and premiums. The principle underlying these arguments is that insurance prices should be the same for all insureds.

While these arguments are no doubt well intentioned, insurance is a risk-priced product and any restriction on an insurer's ability to price a policyholder's risk will result in more, not fewer, access and affordability issues. Fairness in insurance must focus on the effort to match price to risk based on economic, insurance, and actuarial principles.

[9] Id.
[10] Id., p. 4.
[11] As sample size grows, randomness and uncertainty decrease.

Because not all risks are the same, neither are all rates. If policymakers and advocates divorce the concept of matching rate to risk from insurance and ignore the fact that insurance must reflect reality, the result is an inevitable break of the insurance product at its core.

Resultingly, access and affordability issues will persist and could even increase under such an approach. To fix any issues of high premiums for high-risk insureds, one must fix or solve the underlying risk, through mitigation or otherwise. Artificially setting all insurance premiums at an equal price for all insureds does nothing to help the underlying problem causing the differences in risks and resulting rates.

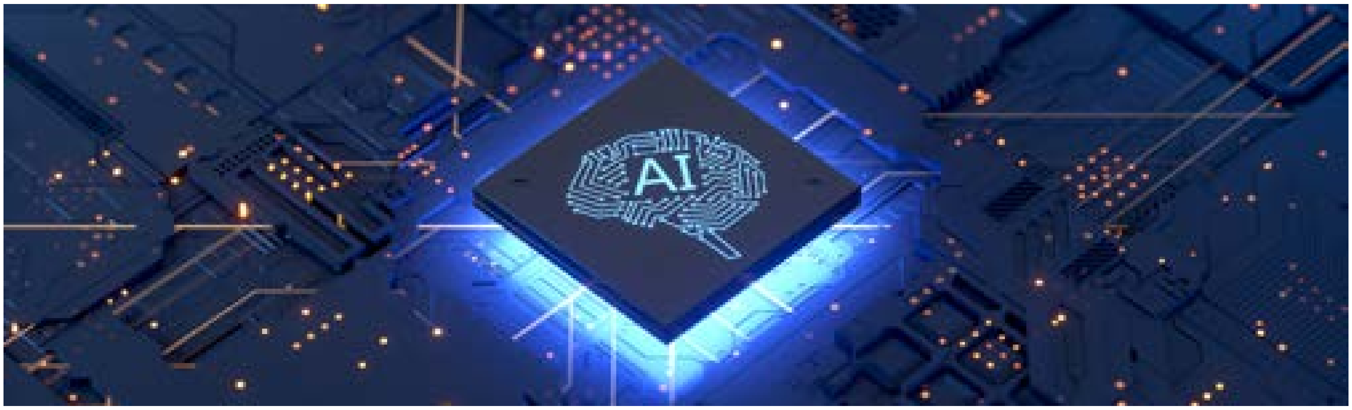
Laws and policies intended to promote social justice typically aim "to achieve greater equality of wealth and income. However, insurance regulation that redistributes risk may have the opposite effect because the level of risk presented by a given individual will not necessarily be correlated with his level of wealth."[12] An unintended consequence of such approach in insurance could be that low-income insureds with low risk will subsidize high-income consumers presenting high risk - think: a middle-class worker driving a Toyota SUV in the suburbs of the Midwest paying the same amount as a billionaire using the self-driving feature on a luxury electronic vehicle on the highways of Los Angeles.

Recognizing the uniqueness of the insurance product and the nature of its pricing, state insurance codes are designed in such a way that "fairness" means a rate or price that is an actuarially sound estimate of future costs based on risk and is not "unfairly discriminatory." Discrimination for insurance purposes has nothing to do with prejudice and everything to do with differentiation among risks. The standard for evaluating the fairness of differential treatment of insurance consumers is statutory, well established, and found in state insurance rating laws, which instruct that "rates shall not be excessive, inadequate, or unfairly discriminatory."[13]

Any discussions involving the industry's use of AI, big data, and predictive analytics should embrace the risk-based history and industry foundation, both of which must be understood fully in the context of what constitutes fairness. To invoke or use a different standard of fairness, such as one that charges all groups similarly, would be catastrophic for policyholders, as it would increase premiums for all policyholders. It forces less-risky consumers to subsidize riskier consumers, which causes market distortions that affect affordability and availability of coverage for consumers, particularly the most vulnerable ones at the center of the fairness discussions being brought forth by some advocates. Thus, the opposite of the advocates' intended affordability and availability occurs, and the underlying problems resulting in concentrated risk with vulnerable populations persist.

[12] Detlefsen, p. 7.

[13] National Association of Insurance Commissioners Property and Casualty Model Rating Law (#1780).



MYTH #4: USE OF AI AND BIG DATA IN RISK-RATING WILL LEAD TO INCREASED BIAS OR DISPARATE IMPACT AND SUCH OUTCOMES SHOULD BE PROHIBITED."

Fact: These concepts are incongruent with the risk-based foundation of insurance, where differential treatment of consumers is based on risk, not on protected class status.

Since the advent of big data and recent advancements in AI, advocates have theorized that use of nontraditional data - newly available data for purposes of risk –rating - and advanced AI models will lead to increased bias and disparate impact in premiums charged. Not only have advocates failed to define what is meant by "bias," the principles of bias and disparate impact are incongruent with the risk-based foundation of insurance.

If one posits that bias means differential treatment resulting in favored or unfavored treatment of a person or group, such differential treatment between consumers is not problematic in insurance regulation, given how the product functions. As explained by the National Association of Insurance Commissioners to the U.S. Supreme Court, "[R]isk selection is the very essence of the business of insurance. In insurance, discrimination is not necessarily a negative term so much as a descriptive one. For insurance, discrimination is not only permitted, but necessary."[14]

To wit, assuming that bias means differential treatment, to prohibit such outcome would be fundamentally incongruent with how the insurance product is priced and how it functions, for insurance discriminates by risk types and charges premiums accordingly based on statistical differences. Discrimination has a different, specific meaning in the business of insurance and its regulatory state than in many other legal contexts, in that insurance "always involves discrimination...based on statistical differences and actuarial tables."[15] State legislatures have recognized this and specifically intended to "only prohibit unfair discrimination in the sale of insurance policies"[16] in state insurance law.

[14] NAIC Amicus Brief to the U.S. Supreme Court in Nationwide v. Cisneros, 1996 WL 33467770.
[15] Thompson v. IDS Life Ins. Co., 274 Or. 649, 654 (1976).
[16] Id., emphasis added.

Separately, if one posits that bias means that the data is skewed or not representative, such concept of bias has not been proven or shown to be an issue in the context of algorithms, data use, or AI for insurance risk-based pricing. Bias, in this sense, results from incomplete or distorted data - for instance, a group that has been historically underrepresented in a particular context. In the context of insurance, however, this concept does not apply in the same way. As this paper has discussed, the data insurers use for risk-based pricing is data that is actuarially sound and correlated with risk and does not include nor use certain protected class attributes. To argue that insurer use of data, algorithms, or AI in risk-based pricing is biased or skewed would be to say that the actuarially sound data is not representative of the risk the policyholder presents, which insurance laws already prohibit. This application of bias in the insurance context is, therefore, nonsensical.

Different from bias, disparate impact is a defined term and legal principle. But, it is a legal principle from the Fair Housing Act that was not developed for use or application in the insurance context. At a high level, disparate impact occurs when there is resulting discrimination without discriminatory intent. If this standard were to apply in the context of insurance, it would prohibit facially neutral risk rating factors that result in disproportionately negative outcomes on certain protected classes. In other words, it would prohibit risk-rating factors that are predictive of risk but result in higher premiums for protected classes. Such application in insurance is antithetical to risk-based pricing. If a disparate impact standard were applied in insurance, eliminating a risk factor found to have a "disparate impact" would result in rates that are likely unfairly discriminatory because rate differences would no longer be based on underlying insurance costs.

Unfair discrimination is the legal standard in insurance, defined as treating similarly situated insureds or similar risks differently through the amount of premium, policy fees, or rates charged for the insurance product,[17] and rates must be actuarially sound based on the policyholder's risk of a claim. No risk classification may be "based upon race, creed, national origin, or the religion of the insured."[18] If an insurer's use of AI or big data in risk rating produces a rating factor that is not predictive of risk or is itself a protected class, then such conduct is not compliant with the existing prohibition on unfair discrimination. Insurers have been diligently applying these objective and understandable rules for many decades and are continuing to do so today. Based on these existing legal standards and the risk-based foundation of the insurance product, a "bad" outcome in insurance occurs when rate diverges from risk regardless of the policyholder's membership in a protected class.

Importing bias and disparate impact principles into insurance law will undoubtedly disrupt markets with negative consequences for the insurance ecosystem and consumers and subject insurer decisions to outcomes testing by regulators, which carries its own issues.

[17] National Association of Insurance Commissioners Unfair Trade Practices Act Model Law (#880-1).
[18] National Association of Insurance Commissioners Property and Casualty Model Rating Law (#1780).

MYTH #5: BIFSG AND MULTIVARIATE ANALYSIS ARE USEFUL TOOLS IN TESTING FOR ALGORITHMIC BIAS.

Fact: Even if an outcomes-testing regime for bias were congruent with insurance and its legal frameworks, there are inherent limitations in reliable demographic data and testing methodologies for such purpose.

Notwithstanding the fact that bias and disparate impact principles are incongruent with the foundation of insurance and existing insurance law, applying such principles runs into another important issue. There are inherent limitations in reliable demographic data for purposes of testing and inherent limitations in reliable testing methodologies for the purpose of identifying any bias or disparate impact.

Property/casualty insurers do not collect or use protected class data - except in circumstances where there is actuarial significance and the traditional underwriting factor has been recognized by law, such as sex in auto insurance risk rating - and insurers are not interested in such collection. Collecting data directly from applicants or insureds creates numerous issues, including credibility of the data, whether the data collection is voluntary and what to do if it is not obtained, how data is stored and preserved, as well as privacy, security, and litigation risk concerns from collecting and storing sensitive data. These issues are present notwithstanding the fact that some states specifically prohibit insurers from collecting protected class data such as race or ethnicity. Any analysis of protected classes mandated by a regulator would require the industry to estimate an applicant's or insured's protected class status, a prospect that creates its own set of concerns and potential liabilities.

Thus far, the only tool suggested by insurance regulators for estimating a protected class has been the Bayesian Improved First Name Surname Geocoding (BIFSG) imputation methodology.[19] The BIFSG methodology "combines estimates of the racial composition of the geographic area in which an individual resides with data on racial and ethnic distributions associated with their first and last names to estimate a set of probabilities that an individual belongs to each of six distinct racial and ethnic groups."[20]

Limitations and Shortcomings of BIFSG use in Insurance:

- Not designed for use in the insurance regulatory context
- Overstates the probability that an individual belongs to the majority group where they live
- Fails to account for multiracial individuals and households

[19] See, DRAFT PROPOSED Algorithm and Predictive Model Quantitative Testing Regulation, Colorado Division of Insurance, https://drive.google.com/file/d/1BMFuRKbh39Q7YckPqrhrCRuWp29vJ44O/view.

[20] Derby, Elana, et al., Statistical Bias in Racial and Ethnic Disparity Estimates Using BIFSG, p. 1 (October 2024).

Studies looking at the utility of BIFSG have found that the "assumptions underlying BIFSG can fail," and result in biased estimates themselves in that "BIFSG suffers from majoritarian bias, overstating the probabilities that non-White individuals are White."[21] To illustrate this, if an individual lives in an area where "the majority of residents are of a different race/ethnicity than the individual, and whose last name is not strongly associated with a certain race/ethnicity, BIFSG assigns a relatively high probability of that individual belonging to the majority group."[22] Further, BIFSG does not account for multiracial households or individuals who identify as more than one race and ethnicity.

It is important to note that BIFSG, while being inherently flawed, was not designed for use in an insurance regulatory context. It is also important to note that even if BIFSG were a reliable methodology for purposes of inferring race or ethnicity, there are not reliable methodologies for purposes of inferring other protected classes. To suggest that there will be other methodologies available in the future to infer protected class status of protected classes other than race is presumptuous at best and raises its own concerns of bias. For instance, it is difficult to contemplate how an insurer would predict someone's religion or sexual orientation with any degree of confidence.

The lack of reliable demographic data is not the end of the matter. The testing methodologies suggested by advocates and some policymakers for assessing output for bias are also flawed. One of the most prominent advocate suggestions for output testing is to introduce a control variable that would represent a protected class in a multivariate analysis.[23] To understand the testing methodology being suggested requires a certain level of working knowledge of how multivariate models work. The argument for a control variable's utility in multivariate analysis is that "[i]f one of the predictive variables, [say, X], is a perfect proxy for race, then when race is added as a control variable, [variable X] loses all of its predictive value."[24] In short, advocates would argue that the predictive variable is more correlated with race than risk and, therefore, its inclusion in risk rating creates disproportionately negative outcomes for race. This argument falls short, however, when looking at the impact of additional variables in a multivariate analysis. The simple fact that a variable - any variable - was added to the equation takes predictive power away from the factors already in the model. As such, introduction of race as a control variable resulting in loss of predictive value for the other variables may be due to the fact that a control variable was added and has nothing to do with race at all. To use a multivariate analysis with race as a control variable would therefore give the illusion that the predictive value of a data element is derived from race when that illusion isn't reflective of reality and the data element may not be deriving any predictive value from race. Using this analysis as a basis for policy would inappropriately lead to a restriction on an insurer's ability to use data elements predictive of risk as rating factors, which would result in inaccurate pricing and harm policyholders.

---

[21] Id., Abstract.
[22] Id., p. 3.
[23] Sclafane, Susanne. Race as a Control Variable (April 6, 2021), https://www.carriermanagement.com/news/2021/04/06/219011.htm.
[24] Id.

CONCLUSION

As illustrated throughout this paper, unfounded notions about the effects of using big data and AI in insurance are running rampant, inappropriately serving as the basis for and inappropriately driving creation of new policy relative to risk-based pricing.

Insurance is about risk transfer, and fairness in insurance is accurately matching price to a policyholder's risk based on economic, insurance, and actuarial principles. Differences in rates are based on differences in risks. State insurance codes explicitly prohibit unfair discrimination in risk-based pricing by requiring rates to be actuarially sound and prohibiting any risk classification based on race, creed, national origin, or religion of an insured. In the context of big data and AI, if an insurer's use of them in risk rating produces a rating factor that is not predictive of risk or is itself a protected class, that is a violation of state insurance law.

It is worth restating that not all risks are the same and neither are all rates. To inject discussions of disparate impact into the insurance policy landscape under the guise of big data and AI concerns is simply a red-herring and will do nothing to solve any underlying problems that contribute to a policyholder's presented risk. Fact must be separated from fiction. If not, the myths discussed in this paper will continue to propel policy efforts, with the only natural consequence being harm to insurance consumers.

NATIONAL ASSOCIATION OF MUTUAL INSURANCE COMPANIES

3601 Vincennes Road | Indianapolis, IN 46268 | 317.875.5250
20 F Street, NW, Suite 510 | Washington, D.C. 20001 | 202.628.1558

NAMIC

NATIONAL ASSOCIATION OF MUTUAL INSURANCE COMPANIES