October 20, 2025

Christian Citarella
Chair, Casualty Actuarial and Statistical (C) Task Force
National Association of Insurance Commissioners

Re: Request for additional comments regarding the issue of "unknown risk characteristics" in personal lines rating

Dear Chair Citarella:

On behalf of the American Academy of Actuaries'[1] Data Science and Analytics Committee (DSAC), we appreciate the opportunity to respond to the Casualty Actuarial and Statistical (C) Task Force's (CASTF) request for additional comments regarding the issue of "unknown risk characteristics" in personal lines rating. We value CASTF's work addressing the problem of rating policies with unknown risk characteristics.

Our overarching perspective is that the appropriate response to missing data is dependent upon the variable in question and the use case. A wide variety of responses are possible, as detailed below. Each response has a context where it is an appropriate balance of practicality and accuracy. Alternately, each response has a different context where it could be highly impractical or negatively impact accuracy. It is important for modelers to be deliberate about their responses and be able to defend their decisions. It is important for regulators to inquire about insurers' approaches and challenge them as needed. Mandating one response in all situations, such as contacting the consumer, would have non-optimal insurance market impacts.

It is beneficial to address the topic as much as possible within the current insurance and regulatory framework, creating new practices where there is a newly identified risk.

Background/Context

Missing data is a well-established field of study in statistics and a common challenge in many databases. It is a subtopic of larger considerations related to data quality. Insurers leverage this knowledge in statistical work related to pricing and underwriting policies, especially when data science is used to develop predictive models for these purposes. This is a challenge in all lines of business.

Insurers have always needed to make decisions about how to handle missing information. For example, determining how to proceed if a consumer refuses to answer an application question.

[1] The American Academy of Actuaries is a 20,000-member professional association whose mission is to serve the public and the U.S. actuarial profession. For 60 years, the Academy has assisted public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.

Insurer practices have developed to handle these situations, and regulators have also developed practices to oversee them. Examples of these practices include filing reviews, filing exhibits/questionnaires, and market conduct exams. This topic is within the scope of current regulatory authority. Regulators can, and do, challenge insurer practices for handling missing information that they find unsatisfactory.

The DSAC cannot comment directly on whether or not missing data is becoming more common. On one hand, it is plausible that increased use of third-party data could lead to an increased frequency of missing data since practical considerations can lead to databases not being fully populated (e.g., a database of car maintenance history would have gather information from a large number of maintenance shops). On the other hand, increased access to external data used for application prefill and policy data verification could decrease the frequency of missing data. The need to handle missing data responsibly is independent of whether it is becoming more common.

Responses to Questions

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?

There are some contexts in which asking a consumer about a rating characteristic is the best course of action. However, there are also situations where consumers may not be the best information source. For example, consumers may not know or be able to obtain the missing information. Consumers may be inclined to misrepresent the information to receive better treatment. The best course of action is dependent upon the nature of the missing information and the use to which it is being put.

Our comments focus on scenarios when requesting consumer input seems reasonable. The following examples identify when consumer input seems *less* reasonable. It's important to consider the broad range of variables and use cases when discussing this topic.

- A query of motor vehicle reports (MVRs) may return "missing" information about whether the consumer had a traffic violation. An insurer would reasonably assume that a lack of information would indicate that the consumer had no violations and group the "Unknown" category with the "no violation" category. However, this data may also be missing for other reasons, such as an error on the MVR. Asking the consumer to verify the accuracy of this data is unlikely to yield any benefit.

- Suppose an insurer appends vehicle safety ratings (i.e., a crash test) to their book of business. They find that 20% of their book is high risk, 20% is low risk, and the remainder have not been subject to the test and are labeled "unknown." The unknown class could be treated neutrally, providing improved risk segmentation to the remaining policies. Alternatively, an insurer could seek to estimate the safety rating (i.e., interpolate) by identifying similar vehicle characteristics as those that underwent the test. Consumers would not be able to supply this information.

- An important note for CASTF is to keep in mind that asking consumers for direct feedback or information that results in less than ideal data is not solely an issue for property and casualty insurance products. This topic is present across all lines of business. For example, in life insurance, data is used to detect risk factors where consumers are unlikely to provide accurate information, like tobacco use or high cholesterol. That data is then used in underwriting and to determine eligibility for specific life insurance products.

A significant consideration in whether to use data in a model is the cost of that data compared to its marginal predictive value. Asking a consumer to verify data increases costs in terms of time spent sending requests, collecting responses, verifying responses, and updating policies. These costs might be significant enough that they override the value of the data, although using less data reduces the accuracy of predictions, negatively impacting the insurance market. Regulators might also incur additional costs associated with filing reviews and market conduct exams of these processes.

Does the insurer have a responsibility to notify the consumer when a risk characteristic is unknown?

This question relates to the broader concern of transparency. Considerations such as accuracy, expense, protection of trade secret business practices, and consumer experience need to be weighed against each other. There is nothing materially different between unknown risk classification and other aspects of risk classification which may or may not be of interest to consumers.

How do the Actuarial Standards of Practice apply when insurers are dealing with unknown rating characteristics in calculating an insured's premium?

Actuarial Standard of Practice No. 12, *Risk Classification (for All Practice Areas)* lists items actuaries should consider when designing, reviewing, or changing risk classification systems. These are included below, along with brief commentary on how each might apply to unknown risk classification:

- Relationship of Risk Characteristics and Expected Outcomes—"unknown" can be treated as its own category if it is correlated with expected losses.
- Causality—It is not necessary to determine causality. Consequently, it is not necessary for insurers to determine whether the reason data is missing is causally related to insurance risk.
- Objectivity (based on readily verifiable facts that cannot be easily manipulated)—Missing data can occur for specific, non-subjective reasons that depend on the type of data.
- Practicality (reflects the tradeoffs between practical and other relevant situations)—The cost, time, and effort to populate missing information is relevant to consider.
- Applicable Law—Laws prohibiting unknown risk classification would need to be followed.
- Industry Practice—Industry practice to address missing data should be considered. The options listed in this letter describe industry practice.

Actuarial Standard of Practice No. 23, *Data Quality* encourages actuaries to balance data limitations against the practicality of obtaining alternative data. Actuaries are prompted to use their professional judgment about the appropriateness of the data for the intended purpose. It lists the following considerations for selecting data which are relevant to this topic:

- whether the data constitute appropriate data
- whether the data are reasonable with particular attention to internal consistency
- whether the data are reasonable given relevant external information that is readily available and known to the actuary
- the degree to which the data are sufficient
- any known significant limitations of the data
- the availability of additional or alternative data and the benefit to be gained from such additional or alternative data, balanced against how practical it is to collect and compile such additional or alternative data

Actuarial Standard of Practice No. 56, *Modeling* states that actuaries should understand the model's intended purposes when determining appropriate design, data, and model risk evaluation. Treatment of unknown/missing data falls within this guidance.

Finally, actuarial standards require transparency in communication (Actuarial Standard of Practice No. 41, *Actuarial Communication*), as well as honesty, integrity, and competence (Precept 1 of the Actuarial Code of Professional Conduct).

What methods are appropriate for determining factors for "unknown" characteristics? Does this differ based on whether or not the data is missing at random?

A wide variety of responses are available, including neutral treatment, best/worst treatment, average treatment, interpolation, requesting information from the consumer, and declining/non-renewing the policy. Each option is appropriate for balancing practicality and accuracy in some contexts but not others. Alternately, each response has a context in which it could be highly impractical or inaccurate. Whether or not the data is missing at random is an example of a consideration which might lead to different responses. It is not desirable, and arguably not possible, to articulate firm standards for how to respond to all reasons why data might be missing.

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?

Actuaries use their professional judgement in these scenarios, taking into consideration the credibility, predictive power, and availability of the information. Some rating variables have high proportions of missing information, such as MVRs and loss history. It is impractical and undesirable to establish a single course of action for all types of data. However, regulators could identify scenarios when they may desire a more in-depth review of insurer practices.

Additionally, the data used to create the rating plan ("development data") can either under- or overestimate the frequency of missingness in production. In an extreme case, data that is missing in the development data might never be missing in production. The nature of the use case and the particular data need to be considered in order to assess what level of missingness is a concern.

How should insurers think about the concept of fairness when using variables for which information is missing for some consumers?

Missing data needs to be handled with care by both insurers and regulators. For rates to be as fair as possible, insurers should be able to use a wide variety of data sources, provided they do so responsibly.

Rates are not unfairly discriminatory if they are actuarially justified estimates of future loss experience. More accurate rates are fairer in that they match the premium paid for the policy with the expected services that will be obtained from the policy. More accurate rates benefit society by improving both the affordability and availability of insurance; this is because insurers can confidently offer coverage to high-risk consumers and reduce margins for uncertainty in their premiums. Also, setting premiums based on the risk profile of the insured discourages anti-selection and encourages risk-mitigation activities. When risk classifications are prohibited or discouraged, lower risk consumers pay more than necessary to fund discounts on higher risk consumers. In order to promote the fairest outcomes in the insurance market, data sources with missing data should evaluated carefully, but should not be discouraged or prohibited.

*****

We appreciate the opportunity to share this feedback and our recommendations in response to CASTF's request for additional comments. If you have any questions or would like to discuss these comments further, please contact Will Behnke, the Academy's Risk Management and Financial Reporting policy project manager (behnke@actuary.org).

Sincerely,

Kirsten Pedersen MAAA, FSA
Chairperson, Data Science and Analytics Committee
American Academy of Actuaries