Sept. 18, 2025

To the Casualty Actuarial and Statistical (C) Task Force, Interested Regulators, and Interested Parties:

After a discussion at the recent open CASTF meeting, we are asking for additonal comments regarding the issue of "unknown risk characteristics" in personal lines rating. Below is a summary of our understanding of the problem. **Please provide written comment by Monday, Oct. 20.**

An insurer's rating plan for property and casualty insurers (e.g., personal auto insurers, homeowners' insurers) includes many risk categories or "classes" that are assigned rating factors. The purpose of these rating factors is to charge premium that correlates with risk. Risk classes are characteristics such as house location, telematics scores, insurance scores, vehicle insured (with VIN), age of house, number of years of driving experience, annual estimated mileage, number of driving convictions, and many more.

In the past, insurers asked consumers for the information that would be used to rate them. This was the purpose of the insurance application. Consumers were asked questions like

- How long have you owned your car?
- Does it have anti-lock brakes?
- How many bathrooms are in your home?

Today, insurers are increasingly using third-party data vendors to gather this and other information about consumers. There are advantages to using third-party vendors for this information in that consumers or agents can miscode or even intentionally attempt to lower rates with false information. In addition, using third-party vendors can lead to expense savings for the insurer due to the cost of communicating with the consumer or agent, sending mailings, properly coding the information once received, etc.

Insurers (or third-party data vendors) may not have all the specific class information for every consumer. When the insurer does not have or obtain the information, the consumer is typically rated using "unknown" for the particular risk category. Sometimes the consumer would be able to supply the needed information, but the insurer does not always contact the consumer to ask. For example, if "number of bathrooms" is a variable in a home insurance

rating algorithm and the insurer and third party do not know the number of bathrooms, "number of bathrooms" for that consumer might be treated as "unknown."

Sometimes the "unknown" category will be assigned a rating factor of 1.00, which means that the premium is the same before and after the associated variable is considered in the premium algorithm. The insurer may contend that assigning 1.00 removes the characteristic from consideration for that particular insured and does not penalize them. However, in the case of a discount with two options – say, a .90 factor for a type 1 discount and a .70 for a type 2 discount – assigning the "unknown" a factor of 1.00 means the consumer does not qualify for a discount and gets the highest rates compared to everyone whose data is not missing.

Other times an "unknown" risk may be slotted in a distinct category or be assigned a weighted average of the factors for the non-missing categories. Using a distinct category may be appropriate if missingness is not random. For example, a lack of credit information likely means that the consumer does not use credit. When such is the case, it makes sense that a "no hit" score would be assigned a distinct factor selected judgmentally by the insurer. By contrast, using a weighted average of the factors for the non-missing categories may be appropriate if missingness is random.

In an attempt to summarize the change occurring through theory, perhaps we could describe that as changing our regulatory view of risk-based pricing from "two consumers with the same risks transferred to the insurer receive the same rate" to "two consumers with the same *identified* risks transferred to the insurer receive the same rate." In reality, the latter statement has always been the theory. Companies have never had 100% of information to conduct rating and have had to make judgmental choices about the rates or further investigate. Although it certainly affects the individual policyholders, the treatment of "unknown" characteristics does not have a significant impact on the book of business if only a small percentage of consumers have missing information. But what if the information for a variable is missing for a larger percentage of consumers, say 10-50%? Some regulators believe rating variables should not be used if there are a large number of unknowns. Some insurers believe using rating variables with a large number of unknowns still obtains a more accurate rate for a majority of consumers.

Some regulators are concerned that the treatment of "unknown" data means that consumers are not being charged a rate conmensurate with their risk. Regulators contend that insurers could sometimes obtain the information and thereby charge fairer rates if they asked the consumer or agent to provide the missing information, or re-queried the third-party vendor.

Not having data is one issue; not receiving data due to system error is another. Consumers might have retrievable data for a risk characteristic, but the insurer or the third party may have a system error that keeps the data from being retrieved at that moment. This likely causes the consumer to be slotted into the "unknown" category.

Adding to the issue, the variables used are often part of a confidential rating model, so consumers do not know what those categories are and would not know whether the data used is correct or not. If the consumer knew that "number of bathrooms in the home" was a rating variable and that the value of this variable was treated as "unknown" for the consumer, the consumer might be able to correct this information by providing the insurer with information about the number of bathrooms. Insurers would be allowed to ask for proof; this is especially important in situations in which there might be an advantageous answer that would lower the consumer's rate.

To be clear, we are seeking input on the problem of rating policies with unknown risk characteristics. This is distinct from the concerns regulators have regarding missing data in the risk modeling process. There is some overlap to be sure, but for the purposes of this letter, we ask that comments be limited to the scenario of determining appropriate premium for an individual applicant where some of the rating characteristics are unknown.

We would appreciate hearing your comments on this issue. Questions you might consider include:
- Does an insurer have a responsibility to ask the consumer about rating characteristics it was unable to collect another way? What risks or challenges might be involved with asking a consumer or agent to supply missing data?
- Does the insurer have a responsibility to notify the consumer when a risk charactistic is unknown?
- How do the Actuarial Standards of Practice apply when insurers are dealing with unknown rating characteristics in calculating an insured's premium?
- What methods are appropriate for determining factors for "unknown" characteristics? Does this differ based on whether or not the data is missing at random?
- If the data is frequently missing for a variable, how should the insurer decide whether or not to keep that variable in the rating algorithm? Are there thresholds that should apply – e.g., if data is missing for more than X% of consumers, that variable should not be used?
- How should insurers think about the concept of fairness when using variables for which information is missing for some consumers?

Thank you for your input.

Christian Citarella, CASTF Chair
Julie Lederer, CASTF Vice Chair


Cc: Kris DeFrain and Roberto Perez (NAIC)