

## **CASUALTY ACTUARIAL AND STATISTICAL (C) TASK FORCE**

Casualty Actuarial and Statistical (C) Task Force Aug. 10, 2022, Minutes  
Casualty Actuarial and Statistical (C) Task Force July 12, 2022, Minutes (Attachment One)  
Random Forest Models and Tree-Based Models Presentation (Attachment One-A)  
Casualty Actuarial and Statistical (C) Task Force June 14, 2022, Minutes (Attachment Two)  
Comments on Loss Cost Multiplier Form (Attachment Two-A)  
Actuarial Opinion (C) Working Group Aug. 2, 2022, Minutes (Attachment Three)  
Referral from Financial Analysis (E) Working Group (Attachment Three-A)  
Statistical Data (C) Working Group Aug. 4, 2022, E-Vote Minutes (Attachment Four)  
Statistical Data (C) Working Group July 20, 2022, Minutes (Attachment Five)  
Statistical Data (C) Working Group June 15, 2022, Minutes (Attachment Six)  
Statistical Data (C) Working Group May 18, 2022, Minutes (Attachment Seven)  
Statistical Data (C) Working Group April 14, 2022, Minutes (Attachment Eight)  
Regulatory Review of Tree-Based Models Guidance (Attachment Nine)  
openIDL Initiative Presentation (Attachment Ten)

[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022\\_Summer/CASTF/Contents.docx](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022_Summer/CASTF/Contents.docx)

## Draft Pending Adoption

Draft: 8/19/22

Casualty Actuarial and Statistical (C) Task Force  
Portland, Oregon  
August 10, 2022

The Casualty Actuarial and Statistical (C) Task Force met in Portland, OR, Aug. 10, 2022. The following Task Force members participated: Mike Kreidler, Chair, represented by Eric Slavich (WA); Grace Arnold, Vice Chair, represented by Phil Vigliaturo (MN); Ricardo Lara represented by Ken Allen and Lynne Wehmueller (CA); Michael Conway represented by Sydney Sloan (CO); Andrew N. Mais represented by Wanchin Chou (CT); Colin M. Hayashida represented by Randy Jacobson (HI); Doug Ommen represented by Travis Grassel (IA); Dana Popish Severinghaus represented by Bruce Sartain and Judy Mottar (IL); Vicki Schmidt represented by Nicole Boyd (KS); Timothy N. Schott represented by Sandra Darby (ME); Chlora Lindley-Myers represented by Cynthia Amann and Julie Lederer (MO); Mike Causey represented by Richard Kohan (NC); Russell Toal and Anna Krylova (NM); Judith L. French represented by Tom Botsko (OH); Glen Mulready represented by Andrew Schallhorn (OK); Michael Humphreys represented by Shannen Logue (PA); Michael Wise represented by Will Davis (SC); Cassie Brown represented by Miriam Fisk (TX); Kevin Gaffney and Rosemary Raszka (VT); and Allan L. McVey represented by Greg Elam (WV).

### 1. Adopted its July 12, June 14, and Spring National Meeting Minutes

Mr. Slavich said the Task Force met July 12, June 14, and March 8. During these meetings, the Task Force took the following action: 1) exposed the loss cost multiplier (LCM) form for a 40-day public comment period ending June 7; and 2) exposed the regulatory review of tree-based model guidance for a 25-day public comment period ending Aug. 5.

The Task Force also met June 21, May 17, April 19, and March 15 in regulator-to-regulator session, pursuant to paragraph 3 (specific companies, entities or individuals) of the NAIC Policy Statement on Open Meetings, to discuss rate filing issues.

The Task Force held Predictive Analytics Book Club meetings on July 26, June 28, April 26, and March 22. On July 26, Betterview presented on “Leveraging Computer Vision and AI for Property and Risk Management.” On June 28, Dorothy L. Andrews (American Academy of Actuaries—Academy) and Julia Romero (Academy) presented on “Correlation vs. Causation.” On April 26, Kris DeFrain (NAIC), Ms. Andrews (NAIC), Sam Kloese (NAIC), and Roberto Perez (NAIC) presented on “Speed to Market.” On March 22, Brian Fannin (Casualty Actuarial Society—CAS) presented on “R for Actuaries and Data Scientists with Application to Insurance.”

Mr. Vigliaturo made a motion, seconded by Mr. Botsko, to adopt the Task Force’s July 12 (Attachment One); June 14 (Attachment Two); and March 8 (*see NAIC Proceedings – Spring 2022, Casualty Actuarial and Statistical (C) Task Force*) minutes. The motion passed unanimously.

### 2. Adopted the Report of the Actuarial Opinion (C) Working Group

Ms. Krylova said the Actuarial Opinion (C) Working Group met Aug. 2. The Working Group discussed a Financial Analysis (E) Working Group referral on predictive analytics in a reserve setting, potential changes to the qualification documentation requirements and disclosures, and potential changes to 2022 Regulatory Guidance and 2023 *Annual Statement Instructions*. The Actuarial Opinion (C) Working Group has begun to draft some financial surveillance questions that could be asked about reserve models. Because the qualification documentation is a burden on actuaries and does not change much year to year, the Working Group is discussing

## Draft Pending Adoption

whether the qualification documentation only needs to be submitted to the Board ever five years and when there have been significant changes in the actuarial qualifications or a company's operations.

Ms. Krylova said the Working Group also met July 22 and June 3 in regulator-to-regulator session, pursuant to paragraph 3 (specific companies, entities or individuals) of the NAIC Policy Statement on Open Meetings, to discuss 2021 Statements of Actuarial Opinion (SAOs). She said no serious issues or trends were identified.

Ms. Krylova made a motion, seconded by Mr. Botsko, to adopt the report of the Actuarial Opinion (C) Working Group, including its Aug. 2 minutes (Attachment Three). The motion passed unanimously.

### 3. Adopted the Report of the Statistical Data (C) Working Group

Ms. Darby said the Statistical Data (C) Working Group met Aug. 4, July 20, June 15, May 18, and April 14. During these meetings, the Working Group: 1) discussed implementing an accelerated timeline for the *Auto Insurance Database Report* (Auto Report); 2) heard a presentation from the Center for Economic Justice (CEJ) on the modernization of statistical data reporting; and 3) discussed proposed changes to the Auto Report, the *Dwelling, Fire, Homeowners Owner-Occupied, and Homeowners Tenant and Condominium/Cooperative Unit Owner's Insurance Report* (Homeowners Report), the *Competition Database Report* (Competition Report), and the *Report on Profitability by Line by State* (Profitability Report) presented by Arthur Schwartz (LA). As previously reported, the Working Group adopted an accelerated timeline for the submission of auto insurance premium and exposure data. Statistical agents will now provide that data by Dec. 1 following the end of the data year. The Working Group will continue to consider the proposed changes to the statistical reports, as well as consider the need for updates to the *Statistical Handbook of Data Available to Insurance Regulators* and the statistical data reporting process.

Ms. Darby made a motion, seconded by Mr. Chou, to adopt the report of the Statistical Data (C) Working Group, including its Aug. 4 (Attachment Four), July 20 (Attachment Five), June 15 (Attachment Six), May 18 (Attachment Seven), and April 14 (Attachment Eight) minutes. The motion passed unanimously.

### 4. Exposed the LCM Form and Memorandum

Mr. Slavich said the idea to create an updated LCM form was brought forward at the Spring National Meeting, after which Larry Steinert (IN) led a volunteer group to update and combine the NAIC's numerous forms. At the July 12 meeting, there were no suggested changes to the form; however, the Task Force decided to wait on adoption until the LCM memorandum could be updated and considered for adoption at the same time. A proposed memorandum was distributed for the call. The Task Force had no questions about the form or memorandum.

The Task Force exposed the LCM form and associated memorandum for a 45-day public comment period ending Sept. 23.

### 5. Adopted the Regulatory Review of Tree-Based Models Guidance

At the Spring National Meeting, the Task Force adopted the appendix for the random forest models as regulatory guidance and noted that the plan would be to combine all similar appendices together for consideration as an attachment to the *Regulatory Review of Predictive Models* white paper.

At the July 14 meeting, Mr. Kloese proposed that the random forest guidance be replaced with guidance for all tree-based models. Mr. Slavich said the guidance did not need to change substantially. The Task Force exposed the regulatory review of tree-based model guidance for a 25-day public comment period ending Aug. 5. No comment letters were received.

## Draft Pending Adoption

Mr. Vigliaturo made a motion, seconded by Mr. Botsko, to adopt the regulatory review of tree-based models document (Attachment Nine) as a replacement for the random forest document. The motion passed unanimously.

Mr. Kloese said he plans to next draft generalized additive modeling (GAM) guidance for Task Force review and consideration.

### 6. Received a Report on the NAIC Algorithmic Bias Training

At the Spring National Meeting, Mr. Slavich said he gave a report about coordination with the Innovation, Cybersecurity, and Technology (H) Committee and the Special (EX) Committee on Race and Insurance Workstream Three. He said that coordination led to the NAIC hosting a Collaboration Forum on Algorithmic Bias in Kansas City, July 18–19. Ms. Andrews gave a brief report on the forum, saying the forum contained numerous sessions on bias, how bias can get into data, governance addressing bias, and data needed for bias detection.

### 7. Heard a Presentation on the openIDL Initiative

Mr. Slavich introduced Jefferson Braswell (openIDL) and said openIDL is the Linux Foundation project that is coordinating the development of a collaborative, open network of insurance carriers, analytical services, and state insurance commissioners to enable more efficient and timely access to insurance industry data on the part of commissioners. He said the Task Force is hearing the presentation because of the work at the Statistical Data (C) Working Group. Mr. Braswell gave his presentation (Attachment Ten).

### 8. Heard from Professional Actuarial Organizations

The Academy, the Actuarial Board for Counseling and Discipline (ABCD), and the Society of Actuaries (SOA) provided reports on current activities.

Having no further business, the Casualty Actuarial and Statistical (C) Task Force adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/08-10 min.docx

Draft: 7/28/22

Casualty Actuarial and Statistical (C) Task Force  
Virtual Meeting  
July 12, 2022

The Casualty Actuarial and Statistical (C) Task Force met July 12, 2022. The following Task Force members participated: Mike Kreidler, Chair, represented by Eric Slavich (WA); Grace Arnold, Vice Chair, represented by Connor Meyer (MN); Ricardo Lara represented by Mitra Sanandajifar (CA); Michael Conway represented by Mitchell Bronson (CO); Andrew N. Mais represented by Wanchin Chou (CT); Karima M. Woods represented by David Christhilf (DC); David Altmaier represented by Greg Jaynes (FL); Colin M. Hayashida represented by Randy Jacobson (HI); Dana Popish Severinghaus represented by Anthony Bredel and Reid McClintock (IL); Amy L. Beard represented by Larry Steinert (IN); Vicki Schmidt represented by Nicole Boyd (KS); James J. Donelon represented by Arthur Schwartz and Nichole Torblaa (LA); Kathleen A. Birrane represented by Ronald Coleman and Walter Dabrowski (MD); Timothy N. Schott represented by Sandra Darby (ME); Chlora Lindley-Myers represented by Cynthia Amann and Julie Lederer (MO); Troy Downing represented by Mari Kindberg (MT); Mike Causey represented by Michelle Osborne (NC); Chris Nicolopoulos represented by Christian Citarella (NH); Russell Toal and Anna Krylova (NM); Barbara D. Richardson represented by Gennady Stolyarov (NV); Judith L. French represented by Tom Botsko (OH); Glen Mulready represented by Andrew Schallhorn (OK); Andrew R. Stolfi represented by David Dahl (OR); Michael Humphreys represented by Michael McKenney (PA); Cassie Brown represented by J'ne Byckovski (TX); Kevin Gaffney represented by Rosemary Raszka (VT); and Allan L. McVey represented by Juanita Wimmer (WV).

1. Adopted the Report of the Statistical Data (C) Working Group

Ms. Darby said the Statistical Data (C) Working Group met June 15 to discuss the timeline of the *Auto Insurance Database Report* (Auto Report) and changes proposed by Mr. Schwartz to the *Dwelling, Fire, Homeowners Owner-Occupied, and Homeowners Tenant and Condominium/Cooperative Unit Owner's Insurance Report* (Homeowners Report); the Auto Report; the *Competition Database Report* (Competition Report); and the *Report on Profitability by Line by State* (Profitability Report). The Working Group voted to accelerate the timeline of the Auto Report. Written premium and exposure data will now be submitted by Dec. 1 following the end of the data year. This will move up the publication of the report by about six months. This is generally the same timeline adopted for the Homeowners Report in a prior meeting.

The Working Group plans to meet July 20 to continue the discussion of Mr. Schwartz's proposed changes. It also plans to meet in August after the Summer National Meeting in regulator-to-regulator session to discuss the data for this year's Homeowners Report and Auto Report.

Ms. Darby made a motion, seconded by Ms. Krylova, to adopt the report of the Statistical Data (C) Working Group. The motion passed unanimously.

2. Discussed the NAIC's LCM Form

The idea to compile and update multiple NAIC loss cost multiplier (LCM) forms into one was brought forward at the Spring National Meeting. Mr. Steinert's group of volunteers submitted an initial proposal, which was exposed for a public comment period that ended Feb. 7. The group revised the proposal.

Mr. Steinert said the group proposes that the form be in an Excel file. State insurance regulators can protect cells and include calculations. Mr. Steinert walked through the revised form and pointed out some programming and

wording changes. He said the group did not include the System for Electronic Rates & Forms Filings (SERFF) filing number and the NAIC group code because those can be found in the filing elsewhere. The profit and contingencies entry was combined with investment income offset. Additional calculations of percent change were added.

Mr. Steinert said there is an instructions document that will need some updates. He suggests that the instructions remain a separate document because it is more applicable to providing guidance about loss costs and loss cost filings rather than technical instructions about the LCM calculations. Mr. Schwartz volunteered to redraft the instructions document. Mr. Slavich asked the group to produce an instructions document to be considered for adoption as a package with the proposed form at the Summer National Meeting.

Ms. Darby asked for the consideration of states that require a portable document format (PDF) version in SERFF. The PDF version is needed for production of the SERFF pipeline document. The group will consider adding some instructions for how to convert an Excel file into a PDF document for SERFF filing purposes.

Mr. Steinert said the Excel file will have a password with locked cells. Only the input cells would be changeable, not the formulas.

### 3. Discussed the Regulatory Review of Tree-Based Rate Models

At its meeting in lieu of the Spring National Meeting, the Task Force adopted a document describing the regulatory review of random forest models and accompanying definitions. Sam Kloese (NAIC) began work to develop similar materials for gradient boosting machines (GBMs) and decided to instead propose a modification to the random forest documents for them to apply more broadly to all tree-based models.

Mr. Kloese provided education on random forest models and tree-based models (Attachment One-A). He presented his proposal for the regulatory review of tree-based models.

With few changes from the already adopted random forest models document, the Task Force agreed to a short exposure. Mr. Slavich said comments would be due Aug. 5 and discussed at the Summer National Meeting.

### 4. Discussed Other Matters

Mr. Slavich said the Task Force discussed the possibility of creating a handbook to help guide NAIC staff in their technical reviews of models during its June 14 call. NAIC staff will discuss the alternative of creating a steering committee with senior NAIC management.

Having no further business, the Casualty Actuarial and Statistical (C) Task Force adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/07-12-22/07-12 min.docx



## Boosted Trees

Sam Kloese, ACAS, CSPA  
 P/C Rate Modeling Actuary  
 June 14, 2022



Image from Introduction to Boosted Trees, The Official Blog of BigM, com

## Introduction

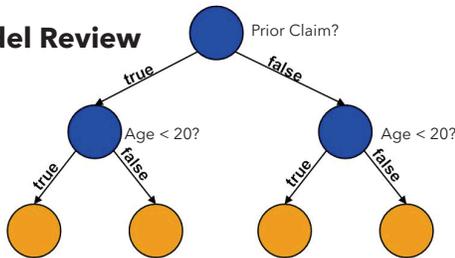
- GLM's are industry standard
- The CASTF White Paper for Predictive Models is focused primarily on GLM's
- New Appendix for Random Forests has been adopted
- Boosted trees are very similar with a couple differences

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 2

## Tree Based Model Review

- Single Decision Tree
- Easy to Understand
- Mimics how people make decisions
- Easily interpreted

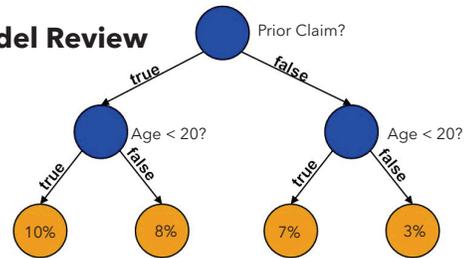


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 3

## Tree Based Model Review

- Single Decision Tree
- Easy to Understand
- Mimics how people make decisions
- Easily interpreted
- Classification returns a likelihood

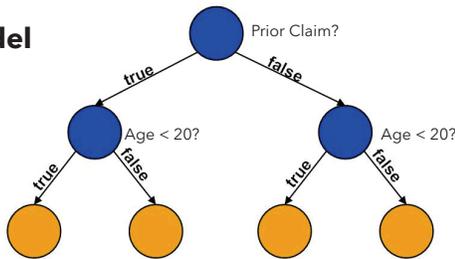


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 4

## Tree Based Model

- Single Decision Tree
- Easy to Understand
- Mimics how people make decisions
- Easily interpreted
- Classification returns a likelihood

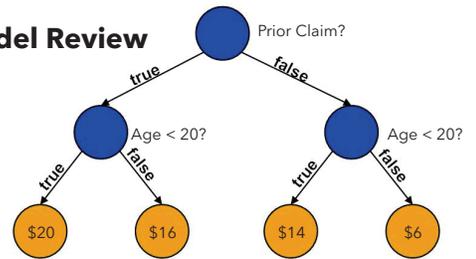


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 5

## Tree Based Model Review

- Single Decision Tree
- Easy to Understand
- Mimics how people make decisions
- Easily interpreted
- Classification returns a likelihood
- Regression returns a predicted amount

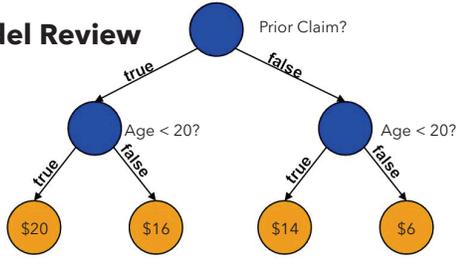


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 6

## Tree Based Model Review

- Terminology
  - Nodes
    - Root
    - Sub-Node
    - Parent/Child
  - Splitting
    - Branch
    - Sub-Tree

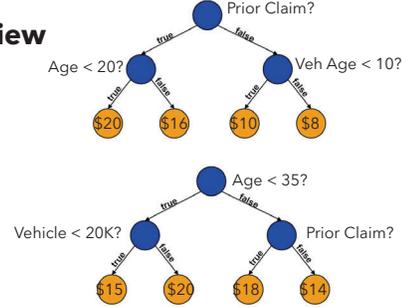


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 7

## Random Forest Review

- Random Forest
  - Each tree is based on a different bootstrap sample
  - Additionally: Randomly chosen variables considered at each split
  - Each tree is grown the same way
  - Final prediction is the average of each trees prediction
- Advantages
  - Trees are substantially different
  - Each tree not based on the same sample
  - Each split not based on the same variables

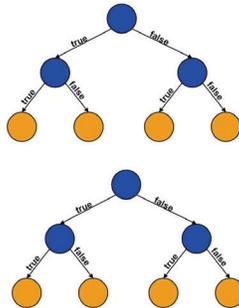


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 8

## Random Forest Review

- Hyperparameters
  - Number of trees
  - Criteria on which to split
  - Bootstrap sample size (% of rows)
  - When to stop splitting
    - Max Tree Depth
    - Minimum Node Size
    - Max Leaf Nodes
  - Random Variables for each split (# of columns)



NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 9

## Differences Between RF and Boosted Trees

- Main differences
  - Boosting is run sequentially
    - Each subsequent tree attempts to refine the model further
    - The data is reweighted after each tree, so subsequent trees focus more on wrong predictions
    - Often takes longer to run than RF, trees can not be run simultaneously
  - New Hyperparameter "Learning Rate"
    - Also known as "shrinkage parameter"
    - Between 0 and 1 (usually small, closer to 0)
    - There is a trade off between number of trees and learning rate

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 10

## Boosted Trees

- Boosted Trees algorithms:
  - Adaboost (Adaptive Boosting)
  - GBM (Gradient Boosting Machines)
  - XGBoost (eXtreme Gradient Boosting)
  - CatBoost (Categorical Boosting)

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 11

## XGBoost

- Particularly Popular Boosting Algorithm
  - Many machine learning competitions won with XGBoost
  - Available in open source software R and Python
- Xgboost() function arguments
  - Common Hyperparameters
    - Max\_depth: how many splits deep can a tree grow
    - Nrounds: Number of trees
    - Eta: Learning Rate
    - Colsample\_bytree: % of columns (predictor variables) to use at each split
    - Subsample: % of rows (records) to use within each tree

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 12

## XGBoost

- Other xgboost() function arguments
  - Specifying the input data
    - Data = matrix containing predictor variables
    - Label = matrix containing target variable
    - Weight = matrix containing weights (earned exposures, claim counts, etc.)
  - Specifying the learning objection (many options)
    - Objective = Reg:squarederror
    - Objective = Reg:logistic
    - Objective = Count:poisson
    - Objective = Reg:gamma
    - Objective = Reg:tweedie
    - Objective = binary:logistic

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 13

## Tree Based Model Challenges Revisited

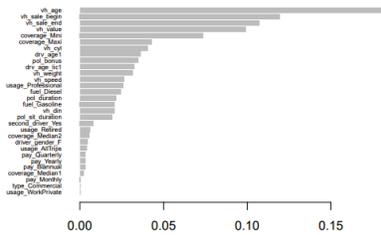
- Interpretability
  - Boosted Trees even less interpretable than RF
  - Individual trees can be counterintuitive, because they target residual from prior trees
- Prone to Overfit
  - Boosted Trees even more likely than RF to be overfit
  - Sequential nature of boosted trees means they can chase after the outliers
- Auditability
  - Same issue with auditability
  - Requires documentation of all trees to reproduce model predictions

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 14

## Variable Triaging

- Variable Importance Plots
  - Provide a measure of which variables are relatively more important than others
  - High importance variables should be evaluated as they will have the greatest impact on consumers
  - Low importance variables should be evaluated for whether there is a good reason to include them
    - Similar to questioning variables with high p-values in a GLM

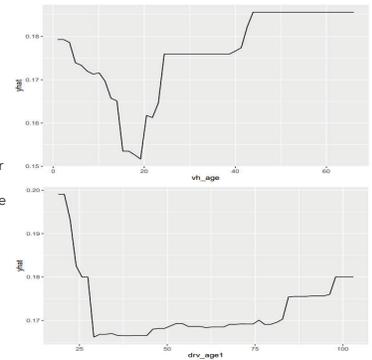


NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 15

## Interpretability Plots

- Partial Dependence Plots
  - Computes the marginal effect of a given feature on the prediction
  - Fixes the value of the predictor variable of interest, calculating the model prediction for each observation using the fixed value
  - Repeat for all values of the predictor variable



NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

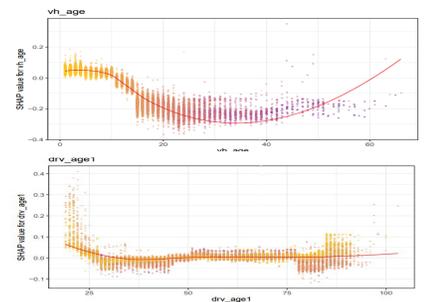
8/19/2022 16

## Interpretability Plots

- Accumulated Local Effects
  - Better option in the case of correlated features
  - Calculates and accumulates incremental changes in the feature effects
  - Shows the expected and centered effects of a feature, like a coefficient in a GLM

## Interpretability Plots

- Shapely Additive Explanations
  - How much that feature moves the prediction away from the overall average prediction.
  - >0, feature increases predicted value higher than average value
  - <0, feature decreases predicted value lower than average value



NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

July 2021 Book Club: <https://www.youtube.com/watch?v=yMdTAlkewk> 8/19/2022 17

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 18

## Challenges - Prone to Overfit

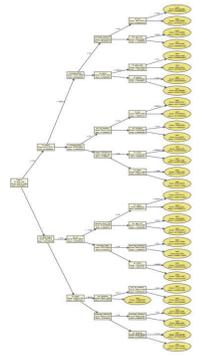
- Review general narrative on setting of hyperparameters
- Hyperparameter Consideration
  - Number of trees should be large enough, but no larger
    - Look at plot to minimize OOB/Test Error or Deviance
  - Learning Rate
    - Greater than 0, commonly < 0.2
  - Tree Complexity
    - Minimum node size should be set high enough for reasonable credibility
    - Rule of Thumb: Max depth of > 8 may be too high
  - Other hyperparameters should be disclosed and briefly commented on
    - Bootstrap sample size (% of rows)
    - Random Variables tried for each split (# of columns)
    - Criteria to split should match the model purpose (classification, regression)
- Review lift charts on test/holdout data

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 19

## Challenges - Auditability

- Comprehensive Documentation
  - Allows for third party audit
    - Makes reproducing predictions from data to decision possible
    - Ideally would not require understanding complex code or specific software
- Document entire tree ensemble
  - Print tree logic for each tree OR
  - Plot all trees OR
  - Provide a table with prediction for all possible values



NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 20

## Challenges - Auditability

- Sample Prediction Examples
  - Exhibits could be made for **spot-checking** against tree documentation
    - Input Predictors
    - Individual Tree Predictions
    - Overall Model Prediction

Sample Risk	Driver Age	Prior Claims	Vehicle Age	...	Tree 1	Tree 2	Tree 3	...	Model Prediction
1	16	0	3	...	\$ 50.00	\$ 40.00	\$ 30.00	...	\$ 40.00
2	17	0	6	...	\$ 49.00	\$ 39.20	\$ 29.40	...	\$ 39.20
3	18	0	2	...	\$ 48.02	\$ 38.42	\$ 28.81	...	\$ 38.42
4	19	1	3	...	\$ 47.06	\$ 37.65	\$ 28.23	...	\$ 37.65
5	20	0	9	...	\$ 46.12	\$ 36.90	\$ 27.67	...	\$ 36.90

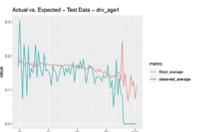
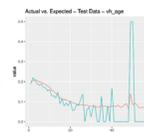
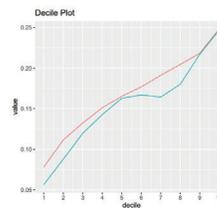
- However, **auditing** every prediction for a book of business would still be **extremely difficult**

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 21

## Review Other Useful Exhibits

- Quantile Plots on Holdout Data
- Actual vs. Expected plots on Holdout Data
- Compares fitted average to observed average by Quantile
- Separate plot by variable
- Demonstrates fit across variable levels



NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 22

## Text References

- Basic Decision Tree Terminology
  - <https://medium.datadriveninvestor.com/the-basics-of-decision-trees-e5837cc2aba7>
- Theoretical Introduction to Random Forest and Boosted Trees
  - Introduction to Statistical Learning (Chapter 8 - 8.2.3)
  - [https://web.stanford.edu/~hastie/ISLRv2\\_website.pdf](https://web.stanford.edu/~hastie/ISLRv2_website.pdf)
- Interpretable Machine Learning (Variable Importance and Interpretability Plots)
  - [https://us.milliman.com/-/media/milliman/pdfs/2021-articles/4-2-21-interpretable-machine-learning\\_ashx](https://us.milliman.com/-/media/milliman/pdfs/2021-articles/4-2-21-interpretable-machine-learning_ashx)

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 23

## Other References

- YouTube Videos:
  - Actuarial Virtual Data Science Seminar: Risk Pricing with XGBoost
    - <https://www.youtube.com/watch?v=sOyMLB1SsFk>
  - CASTF Book Club Call: Tree Based Models
    - <https://www.youtube.com/watch?v=6UCbpAt4r9M>
  - CASTF Book Club Call: Interpretable Machine Learning
    - <https://www.youtube.com/watch?v=yMdTAlkewk>
- Sample Model R Markdown
  - Illustration purposes only, primarily created to demonstrate sample plot types
  - <https://github.com/skloese/Insurance-Pricing-XGBoost-Example/tree/main>
  - E-mail [skloese@naic.org](mailto:skloese@naic.org) if you'd like to see the plots in PDF

NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

8/19/2022 24

[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022\\_Summer/CASTF/07-12-22/Boosted%20Trees%20Background%20Info%20-%20Full.pptx](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022_Summer/CASTF/07-12-22/Boosted%20Trees%20Background%20Info%20-%20Full.pptx)

Draft: 7/27/22

Casualty Actuarial and Statistical (C) Task Force  
Virtual Meeting  
June 14, 2022

The Casualty Actuarial and Statistical (C) Task Force met June 14, 2022. The following Task Force members participated: Mike Kreidler, Chair, represented by Eric Slavich (WA); Grace Arnold, Vice Chair, represented by Connor Meyer and Phil Vigliaturo (MN); Jim L. Ridling represented by Daniel Davis (AL); Ricardo Lara represented by Lynne Wehmueller (CA); Michael Conway represented by Mitchell Bronson (CO); Andrew N. Mais represented by Wanchin Chou (CT); Karima M. Woods represented by David Christhilf (DC); David Altmaier represented by Greg Jaynes (FL); Colin M. Hayashida represented by Randy Jacobson (HI); Doug Ommen represented by Travis Grassel (IA); Dana Popish Severinghaus represented by Chantel Long, Reid McClintock, and Judy Mottar (IL); Amy L. Beard represented by Larry Steinert (IN); Vicki Schmidt represented by Nicole Boyd (KS); James J. Donelon represented by Nichole Torblaa and Arthur Schwartz (LA); Kathleen A. Birrane represented by Ron Coleman and Walter Dabrowski (MD); Timothy N. Schott represented by Sandra Darby (ME); Chlora Lindley-Myers represented by Julie Lederer (MO); Troy Downing represented by Mari Kindberg (MT); Mike Causey represented by Michelle Osborne (NC); Chris Nicolopoulos represented by Christian Citarella (NH); Russell Toal and Anna Krylova (NM); Barbara D. Richardson represented by Gennady Stolyarov (NV); Judith L. French represented by Tom Botsko (OH); Glen Mulready represented by Andrew Schallhorn and Kate Yang (OK); Andrew R. Stolfi represented by David Dahl and Ying Liu (OR); Michael Humphreys represented by Kevin Clark and Michael McKenney (PA); Michael Wise represented by Will Davis (SC); Cassie Brown represented by J'ne Byckovski (TX); and Kevin Gaffney represented by Rosemary Raszka (VT). Also participating was: Kevin Dyke (MI).

1. Adopted the Report of the Actuarial Opinion (C) Working Group

Ms. Krylova said the Actuarial Opinion (C) Working Group met in regulator-to-regulator sessions to discuss individual Statements of Actuarial Opinion (SAOs). The Working Group also received a referral from the Financial Analysis (E) Working Group.

Ms. Krylova made a motion, seconded by Ms. Lederer, to adopt the report of the Actuarial Opinion (C) Working Group. The motion passed unanimously.

2. Adopted the Report of the Statistical Data (C) Working Group

Ms. Darby said the Statistical Data (C) Working Group is making progress on its charge to accelerate the reporting of average premiums for auto and homeowners insurance. For both auto and homeowners insurance, 2020 data is being collected. For homeowners insurance, 2021 data is also being collected. The aim for homeowners is to produce two reports, with the 2020 data being released in late 2022 and the 2021 data being released in the early months of 2023. For auto insurance, similar timelines are planned. The Working Group will hear from Louisiana during the June 15 meeting on suggested revisions to future statistical reports.

The Working Group is collecting information on state data calls. The goal is to find common data elements that are being requested in these data calls to inform updates to the *Statistical Handbook of Data Available to Insurance Regulators*. The Working Group is trying to determine the type of data state insurance regulators would find useful and include those data elements in the required statistical reporting. Members of the Task Force were asked to send any data call templates to Libby Crews (NAIC). Ms. Crews said there is no need to send information about past data calls when the NAIC was involved.

Ms. Darby made a motion, seconded by Mr. Chou, to adopt the report of the Statistical Data (C) Working Group. The motion passed unanimously.

### 3. Discussed the NAIC LCM Form

In March, Mr. Steinert asked if there was interest in updating the NAIC's loss cost multiplier (LCM) forms. After the Task Force agreed, a Subgroup was formed to create a draft for exposure. The Subgroup's draft was then exposed for a public comment period ending June 7. Parties who submitted written comments (Attachment Two-A) presented their comments orally to the Task Force. Mr. Slavich asked the Subgroup to consider the written comments and discussion to produce a revised draft.

### 4. Discussed the Creation of a Rate Review Support Services Handbook

The NAIC's Rate Model Review Team was created by the Executive (EX) Committee after the former Big Data (EX) Working Group asked the Committee to research the best option to provide rate model assistance to states. Around the same time, the Task Force was asked to write a white paper about the regulatory review of predictive analytics, which was adopted in 2020.

With the completion of its white paper, which contains the recommendations for the priority of information items needed to review a rate filing's risk classification model, NAIC staff were able to begin to implement the NAIC rate model review process based largely on the white paper.

The NAIC Rate Model Review Team has now built processes, a shared model database, and technical review templates based on the white paper, advice from a few state insurance regulators who would use the database, and requests from individual users of the NAIC rate model technical reviews. Those processes and procedures have largely stabilized.

Mr. Slavich said the Blanks (E) Working Group has a handbook for the adoption of blanks changes, the Financial Analysis (E) Working Group has a handbook for its operations, and other NAIC committee activities have handbooks or guidance manuals. He asked the Task Force whether it should create a handbook to explain the NAIC staff technical reviews of models and help guide staff over time. He said such a handbook could contain current processes and procedures for NAIC rate model review; include state insurance regulators' guidance to the NAIC Rate Model Review Team, perhaps including the appendices of the Task Force white paper; explain how regulators want the NAIC to prioritize requests if not on a first come, first served basis; and include a process for requesting changes in the process.

Kris DeFrain (NAIC) said her vision of a handbook would not be technical, such as the development of templates, and would not put restraints on NAIC staff who might need to make changes depending on the individual filing. She said the handbook could document the history and purpose and explain confidentiality. She said most of the other NAIC staff operations have a direct tie to a specific committee that oversees the work.

Mr. Stolyarov said it might be best to keep internal documentation rather than a Task Force handbook.

Mr. Dyke said although the nature of this work is different from other committee/NAIC staff work, there might be some information in other NAIC regulator-only handbooks that could prove useful. He said the handbook for blanks is a public-facing document, and that does not seem appropriate for this case.

Mr. Davis said he sometimes just wants to ask questions to the Rate Model Review Team. He said a compilation of questions and answers (Q&As) might prove useful. He suggested creating an academic resource and Q&A

document for state insurance regulators' access. Ms. DeFrain said there is no Q&A document at this time, but academic resources are available to regulators on I-SITE. Mr. Steinert said a collection of common pitfalls, such as collinearity or things that could go wrong when modeling, would be helpful.

Mr. Slavich said it would be important to record practice and procedures in case of staff turnover. He said he does not believe the Task Force needs to exert control over the process. Ms. Darby agreed, saying that when she used the resource, she had control over how the model was reviewed and what questions were asked for the companies to answer.

The group seemed to agree that the individual state should control the process it uses with NAIC staff. There was no interest in directing the content of the NAIC's technical reports, generally because that was thought to be best handled between the individual state and NAIC staff.

Ms. DeFrain said there are some decisions where it would be helpful to have state agreement. For example, should NAIC staff work on a first come, first served basis? Ms. DeFrain asked if it is fair if one state dominates the staff resource given every state can benefit from every review due to the sharing of information in the shared model database or would regulators prefer a more even distribution of requesting states? Another example is whether there are enough NAIC resources dedicated to the project. Mr. Vigliaturo said determining priorities may be helpful, but he would suggest handling questions on a case-by-case basis. He said things appear to work fine now, and nothing may be needed unless things start to break down. Mr. Dyke said it sounds like NAIC staff may want assistance before things break down.

Mr. Dyke said another approach might be the creation of a steering committee like System for Electronic Rates & Forms Filing (SERFF) uses. That might handle what sounds like a governance role. Ms. DeFrain said she would gather information about whether a steering committee could be created.

##### 5. Discussed the Insurance Summit's Predictive Analytics Session

The Insurance Summit will be a hybrid event hosted in Kansas City, MO, at the Loews Hotel and Convention Center. Mr. Slavich asked what content would be beneficial in a half-day predictive analytics regulator-to-regulator session on Friday, Sept. 23 all via webinar. Given the time, he asked members to send suggestions to Ms. DeFrain.

Having no further business, the Casualty Actuarial and Statistical (C) Task Force adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/06-14-22/06-14 min.docx

**NAIC LOSS COST FILING DOCUMENT**  
**CALCULATION OF COMPANY LOSS COST MULTIPLIER**

Attachment Two-A  
 Casualty Actuarial and Statistical (C) Task Force  
 8/10/22

(EFFECTIVE \_\_\_\_)

Company Name	
NAIC Company Code	
Line, subline, coverage, territory, etc. combination to which this form applies	
Does this form apply to all class codes? (Yes/No) (If no, list class codes in a specifically identified attachment.)	
SERFF Filing # <b>[NEEDED? Group Code needed too?]</b>	
Loss Cost Reference Filing (Advisory Org. and Reference Filing #)	
Expense constant(s) (0 if no expense constant is used): Justify any expense constant(s) in a specifically identified attachment.)	

**1. Declaration**

The above insurer hereby declares that it is a member, subscriber, or service purchaser of the named advisory organization for this line of insurance and is filing the prospective loss costs in the captioned Loss Cost Reference Filing. The insurer's rates will be the combination of the prospective loss costs and the loss cost multipliers and, if utilized, the expense constants.

**2. Rule of Application (Check One)**

- Current and future loss cost reference filings** -- The insurer hereby files to have its loss cost multipliers and, if utilized, expense constants be applicable to future revisions of the advisory organization's prospective loss costs for this line of insurance. The insurer's rates will be the combination of the advisory organization's prospective loss costs and the insurer's loss cost multipliers and if utilized, expense constants. The rates will apply to policies written on or after the effective date of the advisory organization's prospective loss costs. This authorization is effective until disapproved by the Commissioner, or until amended or withdrawn by the insurer. **Note: Some states prohibit this option.**
- Current loss cost reference filing only** -- The insurer hereby files to have its loss cost multipliers and, if utilized, expense constants be applicable only to the above Loss Cost Reference Filing.

**3. Loss Cost Modification/Deviation**

(See examples below. Attach supporting data and/or rationale for the modification(s) in a specifically identified attachment.)

Loss Cost Modification Factor Examples:

- If the loss cost modification is 0%, the Loss Cost Modification Factor is 1.00.
- If your company's loss cost modification is -10%, the Loss Cost Modification factor is 0.900. The calculation is (1.000 - 0.100).
- If your company's loss cost modification is +15%, the Loss Cost Modification Factor is 1.150. The calculation is (1.000 + 0.150).

	Current	Proposed	% Change
Loss Cost Modification Factor			

**NAIC LOSS COST FILING DOCUMENT**

**4. Expense Provisions**

Projected expenses should be relative to charged premium (for non-workers' compensation lines) and standard premium (for workers' compensation) using the company's rates in effect. (Attach exhibit detailing insurer expense data, impact of premium discount plans, and/or other supporting information in a specifically identified attachment.)

		Current	Proposed
A.	Commission and Brokerage	%	%
B.	Other Acquisition	%	%
C.	General Expenses	%	%
D.	Taxes, Licenses & Fees	%	%
E.	Underwriting Profit & Contingencies (before investment income offset)	%	%
F.	Investment Income Offset (show as negative value)	%	%
G.	Average Premium Discount (i.e., for workers' compensation)	%	%
H.	Other 1 (If used, explain in Section 9.)	%	%
I.	Other 2 (If used, explain in Section 9.)	%	%
J.	Total (sum A through I)	%	%

**5. Calculation of Permissible Loss (and Loss Adjustment Expense) Ratio**

		Current	Proposed
A.	Permissible Loss Ratio: $PLR = 100.0\% - 4J =$	%	%
B.	PLR in Decimal Form =		

**6. Additional Adjustments**

(Use 1.000 where not applicable.)

		Current	Proposed
A.	Loading Factor Relative to Loss (for states where LAE is not included in loss costs)		
B.	Overall Impact of Expense Constant and Minimum Premiums (e.g., a 2.3% impact would be expressed as 1.023)		

**7. Calculation and Selection of Loss Cost Multiplier**

(Explain any differences, other than rounding, between 7A and 7B in Section 9.)

		Current	Proposed
A.	Company Formula Loss Cost Multiplier [(3 x 6A) / (5B x 6B)]		
B.	Company Selected Loss Cost Multiplier		

**NAIC LOSS COST FILING DOCUMENT**

**8. Percent Change (from Current to Proposed)**

		% Change
A.	Percent Change in Loss Cost Multiplier (7B proposed / 7B Current – 1.000)	%
B.	Percent Change in Loss Costs (Weighted on company’s own book of business and not the advisory organization unless company has zero premium volume)	%
C.	Percent Change in Other Rating Items (as identified in Section 9.)	%
D.	Total Percent Change [(1.000 + 8A) x (1.000 + 8B) x (1.000 + 8C) - 1.000]	%

**9. Additional Comments**

(If needed, attach a specifically identified attachment.)

[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022\\_Summer/CASTF/06-14-22/LCM%20Proposal\\_subgroup%20042022.doc](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022_Summer/CASTF/06-14-22/LCM%20Proposal_subgroup%20042022.doc)

6/7/22

Attachment Two-A  
Casualty Actuarial and Statistical (C) Task Force  
8/10/22

Comments about LCM Form Exposure

## CONNECTICUT

**From:** Gozzo Andrews, Susan <Susan.Gozzo.Andrews@ct.gov>  
**Sent:** Wednesday, May 4, 2022 1:44 PM  
**To:** Steinert, Larry <lsteinert@idoi.in.gov>; DeFrain, Kris <kdefrain@naic.org>  
**Cc:** Thomas, Lia C <LCThomas@naic.org>  
**Subject:** RE: CASTF - LCM Form Exposed for Comment

Thanks for the response Larry:  
I didn't have you email. Now I do.

To clarify where I was going with that is I want to see three things always:  
% in expense component of LCM  
% change in loss modification  
% total LCM change

This is what I ask for in our actuarial checklist in CT.

- e. Support for changing LCM (Lost Cost Multipliers) – Exhibits and narrative has been provided to explain the LCM changes.

**Note 1:** For LCMs, this includes a comparison of the proposed and current LCMs total broken out by the "pure LCM" component based on filed expense ratios ( $1 / (1 - \text{total expense ratio})$ ), and the "loss modification factors".

**Note 2:** Exhibits need to show % changes for the proposed LCMs in total and separately by "pure LCM" and "loss modification" components.

---

Seems to me that a lot of companies, maybe do to changing staffs, are changing LCM in total to get to their rate change, but are not connecting the total LCM change to the two components.

I also think sending out updated instructions with the new form is going to help us re-educate industry on this.

Thanks for all your good work on this.

Regards,

*Sue*

Susan Gozzo Andrews, FCAS, MAAA, PIR, RPLU+  
Property Casualty Actuary  
State of Connecticut Insurance Department

**From:** Steinert, Lawrence <LSteinert@idoi.IN.gov>  
**Sent:** Wednesday, May 4, 2022 8:12 AM  
**To:** DeFrain, Kris <kdefrain@naic.org>; Gozzo Andrews, Susan <Susan.Gozzo.Andrews@ct.gov>  
**Cc:** Thomas, Lia C <LCThomas@naic.org>  
**Subject:** RE: CASTF - LCM Form Exposed for Comment

6/7/22

Attachment Two-A  
Casualty Actuarial and Statistical (C) Task Force  
8/10/22

Comments about LCM Form Exposure

Thanks Susan. I would assume sections 4 and 5 would show additive differences, while sections 6 and 7 would show multiplicative differences (% change)?

The drafting group did consider this, but we may have felt that if the total expenses were changing significantly, it would be easy to “eyeball” the current and proposed columns and see what is changing most significantly. But we can certainly ask the larger group what they think.

Thanks again !

Lawrence Steinert, FCAS  
Actuary  
Indiana Dept. of Insurance  
317-234-6622

On May 2, 2022, at 3:20 PM, Gozzo Andrews, Susan <[Susan.Gozzo.Andrews@ct.gov](mailto:Susan.Gozzo.Andrews@ct.gov)> wrote:

I would like to comment that it would be helpful if we could add a column of “% change” to all the elements in Page 2.

With the loss modification percentage change on bottom of Page 1, and all the expense components on Page 2, adding % change to those elements on Page 2 would be helpful to regulators to see major changes in expenses.

And it would also help us understand what is driving the “LCM change” in 8A; that is the Loss Modification Factor or Expense change (and which expense change even).

Thank you for considering this.

*Sue*

[Susan Gozzo Andrews, FCAS, MAAA, PIR, RPLU+](#)  
[Property Casualty Actuary](#)  
[State of Connecticut Insurance Department](#)

## HAWAII

June 7, 2022

To: Erich Slavich, CASTF Chair  
CC: Kris DeFrain, Lia Thomas, and Larry Steinert

From: Randy Jacobson, FCAS, MAAA  
Member of the CASTF (representing Hawaii)

Subject: NAIC LCM Form – Exposed on 4/28/2022 for Comments

6/7/22

Attachment Two-A  
Casualty Actuarial and Statistical (C) Task Force  
8/10/22

Comments about LCM Form Exposure

Dear Mr. Slavich:

Thank you for providing me this opportunity to review and comment on the proposed LCM form. After a thorough review of it, I believe it does reflect some good improvements compared to the currently published forms. With that said, I have just one comment (which I title as “*Investment Income Offset Line Item*”):

**Investment Income Offset Line Item,**

It is suggested that Line 4.F. “*Investment Income Offset ...*”, should be removed. It is confusing and awkward to put this item on the form. It appears to imply that an investment income offset is a result of the profit & contingency load. That is materially untrue. (Investment income comes from the investment return that is expected to be earned on the “float” -- i.e. unearned premium reserve & loss and LAE reserves – as well as earmarked surplus).

With this removal, Line 4.E. would appropriately be modified. If it is wished, Line 4.E. could be asterisked to still say, “*Show corresponding support, including how investment income is accounted for.*”

Once again, thank you for the opportunity to provide comments in this matter.

Very truly yours,

*Randy Jacobson*

Randy Jacobson, FCAS, MAAA  
Property & Casualty Actuary  
State of Hawaii Insurance Division  
Department of Commerce & Consumer Affairs  
335 Merchant Street, Room 213  
Honolulu, HI 96813  
Phone: 808-587-6744  
Email: rjacobso@dcca.hawaii.gov

## MICHIGAN

**From:** Nancy, Tina (DIFS) <[NacyT@michigan.gov](mailto:NacyT@michigan.gov)>  
**Sent:** Friday, May 27, 2022 2:46 PM  
**To:** Robben, Sara <[srobben@naic.org](mailto:srobben@naic.org)>  
**Subject:** RE: Comment Period for CASTF Proposed NAIC Loss Cost Multiplier Form

Sara,

One suggestion I had on this would be to make the file an Excel file that has certain fields locked down. That way, you could include some formulas that calculate certain fields and the company would just

## Comments about LCM Form Exposure

input the initial data. We did that on the version we have. Please see the LCM and Expense Constant tab of the attached. This was created by one of our actuarial resources.

Regards,

Tina M. Nancy, AINS - State Administrative Manager  
**Office of Insurance Rate and Forms - Property & Casualty Section**  
 Michigan Department of Insurance and Financial Services  
 530 W. Allegan St, 7th fl, Lansing, MI 48933  
 Phone: 517-284-8709  
 Email: [nacyt@michigan.gov](mailto:nacyt@michigan.gov)

## MISSOURI

**From:** Lederer, Julie <Julie.Lederer@insurance.mo.gov>  
**Sent:** Monday, June 6, 2022 4:50 PM  
**To:** DeFrain, Kris <kdefrain@naic.org>  
**Cc:** LeDuc, Jo <Jo.LeDuc@insurance.mo.gov>; Lennon, Patrick <Patrick.lennon@insurance.mo.gov>  
**Subject:** RE: CASTF - LCM Form Exposed for Comment

Dear Kris,

Thank you to Larry Steinert and the other volunteer drafters for their work on the LCM form. Missouri appreciates their efforts to modernize the form and combine the two forms into one. We would like to offer a few minor comments.

### Minor comments for consideration as we finalize the form:

1. Regarding the line "Does this form apply to all class codes? (Yes/No)": We might consider adding an "N/A" to the options for non-WC filings.
1. Regarding the question on whether the SERFF tracking number should be requested: If the company submits this form with the filing, the company may not know the SERFF number.

**Other comments from Missouri** – *these likely would **not** lead to any changes in the proposed form, but we were hoping to discuss them on a CASTF call and get insights from others:*

1. Regarding line 4.G ("Average Premium Discount"): It appears that premium discount may be treated differently in the proposed form versus the current WC form. The different treatment appears to lead to a change in the calculated LCM.
  - In the current WC form, the average discount is expressed as a factor in line 7, and this factor is in the denominator of the LCM calculation in line 8. In the proposed form, the average discount is treated as a component of the expenses and therefore factors into the PLR calculation in line 5.

## Comments about LCM Form Exposure

- Suppose that the expenses (excluding average discount) are 30%, the average discount is 10%, and the loss cost modification as a factor is 1.00.
    - Then, using the current form, the LCM would be  $1.00/(0.70*0.90) = 1.587$ .
    - Using the proposed form, the LCM would be  $1.00/(0.60) = 1.667$ .
1. Regarding line 6.B (“Overall Impact of Expense Constant and Minimum Premiums”): This line item is unchanged from the current WC form but we are wondering how companies generally calculate this item.

Missouri appreciates the opportunity to comment.  
Sincerely,

Julie Lederer, FCAS, MAAA  
Property and Casualty Actuary  
Missouri Department of Commerce and Insurance  
816-889-2219  
[Julie.Lederer@insurance.mo.gov](mailto:Julie.Lederer@insurance.mo.gov)

## WEST VIRGINIA

**From:** Wimmer, Juanita D <juanita.d.wimmer@wv.gov>  
**Sent:** Thursday, May 19, 2022 3:17 PM  
**To:** Thomas, Lia C <LCThomas@naic.org>  
**Cc:** DeFrain, Kris <kdefrain@naic.org>; Gillespie, Tonya <tonya.l.gillespie@wv.gov>  
**Subject:** Re: CASTF - LCM Form Exposed for Comment

Good afternoon,

I'm providing comments for West Virginia on the proposed LCM form.

West Virginia could use the proposed NAIC form for all lines but would like to add the following additions or comments:

- We would like to keep the SERFF Filing number.
- We would like to add the NAIC group code.
- Can we include a field for the policy count affected by the proposed LCM?
- Can we include the effective date for the current factor so we know how long it's been in effect or when it was last changed?
- The check boxes in Question 2: Rule of Application didn't work properly for my test and may need revision before distribution or utilization.
- The current West Virginia loss cost form is located here: <https://www.wvinsurance.gov/Portals/4/pdf/rates/PCA-LCR-2009.pdf>

and the current West Virginia Rate Filing abstract is located here: <https://www.wvinsurance.gov/Portals/4/pdf/rates/PCA-R-2009.pdf>

I appreciate the chance to provide feedback and would like to express my thanks to the volunteer drafters for their work on the form.

Please let me know if you have any questions or need additional information.

Thanks,

**Juanita Wimmer**  
**Actuarial Analyst**  
West Virginia Offices of the Insurance Commissioner  
900 Pennsylvania Ave., 9th Floor  
Charleston, WV 25302  
[Juanita.D.Wimmer@wv.gov](mailto:Juanita.D.Wimmer@wv.gov)  
Phone: 304-414-8491  
Fax: 304-558-0412

## AMERICAN ASSOCIATION OF INSURANCE SERVICES (AAIS)

**From:** Lori Dreaver Munn <lorim@AAISonline.com>  
**Sent:** Tuesday, May 24, 2022 3:45 PM  
**To:** Thomas, Lia C <LCThomas@naic.org>; DeFrain, Kris <kdefrain@naic.org>  
**Cc:** Michael Payne <michaelpa@AAISonline.com>; Robin Westcott <robinw@AAISonline.com>  
**Subject:** RE: CASTF - LCM Form Exposed for Comment

Good afternoon Kris and Lia,

On behalf of the American Association of Insurance Services (AAIS) we offer the following comments in response to your email dated April 28, 2022. Overall, we feel that the additional level of detail and explanations are great improvements. We have provided feedback specific to each section as indicated below:

### General

- SERFF Filing # would be good information to capture.

### 3. Loss Cost Modification/Deviation

- For complete transparency, it may be helpful to add another row in this section for the factors instead of using the descriptive examples.

	Current	Proposed	% Change
A. Loss Cost Modification	%	%	N/A
A. Loss Cost Modification Factor (1.000 + 3A)			

## Comments about LCM Form Exposure

- A formula for the % Change field may also be beneficial.

## 5. Calculation of Permissible Loss (and Loss Adjustment Expense) Ratio

- In item 5.B., consideration should be given to displaying the formula, perhaps replacing the '=' with  $[5A/100]$  following the word Form. For consistency, perhaps in item 5.A., the '=' following 4J should be removed as well. (Items 7 and 8 do not display '=' in their formulas).

## 8. Percent Change (from Current to Proposed)

- In item 8.A., for clarity of the formula, it may be more meaningful to move the ')' after the word 'Current' or to use a display similar to 8.D, like  $[(7B \text{ Prop} / 7B \text{ Curr}) - 1.000]$

## Expense Constant

- Are there instructions for how to calculate an expense constant, if applicable?

Please do not hesitate to contact me if you have any questions or if you require any additional information.

Thank you.

Lori

**Lori Dreaver Munn**

Manager, Government Relations and Compliance

 **Email:** [lorim@AAISonline.com](mailto:lorim@AAISonline.com)

 **Tel:** [800.564.AAIS](tel:800.564.AAIS) x3276 | 630-457-3276

 **Web:** [www.aaisonline.com](http://www.aaisonline.com)

**ORIGINAL EXPOSURE EMAIL**

On Thu, Apr 28, 2022 at 2:03 PM Thomas, Lia C <[LCThomas@naic.org](mailto:LCThomas@naic.org)> wrote:

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

Eric Slavich, CASTF Chair, exposes the attached proposed NAIC loss cost multiplier form for a 40-day comment period, with comments due by **Tuesday, June 7**. The volunteer drafters, led by Larry Steinert, propose this revised form could apply to any line of business (so there would no longer be a separate form for work comp). There is also an aim for this form to meet the needs of almost all states, if possible, whether LAE is included in loss costs or not. If your state has a special need that could be met with a proposed NAIC form, please submit comments. The

Comments about LCM Form Exposure

current forms are located at this

link: [https://content.naic.org/industry\\_rates\\_forms\\_loss\\_cost.htm](https://content.naic.org/industry_rates_forms_loss_cost.htm)

The plan is to be able to discuss comments received on the June 14 CASTF conference call.

**Kris DeFrain**

Director

Research & Actuarial Services



**O:** 816-783-8229

**M:** 816-392-6140

**W:** [www.naic.org](http://www.naic.org)

Follow the NAIC on



[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member Meetings/C CMTE/2022\\_Summer/CASTF/06-14-22/LCM Comments 060722.docx](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20MTE/2022_Summer/CASTF/06-14-22/LCM%20Comments%20060722.docx)

Draft: 8/9/22

Actuarial Opinion (C) Working Group  
Virtual Meeting  
August 2, 2022

The Actuarial Opinion (C) Working Group of the Casualty Actuarial and Statistical (C) Task Force met Aug. 2, 2022. The following Working Group members participated: Anna Krylova, Chair (NM); Miriam Fisk, Vice Chair (TX); Qing He (CT); David Christhlf (DC); Judy Mottar (IL); Sandra Darby (ME); Julie Lederer (MO); Gordon Hay (NE); Tom Botsko (OH); Andrew Schallhorn (OK); and Kevin Clark, James DiSanto, and Jeffery Smith (PA).

1. Discussed a Financial Analysis (E) Working Group Referral on Predictive Analytics in Reserve Setting

The Working Group received a referral from the Financial Analysis (E) Working Group (Attachment Three-A) asking for discussion of the use of predictive analytics in reserve setting and consideration of drafting guidance.

Ms. Lederer said it might be helpful to provide financial examiners with questions they might want to ask about any type of model. The Working Group discussed creating questions general in nature; using Actuarial Standard of Practice (ASOP) No. 56, Modeling; paring down questions for rating models in the *Regulatory Review of Predictive Models* white paper; investigating whether models are being used for case reserves, not just incurred but not reported claim reserves; separating out questions based on types of reserves being established; and adding subsections specific to the type of model in addition to the general questions.

2. Discussed Potential Changes to the Qualification Documentation Requirements and Disclosures

Ms. Krylova said the Working Group met June 3 and May 26 in regulator-to-regulator session, pursuant to paragraph 3 (specific companies, entities or individuals) of the NAIC Policy Statement on Open Meetings, to discuss the individual companies' Statements of Actuarial Opinion (SAOs). The Working Group expressed agreement to reconsider qualification documentation requirements. One idea is to only require qualification documentation every five years or if the company has significant changes. Ms. Lederer said the risk-based capital (RBC) instructions contain a list of requirements for filing the catastrophe risk charge documentation that may be useful as sample language to adjust for the qualification documentation. Interested parties said there would likely be numerous issues that would need to be detailed in the instructions if they are changed.

3. Discussed Proposed Changes to Regulatory Guidance and Annual Statement Instructions

State insurance regulators noticed that COVID-19 exposure was not always mentioned in 2021 actuarial opinions. Ms. Krylova suggested that the Working Group discuss whether COVID-19 exposure should continue to be in the opinions or whether it may be appropriate to eliminate COVID-19 disclosure in the opinion at some point.

Ms. Krylova said the Working Group previously discussed which ASOPs to reference in the instructions or whether to refer to the actuarial matrix. She said some potential wording would be as follows: "Although it is the responsibility of the Appointed Actuary to identify the ASOPs applicable to the Actuarial Opinion, Actuarial Opinion Summary, and the Actuarial Report, the Appointed Actuary may find it useful to review the Applicability Guidelines for Actuarial Standards of Practice published by the Actuarial Standards Board (ASB)."

Ms. Krylova asked whether the applicable ASOPs should be disclosed in the Actuarial Report. Working Group members did not identify significant value in having such a list.

Ms. Krylova said the Working Group previously discussed defining “conclusions” and “actuary’s conclusions” and some additional proposals regarding Section 7. With limited discussion on the call, Ms. Krylova said the changes could be discussed on the next call.

Having no further business, the Actuarial Opinion (C) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/AOWG/AOWG 080222 min.docx

---

**MEMORANDUM**

TO: Anna Krylova, Chair, Actuarial Opinion (C) Working Group

FROM: Judy Weaver, Chair, Financial Analysis (E) Working Group

DATE: May 9, 2022

RE: Enhanced Regulatory Guidance

---

As you may be aware, the Financial Analysis (E) Working Group (FAWG) meets annually in Kansas City to discuss among other things, potentially troubled insurers and insurance groups. During this meeting, FAWG also discusses issues and industry trends, including identifying any that are potentially adverse or might warrant communication and coordination with other NAIC groups. As a result of the issues and trends discussed, FAWG would like to refer the following items to the attention of your group.

1. **Use of Predictive Analytics in Reserve Setting** – In conducting “post-mortem” analysis of a recently failed insurer, we discussed whether the insurer's use of "predictive analytical models" could have been a contributing factor. We would like you to a) discuss the use of predictive analytical models in setting reserves and b) recommend whether additional guidance should be offered to regulators who are reviewing predictive models in reserving.

In considering these issues, FAWG recommends consideration of additional guidance for regulatory actuaries and/or financial analysts/examiners to ensure these concerns are adequately addressed. If necessary and appropriate, the Working Group should consider working with the Financial Analysis Solvency Tools (E) Working Group and/or the Financial Examiners Handbook (E) Technical Group to address. If there are any questions regarding the proposed recommendation, please contact me or NAIC staff (Bruce Jensen at [bjensen@naic.org](mailto:bjensen@naic.org)) for clarification.

Thank you for your consideration.

[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member Meetings/C CMTE/2022\\_Summer/CASTF/AOWG/2022 FAWG Referral to AOWG.docx](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022_Summer/CASTF/AOWG/2022%20FAWG%20Referral%20to%20AOWG.docx)

---

EXECUTIVE OFFICE • 444 North Capitol Street NW, Suite 700 • Washington, DC 20001-1509

p | 202 471 3990 f | 816 460 7493

CENTRAL OFFICE • 1100 Walnut Street, Suite 1500 • Kansas City, MO 64106-2197

p | 816 842 3600 f | 816 783 8175

CAPITAL MARKETS & INVESTMENT ANALYSIS OFFICE • One New York Plaza, Suite 4210 • New York, NY 10004

p | 212 398 9000 f | 212 382 4207

---

[www.naic.org](http://www.naic.org)

Draft: 8/4/22

Statistical Data (C) Working Group  
E-Vote  
August 4, 2022

The Statistical Data (C) Working Group of the Casualty Actuarial and Statistical (C) Task Force conducted an e-vote that concluded August 4, 2022. The following Working Group members participated: Sandra Darby, Chair (ME); Wanchin Chou, Vice Chair (CT); David Christhilf (DC); Arthur Schwartz (LA); Alexander Vajda (NY); Tom Botsko (OH); Landon Hubbard (OK); and David Dahl (OR).

1. Adopted its July 20 Minutes

The Working Group met July 20 to discuss proposed changes to the *Competition Database Report* (Competition Report) and the *Dwelling, Fire, Homeowners Owner-Occupied, and Homeowners Tenant and Condominium/Cooperative Unit Owner's Insurance Report* (Homeowners Report).

A majority of the Working Group members voted in favor of adopting the July 20 meeting minutes. The motion passed.

Having no further business, the Statistical Data (C) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/SDWG/7-20-22 /StatDataWGevote\_0804.docx

Draft: 8/1/22

Statistical Data (C) Working Group  
Virtual Meeting  
July 20, 2022

The Statistical Data (C) Working Group of the Casualty Actuarial and Statistical (C) Task Force met July 20, 2022. The following Working Group members participated: Sandra Darby, Chair (ME); Wanchin Chou, Vice Chair, and Qing He (CT); David Christhif (DC); Arthur Schwartz (LA); Cynthia Amann and Brent Kabler (MO); Christian Citarella (NH); Alexander Vajda (NY); Tom Botsko (OH); Landon Hubbart (OK); David Dahl and Ying Liu (OR); and Brian Ryder and Ken Burton (TX). Also participating were: David Dombrowski (MT); Martin Swanson (NE); and Mike Andring (ND).

1. Adopted its June 15 Minutes

The Working Group met June 15 to: 1) adopt an accelerated timeline for auto premium and exposure data collection; and 2) discuss proposed changes to the *Competition Database Report* (Competition Report) and the *Dwelling, Fire, Homeowners Owner-Occupied, and Homeowners Tenant and Condominium/Cooperative Unit Owner's Insurance Report* (Homeowners Report).

Mr. Botsko made a motion, seconded by Ms. Amann, to adopt the Working Group's June 15 minutes. The motion passed unanimously.

2. Discussed Proposed Changes to NAIC Statistical Reports

Ms. Darby said the Working Group discussed the proposed changes to the Competition Report and the Homeowners Report during its June 15 meeting. She said the Working Group will discuss proposed changes to the *Auto Insurance Database Report* (Auto Report) and the *Report on Profitability By Line By State* (Profitability Report).

Mr. Schwartz said in the Profitability Report, he would like to see a breakout of the rate of return by different types of insurers, including mutual reciprocals, risk retention groups (RRGs), surplus lines, and stock insurers. He said this would be valuable for state insurance regulators when looking at rate filings and company reserves. Birny Birnbaum (Center for Economic Justice—CEJ) agreed that it is useful to look at the profitability of investor-owned insurers versus policyholder-owned insurers because the profitability of the industry is distorted in the aggregate. He said different types of companies have different goals for their rate of return. He said he does not see the need for a breakout of surplus lines insurers. Ms. Darby said she would like NAIC staff to see if Mr. Schwartz's suggested changes can be incorporated into the report.

Mr. Schwartz said the Auto Report is useful and often quoted in the media. He said he would like to improve on the report with additional useful information. He said the report currently shows the average premium by state, but he would like the report to show the average premium by metropolitan statistical area (MSA). He said this would be useful in comparing states and areas with similar populations. Ms. Darby asked if MSA or ZIP code data would be better to include. Mr. Schwartz said he wants to be able to drill down into premiums in heavily populated areas versus less populated areas, but he does not have a preference whether that data is collected by ZIP code or MSA. Mr. Citarella said MSAs can cross state lines, so they may not directly correlate to state totals. Mr. Birnbaum said insurance companies would have to convert existing data into geographic areas that correlate to the MSA. He said companies would more easily be able to provide ZIP code data.

Mr. Schwartz said he would like the Auto Report to include additional information such as the average age of private passenger autos (PPAs) on the road, median value of cars on the road, median liability policy limit, median per capita disposable income, population, number of insured drivers, number of registered personal autos, number of earned exposures, minimum liability limits, number of accidents, number of DUI arrests, and accident count and frequency with autonomous emergency braking (AEB) systems. He said some of this information can be collected from insurers, and other information would need to come from external sources such as government databases. He said adding these metrics would allow the report to become a warehouse of extremely useful data.

Ms. Darby asked if the report should include data on other driver assistance systems beyond the AEB system data. Mr. Schwartz said the AEB data would be most important to collect, but the other data would also be useful.

Mr. Chou said the Working Group should weigh the costs and benefits of adding additional information to the report. He said it should consider the cost of adding data that may delay the publication of the report. He said the Working Group should also consider who the customer is and what benefits this additional data would provide them.

Mr. Citarella said some of the external data would be hard to obtain, such as number of licensed drivers and number of crashes.

Mr. Birnbaum said the Working Group needs to consider how the Auto Report is used and what kind of data state insurance regulators and other users would find useful. He said the report should be a high-level summary of premiums and exposures across states. He said the report should only include insurance-relevant information. He said to get a more granular level of detail, statistical reporting would have to be required on a transactional level. He said some of the additional data elements that are being discussed for this report should be included in a re-engineered statistical plan.

Tony Cotto (National Association of Mutual Insurance Companies—NAMIC) said NAMIC is committed to working with state insurance regulators to provide the data that is needed. He said many of the data found in government databases are estimates that are compiled from state agencies. He encouraged the Working Group to look at information from the Insurance Institute of Highway Safety (IIHS), which may be helpful for its review of auto insurance data. He agreed that the purpose of this Auto Report is to be a high-level public consumption summary.

Mr. Schwartz said he would like the Working Group to consider changing the names of certain statistical reports so that all reports coming from this Working Group have a consistent naming structure. He said, for example, the Homeowners Report should be titled the Homeowners Insurance Database. He said this is more consistent with the name of the Auto Report and allows people to easily search for the report.

Ms. Darby said NAIC staff will summarize the proposed changes and that the Working Group would continue to consider these changes in future meetings.

Having no further business, the Statistical Data (C) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/SDWG/7-20-22 /StatDataWGmin\_720.docx

Draft: 7/18/22

Statistical Data (C) Working Group  
Virtual Meeting  
June 15, 2022

The Statistical Data (C) Working Group of the Casualty Actuarial and Statistical (C) Task Force met June 15, 2022. The following Working Group members participated: Sandra Darby, Chair (ME); Wanchin Chou, Vice Chair, George Bradner, and Qing He (CT); David Christhlf (DC); Arthur Schwartz (LA); Cynthia Amann and Jo LeDuc (MO); Christian Citarella (NH); Alexander Vajda (NY); Andrew Schallhorn and Landon Hubbart (OK); Ying Liu (OR); and Brian Ryder (TX). Also participating were: Luciano Gobbo (CA); Tate Flott (KS); Mari Kindberg and David Dombrowski (MT); and Mike Andring (ND).

1. Adopted its May 18 Minutes

The Working Group met May 18 to hear a presentation from the Center for Economic Justice (CEJ) on statistical data collection.

Ms. Amann made a motion, seconded by Mr. Vajda, to adopt the Working Group's May 18 minutes. The motion passed unanimously.

2. Adopted an Accelerated Timeline for Auto Premium and Exposure Data Collection

Ms. Darby said discussion with statistical agents has found that most submitting statistical agents could provide the written premium and exposure data for auto insurance by November following the end of the data year. She said other statistical agents indicated that they would be able to adjust their timelines to the Working Group's requested date. She said Texas data would be available toward the end of November following the end of the data year. Ms. Darby said NAIC staff would need about eight weeks to check the submitted data and compile the full report. She said that after review from this Working Group and the Casualty Actuarial and Statistical (C) Task Force, the publication would be available for release about three months into the following year.

Ms. Darby said the data is currently provided about 18 months after the end the data period. She said the accelerated timeline would speed up the publishing of the *Auto Insurance Database Report* (Auto Report) to about nine months sooner.

Mr. Chou made a motion, seconded by Ms. Amann, to move up the reporting timeline for auto premium and exposure data to Dec. 1 following the end of the data year.

3. Discussed Proposed Changes to NAIC Statistical Reports

Mr. Schwartz said he has proposed changes to the statistical reports that would include data that would be more useful than what is currently in the reports. He said he is open to these additional elements being published after the average premium reports, as the average premium is much more important to have in a timely manner.

Mr. Schwartz said the *Competition Database Report* (Competition Report) should show the breakout of the market share by different types of insurance companies. He said whether a company is a stock, mutual, risk retention group (RRG), or surplus carrier would affect how it is viewed in an analysis of market competition. Mr. Vajda asked where residual market insurers would fit in to the new proposed columns. Jennifer Gardner (NAIC) said the report does currently include RRGs and surplus lines carriers. She said residual markets are not included because most

of them do not report on the NAIC Annual Statement. Mr. Schwartz said he would still like to see additional columns for breaking out stock and mutual companies. Ms. Gardner said those columns replace the columns that currently show the last five-year average.

Mr. Bradner agreed that he would like to see a breakout of residual markets in this report. He said that states could require the residual markets to report their data for this report. Ms. Gardner said she would have to research which residual markets might already report on the NAIC annual statement, but the Working Group could look at collecting the data separate from the annual statement and add it into the Competition Report.

Birny Birnbaum (CEJ) said the Property Insurance Plans Service Office (PIPSO) and AIPSO collect the residual market information. Mr. Bradner said not all state residual markets may provide that data currently. He said if PIPSO and AIPSO could provide that data, it would be a simple way to add the data into the report. He said there is an issue that there are different kinds of residual markets. He said some are assigned risk plans where the risk is assigned to different companies. He said in that case, companies would need to break out their voluntary experience from their assigned risk experience. He said having the residual market data is a good indicator of disruption in the market if a residual market is growing in a certain state. Ms. Gardner said the residual market should be an appendix to the current report in order to not skew the number of insurers shown in the report. Mr. Citarella said states would likely see indicators of an increase in residual market growth before this data would even be available. Mr. Bradner agreed and said this data would just be in addition to other data the department would receive, such as consumer complains. Ms. LeDuc said Missouri receives updates from its residual markets on at least an annual basis, if not quarterly.

Mr. Schwartz said in the *Dwelling Fire, Homeowners Owner-Occupied, and Homeowners Tenant and Condominium/Cooperative United Owners Insurance* (Homeowners Report), he would like to see a chart of average premium by state. He said these premiums would need to be tempered by the average policy limit, as well as the number of homes insured in that particular state. He said the Census Bureau can provide the number of owner-occupied homes versus the number of rental homes, as well as the number of homes insured under a Fair Access to Insurance Requirements Plan (FAIR Plan). He said he would also like to include flood insurance data, including the number of homes insured in a flood zone and how many homes are insured via the National Flood Insurance Program (NFIP) and how many homes are insured by private flood insurance. Mr. Bradner said that the Federal Emergency Management Agency (FEMA) flood data is hard to obtain and that much of the private flood insurance is written by surplus lines carriers, but he would like to be able to see that information included in the report. He said it would be most useful to state insurance regulators to have all of this information available in a Tableau dashboard that each state insurance regulator can use to drill down the data in their state.

Having no further business, the Statistical Data (C) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C Cmte/2022\_Summer/CASTF/SDWG/06-15-22/StatDataWGmin\_615.docx

Draft: 6/7/22

Statistical Data (C) Working Group  
Virtual Meeting  
May 18, 2022

The Statistical Data (C) Working Group of the Casualty Actuarial and Statistical (C) Task Force met May 18, 2022. The following Working Group members participated: Sandra Darby, Chair (ME); Wanchin Chou, Vice Chair, George Bradner, and Amy Waldhauer (CT); Daniel Davis (AL); Arthur Schwartz (LA); Cynthia Amann (MO); Christian Citarella (NH); Ed Scanlon (NJ); Tom Botsko (OH); Andrew Schallhorn (OK); David Dahl and Ying Liu (OR); and Brian Ryder (TX). Also participating were: Luciano Gobbo (CA); Anthony Bredel (IL); Mari Kindberg (MT); Chris Aufenthie and Mike Andring (ND); and Mary Block (VT).

1. Adopted its April 14 Minutes

The Working Group met April 14 to: 1) discuss the timeline of data collection for the *Auto Insurance Database Report* (Auto Report); and 2) discuss updating the *Statistical Handbook of Data Available to Regulators* (Handbook).

Ms. Amann made a motion, seconded by Mr. Botsko, to adopt the Working Group's April 14 minutes. The motion passed unanimously.

2. Heard a Presentation from the Center for Economic Justice on Statistical Data Collection

Birny Birnbaum (Center for Economic Justice—CEJ) said the topic of his presentation is modernizing statistical data reporting for personal lines insurance. He said the availability of the data relating to the Covid-19 pandemic is a good place to begin looking at how data can be collected. He said Workers' Compensation insurance data related to Covid-19 claims, as well as mortgage lending and other financial data on CARES Act funds were made available during the pandemic. He said the data was available on a one to two month lag. Mr. Birnbaum said there was not available data for regulators for personal lines insurance like auto and homeowners.

Mr. Birnbaum said the NAIC Statistical Handbook says responsibilities most relevant to statistical data collection include ensuring rates meet statutory standards and monitoring market structure and performance. He said currently the statistical data system fails to provide timely and relevant data for most Property and Casualty lines of business to assist regulators in carrying out these responsibilities. He said the statistical agency system has not been updated in most states in 40 years and that has led to the NAIC producing Auto and Homeowners Reports three years after the experience period begins.

Mr. Birnbaum said there are three differences in the reporting of Workers' Compensation data and personal lines insurance data: 1) Workers' Compensation data is collected by a single statistical agent in each state while insurers for other property and casualty lines have about four different statistical agents to choose from; 2) Workers' Compensation data is collected at a transaction level; and 3) Workers' Compensation data is collected on a monthly basis which offers a much faster turn around time for data analysis.

Mr. Birnbaum said the solution to modernizing the collection of statistical data is to use existing regulatory authority to update statistical plans, designate a single statistical agent through a competitive bidding process, and establish requirements that the primary duty of the statistical agent is to serve the regulator. He said updated statistical plans should require reporting on at least a quarterly basis.

Mr. Birnbaum said statistical agents that collect transaction level data are in a better position to provide the data in a timely manner and to do better data quality checks. He said moving to a single statistical agent approach would increase the efficiency of data reporting.

Mr. Scanlon said some statistical agents have more relaxed reporting requirements and therefore companies are reporting skeletal data that still satisfies the requirements. He asked if changing the reporting requirements would increase costs. Mr. Birnbaum said that a new statistical plan and reporting system would increase efficiencies and lower cost. He said companies that are reporting in Texas are already doing this kind of data collection and reporting. He said a more robust statistical plan would decrease the amount of special data calls for companies. Mr. Scanlon asked to what extent should stricter reporting requirements be seen as a barrier to entry. Mr. Birnbaum said reporting transaction data should not be difficult for companies because it is essentially a data dump, whereas a summary report would require a program to pull and aggregate the data.

Mr. Chou asked how to go about getting all statistical agents to compile with transaction reporting. Mr. Birnbaum said Commissioner's have the authority to designate statistical plans and even to designate a single statistical agent. He said that process begins with this working group modernizing the statistical reporting system.

Ms. Darby asked if it is possible to request that all statistical agents provide transaction data. Mr. Birnbaum said at least 2 large statistical agents only collect summary data and would need to transition their systems.

Albert Burton (Independent Statistical Service—ISS) said some companies do not have a need for the expansive reporting required by Insurance Services Office (ISO) and that ISS is able to fill their need with less requirements. He said if reporting requirements in the NAIC Statistical Handbook change, then ISS will comply with those changes.

Mr. Schwartz said auto and homeowners data was coming in during the pandemic via fast track reports and regulators were actively looking at those reports. He said the fast track reports could be improved as an alternative to the idea of a new, single statistical agent.

Steve Clarke (ISO) said the way the Workers' Compensation industry developed led to one statistical agent collecting the data for purposes of experience modification. He said ratemaking is the main reason for statistical agents that are also advisory organization to collect robust data. He said many states that have adopted regulations around statistical reporting have language that while the Commissioner appoints statistical agents, the companies get to choose a statistical agent to act on its behalf. He said instead of making the move to a single statistical agent, regulators should decide what data they need and why they need that data.

Ms. Darby said the discussion of specific data elements and what data regulators need will be a topic during the next meeting.

Having no further business, the Statistical Data (C) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C Cmte/2022\_Summer/CASTF/SDWG/05-18-22/StatDataWGmin\_518.docx

Draft: 4/29/22

Statistical Data (C) Working Group  
Virtual Meeting  
April 14, 2022

The Statistical Data (C) Working Group of the Casualty Actuarial and Statistical (C) Task Force met April 14, 2022. The following Working Group members participated: Sandra Darby, Chair (ME); Wanchin Chou, Vice Chair, George Bradner, and Qing He (CT); David Christhilf (DC); Cynthia Amann (MO); Alexander Vajda (NY); Tom Botsko (OH); Landon Hubbart (OK); David Dahl (OR); and Brian Ryder (TX). Also participating were: Luciano Gobbo (CA); Randy Jacobson (HI); Anthony Bredel (IL); Brenda Johnson (KS); Regan Hess (MT); and Michael Muldoon (NE).

1. Adopted its March 10 Minutes

The Working Group met March 10 to: 1) discuss the timeline of data collection for the *Auto Insurance Database Report* (Auto Report); and 2) discuss updating the *Statistical Handbook of Data Available to Regulators* (Handbook).

Ms. Amann made a motion, seconded by Mr. Chou, to adopt the Working Group's March 10 minutes. The motion passed unanimously.

2. Discussed the Timeline of Data Collection for the Auto Report

Ms. Darby said the Working Group has been discussing the possibility of requesting data on a faster timeline for the Auto Report. She said the NAIC sent questions to the submitting statistical agents to get a better sense of each statistical agent's timeline for submitting data for written premium and exposures, earned premium and exposures, and losses.

Ms. Darby said the first question asked was when statistical agents currently submit their data. She said the responses reflect that the data is submitted between March and May. She said the data is submitted on a two-year lag for written premium and a three-year lag for earned premium and losses.

Ms. Darby said the second question asked was when statistical agents could submit written premium and exposure data for the previous year. She said the Insurance Service Office (ISO) and the Commonwealth Automobile Reinsurers (CAR) indicated that they could submit that data in May following the end of the data year. She said the National Independent Statistical Service (NISS), the American Association of Insurance Services (AAIS), and the California Department of Insurance (DOI) indicated that they could submit that data from October to November following the end of the data year. She said the Independent Statistical Services Inc. (ISS) indicated that it would be able to submit the data on the same timeline that it is currently submitted.

Mr. Bradner asked why certain statistical agents are not able to provide the data as quickly and what they would need to do to get the data in faster. Mr. Gobbo said California collects all the data in one data call, and losses are evaluated as of June 30. He said the deadline for data submission from the companies is Sept. 1. Truman Esmond (AAIS) said the AAIS has similar restraints with respect to when the data is collected. He said the AAIS is also required to reconcile the data to the finalized financial statement data. Theresa Szwast (NISS) said companies submit their auto data by June 15 following the end of the data year. She said the NISS would need until October to do data quality checks before sending it to the NAIC.

Birny Birnbaum (Center for Economic Justice—CEJ) asked what is different about the data collections and systems for the statistical agents that can provide the data much sooner than the others. Laura Panesso (ISO) said the ISO collects data on a quarterly basis as opposed to an annual basis. Mr. Birnbaum asked if the quarterly data was reconciled to the financial statement. Ms. Panesso said the ISO performs several data quality checks on the quarterly data, but the reconciliation to the financial statement happens on an annual basis. Mr. Esmond said the AAIS receives quarterly and some monthly data, but the financial statement reconciliation only happens on an annual basis. He said if the financial statement data can be finalized faster, the AAIS would be able to provide the data faster. Mr. Birnbaum asked why the reconciliation of the annual data would take seven months if the quarterly data is reconciled throughout the year. Mr. Esmond said the AAIS would not need seven months. He said if it relies on only quarterly reports, the data could change in the final annual data. He said the AAIS would be able to meet accelerated requirements as it is able to.

Ms. Darby said she does not want to give statistical agents an unreasonable timeline for data submission, but it seems most statistical agents can speed up the data submission.

Mr. Birnbaum said reconciliation to the financial annual statement is not the only tool to check data quality. He said the reconciliation is not a requirement of the Handbook. Susan Chudwick (Travelers Insurance) said if the reconciliation is not required, that would speed up the reporting of the data. Mary Anese (ISO) said with the accelerated timeline, the ISO would not be able to fully reconcile the data to the financial annual statement. She said the ISO has many other data checks that allow them to be confident in the data quality. Mr. Bradner said it makes sense to speed up the timeline and worry about the reconciliation of the data later since there are other data checks in place. Ms. Szwast said the NISS only receives the data annually in June, and it would still need time to do data quality checks before sending the data to the NAIC. She said the June deadline is set to meet reporting requirements for the statistical data that is sent to states.

Ms. Darby asked if the statistical agents would similarly be able to move up the earned premium and earned exposure data. Ms. Panesso said the ISO would be able to move up the earned premium to the same timeline as written premium. Ms. Szwast said the NISS earned premium data is tied to the loss data, and that could be provided in May, two years after the data year.

Mr. Birnbaum said the Working Group should not aim for a timeline of the lowest common denominator. Ms. Darby said the goal is to really explore the limitations of the statistical agent's ability to send data faster, and once that has been uncovered, state insurance regulators can push for a faster timeline that is feasible for the statistical agents.

### 3. Discussed Updating the Handbook

Ms. Darby said during the last meeting, Section 1 of the Handbook was opened for comment. She said before beginning specific updates, the Working Group should dig in to see if and how states are currently using the statistical data. She said the Working Group needs to determine what data is useful to state insurance regulators.

Mr. Dahl said Oregon is looking at data reporting because it has questions from the legislature about sub-markets in Oregon. He said Oregon is looking at developing a specialized data call at the zip code level to answer questions about localized markets.

Mr. Birnbaum said the Handbook says the responsibilities most relevant to statistical collection are to: 1) ensure rates meet statutory standards, meaning they are not inadequate, excessive, or unfairly discriminatory; and 2) monitor market structure and performance, and act if necessary to restore competition or remedy problems. He said states cannot use the statistical reports to determine if companies' rates do not meet statutory standards

because there is no individual company data, and the data is too old to be helpful. Robin Westcott (AAIS) said most states do not have specific requirements for statistical reporting in their legislation. She said the AAIS is looking at what data state insurance regulators need, why and how they use it, and how industry can facilitate getting useful data to them.

Ms. Darby said conversations around the timeline of auto data reporting and updating the Handbook would continue in the next meeting.

Having no further business, the Statistical Data (C) Working Group adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/C CMTE/2022\_Summer/CASTF/SDWG/04-14-22/StatDataWGmin\_414.docx

## APPENDIX B-RF – INFORMATION ELEMENTS AND GUIDANCE FOR A REGULATOR TO MEET BEST PRACTICES’ OBJECTIVES (WHEN REVIEWING RANDOM FORESTS)

This appendix identifies the information a state insurance regulator may need to review a Random Forest predictive model used by an insurer to support a personal automobile or home insurance rating plan. The list is lengthy but not exhaustive. It is not intended to limit the authority of a regulator to request additional information in support of the model or filed rating plan. Nor is every item on the list intended to be a requirement for every filing. However, the items listed should help guide a regulator to sufficient information that helps determine if the rating plan meets state-specific filing and legal requirements. Documentation of the design and operational details of the model will help ensure the business continuity and transparency of the models used. Documentation should be sufficiently detailed and complete to enable a qualified third party to form a sound judgment on the suitability of the model for the intended purpose. The theory, assumptions, methodologies, software, and empirical bases should be explained, as well as the data used in developing and implementing the model. Relevant testing and ongoing performance testing need to be documented. Key model limitations and overrides need to be pointed out so that stakeholders understand the circumstances under which the model does not work effectively. End-user documentation should be provided and key reports using the model results described. Major changes to the model need to be documented and shared with regulators in a timely and appropriate manner. Information technology (IT) controls should be in place, such as a record of versions, change control, and access to the model.<sup>1</sup>

Many information elements listed below are probably confidential, proprietary, or trade secret and should be treated as such, in accordance with state laws and/or regulations. Regulators should be aware of their state laws and/or regulations on confidentiality when requesting data from insurers that may be proprietary or trade secret. For example, some proprietary models may have contractual terms (with the insurer) that prevent disclosure to the public. Without clear necessity, exposing this data to additional dissemination may compromise the model’s protection.<sup>2</sup> Although the list of information is long, the insurer should already have internal documentation on the model for more than half of the information listed. The remaining items on the list require either minimal analysis (approximately 25%) or deeper analysis to generate for a regulator (approximately 25%).

The “Level of Importance to the Regulator’s Review” is a ranking of information a regulator may need to review, which is based on the following level criteria:

Level 1 – This information is necessary to begin the review of a predictive model. These data elements pertain to basic information about the type and structure of the model, the data and variables used, the assumptions made, and the goodness of fit. Ideally, this information would be included in the filing documentation with the initial submission of a filing made based on a predictive model.

Level 2 – This information is necessary to continue the review of all but the most basic models, such as those based only on the filer’s internal data and only including variables that are in the filed rating plan. These data elements provide more detailed information about the model and address questions arising from review of the information in Level 1. Insurers concerned with speed to market may also want to include this information in the filing documentation.

Level 3 – This information is necessary to continue the review of a model where concerns have been raised and not resolved based on review of the information in Level 1 and Level 2. These data elements address even more detailed aspects of the model. This information does not necessarily need to be included with the initial submission, unless specifically requested by a particular state, as it is typically requested only if the reviewer has concerns that the model may not comply with state laws and/or regulations.

Level 4 – This information is necessary to continue the review of a model where concerns have been raised and not resolved based on the information in Level 1, Level 2, and Level 3. This most granular level of detail is addressing the basic building blocks of the model and does not necessarily need to be included by the filer with the initial submission,

---

<sup>1</sup> Bourdeau, M., 2016. “Model Risk Management: An Overview,” The Modeling Platform, Issue 4, December. Accessed online at <https://www.soa.org/globalassets/assets/library/newsletters/the-modeling-platform/2016/december/mp-2016-iss4.pdf>

<sup>2</sup> There are some models that are made public by the vendor and would not result in a hindrance of the model’s protection.

unless specifically requested by a particular state. It is typically requested only if the reviewer has serious concerns that the model may produce rates or rating factors that are excessive, inadequate, and/or unfairly discriminatory.

Appendix B-RF is focused on Random Forest models and should not be referenced in the review of other model types. Random Forest models are a tree-based approach with many significant differences from GLMs. This Appendix B-RF is intended to provide state guidance for the review of rate filings based on Random Forest models.

**A. SELECTING MODEL INPUT**

Section	Information Element	Level of Importance to the Regulator’s Review	Comments
<b>1. Available Data Sources</b>			
A.1.a	Review the details of sources for both insurance and non-insurance data used as input to the model (only need sources for filed input characteristics included in the filed model).	1	<p>Request details of data sources, whether internal to the company or from external sources. For insurance experience (policy or claim), determine whether data is aggregated by calendar, accident, fiscal, or policy year and when it was last evaluated. For each data source, get a list of all data elements used as input to the model that came from that source. For insurance data, get a list all companies whose data is included in the datasets.</p> <p>Request details of any non-insurance data used (customer-provided or other), whether the data was collected by use of a questionnaire/checklist, whether data was voluntarily reported by the applicant, and whether any of the data is subject to the federal Fair Credit Reporting Act (FCRA). If the data is from an outside source, find out what steps were taken to verify the data was accurate, complete, and unbiased in terms of a relevant and representative time frame, representative of potential exposures, and lacking in obvious correlation to protected classes.</p> <p>Note: Reviewing source details should not make a difference when the model is new or refreshed; refreshed models would report the prior version list with the incremental changes due to the refresh.</p>
A.1.b	Reconcile aggregated insurance data underlying the model with available external insurance reports.	4	<p>Accuracy of insurance data should be reviewed. It is assumed that the data in the insurer’s data banks is subject to routine internal company audits and reconciliation. “Aggregated data” is straight from the insurer’s data banks without further modification (i.e., not scrubbed or transformed for the purposes of modeling). In other words, the data would not have been specifically modified for the purpose of model building. The company should provide some form of reasonability check that the data makes sense when checked against other audited sources.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.1.c	Review the geographic scope and geographic exposure distribution of the raw data for relevance to the state where the model is filed.	2	Many models are developed using a countrywide or a regional dataset. The company should explain how the data used to build the model makes sense for a specific state. The regulator should inquire which states were included in the data underlying the model build, testing, and validation. The company should explain why any states were excluded from the countrywide data. The company should provide an explanation where the data came from geographically and that it is a good representation for a state; i.e., the distribution by state should not introduce a geographic bias. However, there could be a bias by peril or wind-resistant building codes. Evaluate whether the data is relevant to the loss potential for which it is being used. For example, verify that hurricane data is only used where hurricanes can occur. The company should provide a demonstration that the model fits well on the specific state or surrounding region.
<b>2. Sub-Models</b>			
A.2.a	Consider the relevance of (i.e., whether there is bias) of overlapping data or variables used in the model and sub-models.	3	Check if the same variables/datasets were used in the model, a sub-model, or as stand-alone rating characteristics. Random Forest models handle redundant variables by splitting on only one of the variables within each component tree. By contrast, generalized linear models (GLMs) struggle with redundant variables as they try to include redundant variables simultaneously. However, best actuarial practice is to keep models as parsimonious as possible and only include additional variables that contribute significant additional predictive power.
A.2.b	Determine if the sub-model was previously approved (or accepted) by the regulatory agency.	1	<p>If the sub-model was previously approved/accepted, that may reduce the extent of the sub-model's review. If approved, obtain the tracking number(s) (e.g., state, System for Electronic Rates &amp; Forms Filing [SERFF]) and verify when and if it was the same model currently under review.</p> <p>Note: A previous approval does not necessarily confer a guarantee of ongoing approval; e.g., when statutes and/or regulations have changed or if a model's indications have been undermined by subsequent empirical experience. However, knowing whether a model has been previously approved can help focus the regulator's efforts and determine whether the prior decision needs to be revisited. In some circumstances, direct dialogue with the vendor could be quicker and more useful.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.2.c	Determine if the sub-model output was used as input to the Random Forest; obtain the vendor name, as well as the name and version of the sub-model.	1	<p>To accelerate the review of the filing, it may be desirable to request (from the company) the name and contact information for a vendor representative. The company should provide the name of the third-party vendor and a contact in the event the regulator has questions. The “contact” can be an intermediary at the insurer (e.g., a filing specialist), who can place the regulator in direct contact with a subject matter expert (SME) at the vendor.</p> <p>Examples of such sub-models include credit/financial scoring algorithms and household composite score models. Sub-models can be evaluated separately and in the same manner as the primary model under evaluation. A sub-model contact for additional information should be provided. Sub-model SMEs may need to be brought into the conversation with regulators (whether in-house or third-party sub-models are used).</p>
A.2.d	If using catastrophe model output, identify the vendor and the model settings/assumptions used when the model was run.	1	<p>To accelerate the review of the filing, get contact information for the SME that ran the model and an SME from the vendor. The “SME” can be an intermediary at the insurer (e.g., a filing specialist), who can place the regulator in direct contact with the appropriate SMEs at the insurer or model vendor.</p> <p>For example, it is important to know hurricane model settings for storm surge, demand surge, and long-term/short-term views.</p>
A.2.e	Obtain an explanation of how catastrophe models are integrated into the model to ensure no double-counting.	1	<p>If a weather-based sub-model is input to the Random Forest under review, loss data used to develop the model should not include loss experience associated with the weather-based sub-model. Doing so could cause distortions in the modeled results by double-counting such losses when determining relativities or loss loads in the filed rating plan.</p> <p>For example, redundant losses in the data may occur when non-hurricane wind losses are included in the data while also using a severe convective storm model in the actuarial indication. Such redundancy may also occur with the inclusion of fluvial or pluvial flood losses when using a flood model or inclusion of freeze losses when using a winter storm model.</p>
A.2.f	If using output of any scoring algorithms, obtain a list of the variables used to determine the score, and provide the source of the data used to calculate the score.	1	Any sub-model should be reviewed in the same manner as the primary model that uses the sub-model's output as input. Depending on the result of item A.2.b, the importance of this item may be decreased.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
<b>3. Adjustments to Data</b>			
A.3.a	Determine if premium, exposure, loss, or expense data were adjusted (e.g., on-leveled, developed, trended, adjusted for catastrophe experience, or capped). If so, how? Do the adjustments vary for different segments of the data? If so, identify the segments and how the data was adjusted.	2	<p>The rating plan or indications underlying the rating plan may provide special treatment of large losses and non-modeled large loss events. If such treatments exist, the company should provide an explanation of how they were handled. These treatments need to be identified, and the company/regulator needs to determine whether model data needs to be adjusted.</p> <p>For example, should large bodily injury (BI) liability losses in the case of personal automobile insurance be excluded, or should large non-catastrophe wind/hail claims in home insurance be excluded from the model's training, test, and validation data? Look for anomalies in the data that should be addressed. For example, is there an extreme loss event in the data? If other processes were used to load rates for specific loss events, how is the impact of those losses considered?</p> <p>Examples of losses that can contribute to anomalies in the data are large losses or flood, hurricane, or severe convective storm losses for personal automobile comprehensive or home insurance.</p> <p>Premium should be brought to current rate level if the target variable is calculated with a premium metric, such as loss ratio. Premium can be brought to current rate level with the extension of exposures method or the parallelogram method. Note that the premium must be on-leveled at a granular variable level for each variable included in the new model if the parallelogram method is used. Statewide on-level factors by coverage are typically sufficient for statewide rate indication development but not sufficient for models that determine rates by variable level.</p>
A.3.b	Identify adjustments that were made to aggregated data (e.g., transformations, binning, and/or categorizations). If any, identify the name of the characteristic/variable, and obtain a description of the adjustment.	1	Pre-modeling binning may be unnecessary in a Random Forest model. The tree model will naturally segment numerical values in the splitting process of the trees. However, if the insurer does bin variables before modeling, the reason should be understood.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
A.3.c	Ask for aggregated data (one dataset of pre-adjusted/scrubbed data and one dataset of post-adjusted/scrubbed data) that allows the regulator to focus on the univariate distributions and compare raw data to adjusted/binned/transformed/etc. data.	4	<p>This is most relevant for variables that have been “scrubbed” or adjusted.</p> <p>Though most regulators may never ask for aggregated data and do not plan to rebuild any models, a regulator may ask for this aggregated data or subsets of it.</p> <p>It would be useful to the regulator if the percentage of exposures and premium for missing information from the model data by category are provided. This data can be displayed in either graphical or tabular formats.</p>
A.3.d	Determine how missing data was handled.	1	<p>This is most relevant for variables that have been “scrubbed” or adjusted. The regulator should be aware of assumptions the modeler made in handling missing, null, or “not available” values in the data.</p> <p>For example, it would be helpful to the reviewer if the modeler were to provide a statement as to whether there is any systemic reason for missing data. If adjustments or recoding of values were made, they should be explained. It may also be useful to the regulator if the percentage of exposures and premium for missing information from the model data are provided. This data can be displayed in either graphical or tabular formats.</p> <p>The modeler should describe the way the tree fitting process handled missing values. The modeler should specify if missing values are treated before running the tree model or if they are allowed to be handled by the tree model.</p> <p>When creating predictions on new datasets (such as hold out datasets), tree-based models may have different approaches for handling missing data or categorical levels not encountered in the training data for a predictor variable. The modeler should specify the process utilized when this occurs.</p>
A.3.e	If duplicate records exist, determine how they were handled.	1	
A.3.f	Determine if there were any material outliers identified and subsequently adjusted during the scrubbing process.	3	<p>Look for a discussion of how outliers were handled. If necessary, the regulator may want to investigate further by getting a list (with description) of the types of outliers, and determine what adjustments were made to each type of outlier. To understand the filer's response, the regulator should ask for the filer's materiality standard.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
<b>4. Data Organization</b>			
A.4.a	Obtain documentation on the methods used to compile and organize data, including procedures to merge data from different sources or filter data based on particular characteristics and a description of any preliminary analyses, data checks, and logical tests performed on the data and the results of those tests.	2	This should explain how data from separate sources was merged and/or how subsets of policies, based on selected characteristics, are filtered to be included in the data underlying the model and the rationale for that filtering.
A.4.b	Obtain documentation on the insurer's process for reviewing the appropriateness, reasonableness, consistency, and comprehensiveness of the data, including a discussion of the rational relationship the data has to the predicted variable.	2	An example is when by-peril or by-coverage modeling is performed; the documentation should be for each peril/coverage and make rational sense.  For example, if "murder" or "theft" data is used to predict the wind peril, the company should provide support and a rational explanation for their use.
A.4.c	Identify material findings the company had during its data review, and obtain an explanation of any potential material limitations, defects, bias, or unresolved concerns found or believed to exist in the data. If issues or limitations in the data influenced modeling analysis and/or results, obtain a description of those concerns and an explanation how modeling analysis was adjusted and/or results were impacted.	1	"None" or "N/A" may be an appropriate response.

**B. BUILDING THE MODEL**

Section	Information Element	Level of Importance to the Regulator’s Review	Comments
<b>1. High-Level Narrative for Building the Model</b>			
B.1.a	Identify the type of model underlying the rate filing (e.g., Random Forest, GLM, decision tree, Bayesian GLM, gradient-boosting machine, neural network, etc.). Understand the model’s role in the rating system and provide the reasons why that type of model is an appropriate choice for that role.	1	<p>It is important to understand if the model in question is a Random Forest and, therefore, these information elements are applicable, or if it is some other model type, in which case other reasonable review approaches may be considered. There should be an explanation of why the model (using the variables included in it) is appropriate for the line of business. If by-peril or by-coverage modeling is used, the explanation should be by- peril/by-coverage.</p> <p><b>Note:</b> If the model is not a Random Forest, the information elements in this appendix may not apply in their entirety.</p>
B.1.b	Identify the software used for model development. Obtain the name of the software vendor/developer, software product, and a software version reference used in model development.	3	<p>Changes in software from one model version to the next may explain if such changes, over time, contribute to changes in the modeled results. The company should provide the name of the third-party vendor and a contact in the event the regulator has questions. The contact can be an intermediary at the insurer (e.g., a filing specialist) who can place the regulator in direct contact with the appropriate SME at the vendor.</p> <p>Open-source software/programs used in model development should be identified by name and version the same as if from a vendor.</p>
B.1.c	Obtain a description of how the available data was divided between model training, test, and/or validation datasets. The description should include an explanation why the selected approach was deemed most appropriate, whether the company made any further subdivisions of available data, and reasons for the subdivisions (e.g., a portion separated from training data to support testing of components during model building). Determine if the validation data was accessed before model training was completed and, if so, obtain an explanation of why that came to occur. Obtain a discussion of whether the model was rebuilt using all the data or if it was only based on the training data.	1	<p>The reviewer should be aware that modelers may break their data into three or just two datasets. Although the term “training” is used with little ambiguity, “test” and “validation” are terms that are sometimes interchanged, or the word “validation” may not be used at all.</p> <p>The reviewer should note whether a company employed cross-validation techniques instead of a training/test/validation dataset approach. If cross-validation techniques were used, the reviewer should request a description of how cross-validation was done and confirm that the final model was not built on any particular subset of the data, but rather the full dataset.</p> <p>The discussion of training, test, and/or validation datasets is a separate discussion from the percentage of observations (rows of data) or percentage of features (columns of data) used within each tree. These splits are based on hyperparameters and are commented on in other sections.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.1.d	Obtain a brief description of the development process, from initial concept to final model and filed rating plan.	1	The narrative should have the same scope as the filing.
B.1.e	Obtain a narrative on whether loss ratio, pure premium, or frequency/severity analyses were performed and, if separate frequency/severity modeling was performed, how pure premiums were determined.	1	
B.1.f	Identify the model's target variable.	1	A clear description of the target variable is key to understanding the purpose of the model. It may also prove useful to obtain a sample calculation of the target variable in Excel format, starting with the "raw" data for a policy, or a small sample of policies, depending on the complexity of the target variable calculation.
B.1.g	Obtain a description of the candidate variable selection process prior to the model building.	1	<p>Candidate variables are the variables used as input to the modeling process. Certain variables may not end up used in the final model if none of the component trees of the model split on the variable. The narrative regarding the candidate variable selection process may address matters such as the criteria upon which variables were selected or omitted, identification of the number of preliminary variables considered in developing the model versus the number of variables that remained, and any statutory or regulatory limitations that were taken into account when making the decisions regarding candidate variable selection.</p> <p>The modeler should comment on the use of automated feature selection algorithms to choose candidate predictor variables and explain how potential overfitting that can arise from these techniques was addressed.</p>
B.1.h	In conjunction with variable selection, obtain a narrative on how the company determined the granularity of the rating variables during model development.	3	The narrative should include discussion of how credibility was considered in the process of determining the level of granularity of the variables selected.
B.1.i	Determine if model input data was segmented in any way (e.g., by-coverage, by-peril, or by-form basis). If so, obtain a description of data segmentation and the reasons for data segmentation.	1	The regulator would use this to follow the logic of the modeling process.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
<b>2. Medium-Level Narrative for Building the Model</b>			
B.2.a	At crucial points in model development, if selections were made among alternatives regarding model assumptions, techniques, or hyperparameters, obtain a narrative on the judgment used to make those selections.	2	
B.2.b	If post-model adjustments were made to the data and the model was rerun, obtain an explanation on the details and the rationale for those adjustments.	2	Evaluate the addition or removal of variables and the model fitting. It is not necessary for the company to discuss each iteration of adding and subtracting variables, but the regulator should gain a general understanding of how these adjustments were done, including any statistical improvement measures relied upon.
B.2.c	Identify which distribution was used for the model (e.g., Regression based on Poisson, Gamma, Logistic, or Tweedie are common choices). Obtain an explanation of why the distribution was chosen. Certain distribution assumptions will involve numerical parameters; i.e., regression with a Tweedie assumed distribution will have a p power value. Obtain the specific numerical parameters associated with the distribution.	1	
B.2.d	Obtain a narrative on how the predictions from the component trees are combined to arrive at a final model prediction.	2	Tree-based methods combine predictions from multiple component trees and aggregate them into a final prediction for each observation. Common methods for combining Random Forest model predictions include the arithmetic or geometric mean of all the component trees.
B.2.e	If there were data situations in which weights were used, obtain an explanation of how and why they were used.	3	Investigate whether identical records were combined to build the model.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.2.f	Obtain the number of component trees comprising the Random Forest model. Obtain a narrative on how this number was chosen.	1	<p>Random Forest models should contain enough trees to reduce error to an acceptable level. Random Forest models should balance this with the concept of parsimony. A model with fewer trees that achieves relatively similar reduction in error is preferable to a model with more trees. Checking the error on a test dataset or out of bag error for different numbers of trees can reveal at what value the error on test data starts to level off.</p> <p>Modelers might rely on early stopping rules within modeling software to arrive at the final number of trees. The narrative on the number of trees should discuss the stopping criterion, which defines what condition is met when the model stopped adding more trees.</p>
B.2.g	Obtain the sampling parameters that apply to both the percent of observations used in each component tree and the number of features tested for each split within each tree. Obtain a narrative on how the sampling parameters were selected.	1	<p>Random Forest models often sample both the observations (typically rows of modeling data) with replacement and sample the features (typically columns of modeling data) This means that each tree has a bootstrapped dataset.</p> <p>The company should discuss the bagging fraction (sample size) applied to observations (typically rows of data). This is often expressed as a percent. For example: perhaps each tree is based on a bootstrapped sample that is 50% of the original dataset.</p> <p>The company should discuss the number of features considered at each split. This is often expressed as an integer. A common choice for the number of features is equal to roughly the square root of the total number of candidate variables. For example: perhaps each split is based on 10 randomly selected features (typically columns of data) when there are 100 candidate variables.</p>
B.2.h	Obtain the maximum depth that applies to the component trees in the model. Obtain a narrative on how this number was chosen.	1	<p>The depth of a tree is the number of splits that are allowed to occur between the root node and the terminal nodes. This number can be set explicitly in modeling software or may be implicitly set if the company applies a splitting constraint, such as a minimum observations per node. Maximum tree depths of eight or higher are considered extremely high.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.2.i	Obtain parameters that determined the volume of data in each tree node and a narrative of how parameters were chosen.	1	<p>Minimum data volume constraints can be applied to a tree-based model, such that the trees will not create a split that would result in terminal nodes with volume below a set amount. The modeler should comment on how the threshold was chosen.</p> <p>If there was no minimum data volume threshold applied to the trees, or if the threshold was exceedingly small, obtain an explanation of any post-modeling adjustments the modeler made to address the credibility considerations and how the adjustments were applied.</p>
B.2.j	Obtain a narrative of the process to select all hyperparameters for the Random Forest. Detail how this process addressed potential overfitting in the model.	2	<p>The narrative should include a description of each hyperparameter, document the values of the hyperparameters, specify the implication of using a higher or lower value for each hyperparameter, and discuss any sensitivity testing completed on the hyperparameters and observations from the sensitivity analysis. Hyperparameter tuning can be done in a variety of ways. The rigor of the tuning process should reflect the risk of overfitting on the specific dataset.</p>
<b>3. Predictor Variables</b>			
B.3.a	Obtain a complete data dictionary, including the names, types, definitions, and rationales for each variable.	1	<p>Types of variables might be continuous, discrete, Boolean, etc. Identify any variable used as an offset or control in the Random Forest model and the offset factor that was applied for each level of the offset variable.</p> <p>For any variable(s) intended to function as a control or offset, obtain an explanation of its purpose and impact. Also, for any use of interaction between variables, obtain an explanation of its rationale and impact.</p>
B.3.b	Obtain a list of predictor variables considered but not used in the final model and the rationale for their removal.	4	<p>The purpose of this requirement is to identify variables the company finds to be predictive but ultimately may reject for reasons other than loss-cost considerations (e.g., price optimization). Also, look for variables the company tested and then rejected. This item could help address concerns about data dredging.</p>
B.3.c	Obtain a correlation matrix for all predictor variables included in the model and sub-model(s).	3	<p>High correlation is less of an issue for tree-based models than it is for GLMs. Tree-based models naturally only use one variable at a time during each split in each tree. However, a correlation matrix still helps the reviewer understand relationships in the data being modeled better. The company should indicate what statistic was used (e.g., Pearson, Cramer's V, etc.) in the correlation matrix. The regulatory reviewer should understand what statistic was used to produce the matrix but should not prescribe the statistic.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.3.d	Obtain plots describing the relationship between each predictor variable and the target variable. Obtain a rational explanation for the observed relationship between each predictor variable and the target variable (frequency, severity, loss costs, expenses, or any element or characteristic being predicted).	1	<p>Partial dependence plots (PDPs), accumulated local effects (ALE) plots, or Shapley plots will help improve model interpretability. There should be at least one plot for every variable used in the model. The plots should be accompanied by commentary on why the visualized relationship is reasonable for variables of concern. Considering possible causation may be relevant, but proving causation is neither practical nor expected. If no rational explanation can be provided, greater scrutiny may be appropriate.</p> <p>For example, the regulator should look for unfamiliar predictor variables and, if found, the regulator should seek to understand the relationship that variable has to the target variable.</p> <p>The regulator should also consider that interpretability plots for tree-based models need to be reviewed with other considerations in mind. For example, partial dependence calculations assume independence with other variables in the model.</p>
B.3.e	If the modeler made use of one or more dimensionality reduction techniques, such as a principal component analysis (PCA), obtain a narrative about that process, an explanation why that technique was chosen, and a description of the step-by-step process used to transform observations (usually correlated) into a set of linearly uncorrelated variables. In each instance, obtain a list of the pre-transformation and post-transformation variable names, as well as an explanation of how the results of the dimensionality reduction technique was used within the model.	2	
B.3.f	Obtain variable importance plots. Obtain a description of how variable importance was calculated.	1	<p>Variable Importance Plots for tree-based methods highlight which variables contributed most to the model. There are multiple ways to calculate variable importance.</p> <p>Variables with the lowest importance measures should be prioritized when identifying variables that may not be contributing significantly to the model. Variables may have a low importance measure due to high correlation with other variables but may still prove useful if they interact with other variables to identify unique subsets of risks.</p> <p>Variables with the highest importance measures should be prioritized when determining which variables have the largest impact on predictions.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
<b>4. Adjusting Data, Model Validation, and Goodness-of-Fit Measures</b>			
B.4.a	Obtain a description of the methods used to assess the statistical significance/goodness-of-fit of the model to validation data, such as lift charts and statistical tests. Compare the model's projected results to historical actual results and verify that modeled results are reasonably similar to actual results from validation data.	1	<p>For models that are built using multistate data, validation data for some segments of risk is likely to have low credibility in individual states. Nevertheless, some regulators require model validation on state-only data, especially when analysis using state-only data contradicts the countrywide results. State-only data might be more applicable, but it could also be impacted by low credibility for some segments of risk.</p> <p><b>Note:</b> It may be useful to consider geographic stability measures for territories within the state.</p>
B.4.b	Obtain evidence that the model fits the training data well by variable and for the overall model.	2	<p>The regulator should ask for the company to provide exhibits or plots that show the fitted average makes sense when compared to the observed average for variables of interest. Regulators would ideally review this comparison for every variable, but time constraints may limit the focus to just variables of interest. Variables of interest should include those with a high importance measure (which will have the most material impact on rates), those with a low importance measure (which may not be contributing significantly to the model), variables without an intuitive relationship to loss, or variables that may be proxies for a protected class attribute.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.4.c	Obtain a description how the model was tested for stability over time.	2	<p>Evaluate the build/test/validation datasets for potential time-sensitive model distortions (e.g., a winter storm in year 3 of 5 can distort the model in both the testing and validation datasets).</p> <p>Obsolescence over time is a model risk (e.g., old data for a variable or a variable itself may no longer be relevant). If a model being introduced now is based on losses from years ago, the reviewer should be interested in knowing whether that model would be predictive in the proposed context. Validation using recent data from the proposed context might be requested. Obsolescence is a risk even for a new model based on recent and relevant loss data.</p> <p>The reviewer may want to inquire as to the following: What steps, if any, were taken during modeling to prevent or delay obsolescence? What controls exist to measure the rate of obsolescence? What is the plan and timeline for updating and ultimately replacing the model?</p> <p>The reviewer should also consider that as newer technologies enter the market (e.g., personal automobile), their impact may change claim activity over time (e.g., lower frequency of loss). So, it is not necessarily a bad thing that the results are not stable over time.</p>
B.4.d	Obtain a narrative on how potential concerns with overfitting were addressed.	2	<p>Tree-based models such as Random Forest models are notorious for overfitting. The company should provide a narrative on how overfitting was addressed. The company should provide a lift chart on training data used to fit the model and a lift chart on testing data that was not used to fit the model. If pruning was used to address overfitting, the narrative should provide commentary on the pruning process.</p>
B.4.e	Obtain support demonstrating that the Random Forest assumptions are appropriate.	3	<p>A visual review of plots of actual errors is usually sufficient.</p> <p>The reviewer should look for a conceptual narrative covering these topics: How does this particular Random Forest work? Why did the rate filer do what it did? Why employ this design instead of alternatives? Why choose this particular distribution function and this particular link function? A company response may be at a fairly high level and reference industry practices.</p> <p>If the reviewer determines that the model makes no assumptions that are considered to be unreasonable, the importance of this item may be reduced.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.4.f	Obtain 5-10 sample records with corresponding output from the model for those records.	2	The company should provide comprehensive documentation of the rating algorithm such that a rate can be reproduced for any theoretical risk. The company should demonstrate the comprehensiveness of the documentation by providing 5-10 sample records with corresponding input variable values and the final model prediction. The company should describe how the final model prediction aggregates the individual tree model predictions. The company should describe how to use other filing exhibits to reproduce the final model prediction for each sample record.
B.4.g	Obtain a deviance analysis by number of trees.	2	The company should provide a plot showing that the deviance of the overall model decreases after each iteration (each additional tree).
<b>5. "Old Model" Versus "New Model"</b>			
B.5.a	Obtain an explanation of why this model is an improvement to the current rating plan.  If it replaces a previous model, find out why it is better than the one it is replacing; determine how the company reached that conclusion and identify metrics relied on in reaching that conclusion. Look for an explanation of any changes in calculations, assumptions, parameters, and data used to build this model from the previous model.	2	The regulator should expect to see improvement in the new class plan's predictive ability or other sufficient reason for the change.
B.5.b	Determine if two Gini coefficients were compared and obtain a narrative on the conclusion drawn from this comparison.	3	This information element requests a comparison of the Lorenz curve and Gini coefficient from the prior model to the Gini coefficient of proposed model. It is expected that there should be improvement in the Gini coefficient. A higher Gini coefficient indicates greater differentiation produced by the model and how well the model fits that data.  This is relevant when one model is being updated or replaced. The regulator should expect to see improvement in the new class plan's predictive ability.  One example of a comparison might be sufficient.  Note: This comparison is not applicable to initial model introduction. The reviewer can look to CAS monograph, "Generalized Linear Models for Insurance Rating."
B.5.c	Determine if double-lift charts were analyzed and obtain a narrative on the conclusion drawn from this analysis.	3	One example of a comparison might be sufficient.  Note: "Not applicable" is an acceptable response.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
B.5.d	<p>If replacing an existing model, obtain a list of any predictor variables used in the old model that are not used in the new model as candidate variables. Obtain an explanation of why these variables were dropped from the new model.</p> <p>Obtain a list of all new predictor variables in the new model that were not in the prior old model.</p>	2	It is useful to differentiate between old and new variables so the regulator can prioritize more time on variables not yet reviewed.
<b>6. Modeler Software</b>			
B.6.a	Request access to SMEs (e.g., modelers) who led the project, compiled the data, and/or built the model.	4	The filing should contain a contact that can put the regulator in touch with appropriate SMEs and key contributors to the model development to discuss the model.

**C. THE FILED RATING PLAN**

Section	Information Element	Level of Importance to the Regulator’s Review	Comments
<b>1. General Impact of Model on Rating Algorithm</b>			
C.1.a	In the actuarial memorandum or explanatory memorandum, for each model and sub-model (including external models), look for a narrative that explains each model and its role (i.e., how it was used) in the rating system.	1	<p>The “role of the model” relates to how the model integrates into the rating plan as a whole and where the effects of the model are manifested within the various components of the rating plan. This is not intended as an overarching statement of the model’s goal, but rather a description of how specifically the model is used.</p> <p>This item is particularly important if the role of the model cannot be immediately discerned by the reviewer from a quick review of the rate and/or rule pages. (Importance is dependent on state requirements and ease of identification by the first layer of review and escalation to the appropriate review staff.)</p>
C.1.b	Obtain an explanation of how the model was used to adjust the filed rating algorithm.	1	The regulator should consider asking for an explanation of how the model was used to adjust the rating algorithm.
C.1.c	Obtain a complete list of characteristics/variables used in the proposed rating plan, including those used as input to the model (including sub-models and composite variables) and all other characteristics/variables (not input to the model) used to calculate a premium. For each characteristic/variable, determine if it is only input to the model, whether it is only a separate univariate rating characteristic, or whether it is both input to the model and a separate univariate rating characteristic. The list should include transparent descriptions (in plain language) of each listed characteristic/variable.	1	Examples of variables used as inputs to the model and used as separate univariate rating characteristics might be criteria used to determine a rating tier or household composite characteristic.
<b>2. Relevance of Variables and Relationship to Risk of Loss</b>			
C.2.a	Obtain a narrative regarding how the characteristics/rating variables included in the filed rating plan relate to the risk of insurance loss (or expense) for the type of insurance product being priced.	2	<p>The narrative should include a discussion of the relevance each characteristic/rating variable has on consumer behavior that would lead to a difference in risk of loss (or expense). The narrative should include a rational relationship to cost, and model visualization plots (such as partial dependence plots, accumulated local effects plots, or Shapley plots) should be consistent with the expected direction of the relationship.</p> <p><b>Note:</b> This explanation would not be needed if the connection between variables and risk of loss (or expense) has already been illustrated.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
<b>3. Comparison of Model Outputs to Current and Selected Rating Factors</b>			
C.3.a	Obtain documentation and support for all calculations, judgments, or adjustments that connect the model's indicated values to the selected rates filed in the rating plan.	1	<p>The documentation should include explanations for the necessity of any such adjustments and each significant difference between the model's indicated values and the selected values. This applies even to models that produce scores, tiers, or ranges of values for which indications can be derived.</p> <p><b>Note:</b> This information is especially important if differences between model-indicated values and selected values are material and/or impact one consumer population more than another.</p>
C.3.b	For each characteristic/variable used as both input to the model (including sub-models and composite variables) and as a separate univariate rating characteristic, obtain a narrative regarding how each characteristic/variable was tempered or adjusted to account for possible overlap or redundancy in what the characteristic/variable measures.	2	<p>The insurer should address this possibility or other considerations; e.g., tier placement models often use risk characteristics/variables that are also used elsewhere in the rating plan.</p> <p>One way to do this would be to model the loss ratios resulting from a process that already uses univariate rating variables. Then the model/composite variables would be attempting to explain the residuals.</p>
<b>4. Responses to Data, Credibility, and Granularity Issues</b>			
C.4.a	Determine what, if any, consideration was given to the credibility of the output data.	2	The regulator should determine at what level of granularity credibility is applied. If modeling was by coverage, by form, or by peril, the company should explain how these were handled when there was not enough credible data by coverage, form, or peril to model.
C.4.b	If the rating plan is less granular than the model, obtain an explanation of why.	2	This is applicable if the company had to combine modeled output in order to reduce the granularity of the rating plan.
C.4.c	If the rating plan is more granular than the model, obtain an explanation of why.	2	A more granular rating plan may imply that the company had to extrapolate certain rating treatments, especially at the tails of a distribution of attributes, in a manner not specified by the model indications. It may be necessary to extrapolate due to data availability or other considerations.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
<b>5. Definitions of Rating Variables</b>			
C.5.a	Obtain a narrative regarding adjustments made to model output (e.g., transformations, binning and/or categorizations). If adjustments were made, obtain the name of the characteristic/variable and a description of the adjustment.	2	If rating tiers or other intermediate rating categories are created from model output, the rate and/or rule pages should present these rating tiers or categories. The company should provide an explanation of how model output was translated into these rating tiers or intermediate rating categories.
<b>6. Supporting Data</b>			
C.6.a	Obtain aggregated state-specific, book-of-business-specific univariate historical experience data, separately for each year included in the model, consisting of loss ratio or pure premium relativities and the data underlying those calculations for each category of model output(s) proposed to be used within the rating plan. For each data element, obtain an explanation of whether it is raw or adjusted and, if the latter, obtain a detailed explanation for the adjustments.	4	For example, were losses developed/undeveloped, trended/untrended, capped/uncapped, etc.?  Univariate indications should not necessarily be used to override more sophisticated multivariate indications. However, they do provide additional context and may serve as a useful reference.
<b>7. Consumer Impacts</b>			
C.7.a	Obtain a listing of the top five rating variables that contribute the most to large swings in renewal premium, both as increases and decreases, as well as the top five rating variables with the largest spread of impact for both new and renewal business.	4	These rating variables may represent changes to rating factors, be newly introduced to the rating plan, or have been removed from the rating plan.
C.7.b	Determine if the company performed sensitivity testing to identify significant changes in premium due to small or incremental change in a single risk characteristic. If such testing was performed, obtain a narrative that discusses the testing and provides the results of that testing.	3	One way to see sensitivity is to analyze a graph of each risk characteristic's/variable's average fitted model prediction. Look for significant variation between the average fitted model predictions for adjacent rating variable levels and evaluate if such variation is reasonable and credible.
C.7.c	For the proposed filing, obtain the impacts on renewal business, and describe the process used by management, if any, to mitigate those impacts.	2	Some mitigation efforts may substantially weaken the connection between premium and expected loss and expense and, hence, may be viewed as unfairly discriminatory by some states.

Section	Information Element	Level of Importance to the Regulator's Review	Comments
C.7.d	Obtain a rate disruption/dislocation analysis demonstrating the distribution of percentage and/or dollar impacts on renewal business (created by rerating the current book of business) and sufficient information to explain the disruptions to individual consumers.	2	<p>The analysis should include the largest dollar and percentage impacts arising from the filing, including the impacts arising specifically from the adoption of the model or changes to the model as they translate into the proposed rating plan.</p> <p>While the default request would typically be for the distribution/dislocation of impacts at the overall filing level, the regulator may need to delve into the more granular variable-specific effects of rate changes if there is concern about particular variables having extreme or disproportionate impacts, or significant impacts that have otherwise yet to be substantiated.</p> <p>See Appendix D for an example of a disruption analysis.</p>
C.7.e	Obtain exposure distributions for the model's output variables and show the effects of rate changes at granular and summary levels, including the overall impact on the book of business.	3	See Appendix D for an example of an exposure distribution.
C.7.f	Identify policy characteristics, used as input to a model or sub-model, that remain "static" over a policy's lifetime versus those that will be updated periodically. Obtain a narrative on how the company handles policy characteristics that are listed as "static," yet change over time.	3	<p>Some examples of "static" policy characteristics are prior carrier tenure, prior carrier type, prior liability limits, claim history over past X years, or lapse of coverage. These are specific policy characteristics usually set at the time new business is written, used to create an insurance score or to place the business in a rating/underwriting tier, and often fixed for the life of the policy.</p> <p>The reviewer should be aware of, and possibly concerned about, how the company treats an insured over time when the insured's risk profile based on "static" variables changes over time, but the rate charged, based on a new business insurance score or tier assignment, no longer reflect the insured's true and current risk profile.</p> <p>A few examples of "non-static" policy characteristics are age of driver, driving record, and credit information (FCRA-related). These are updated automatically by the company on a periodic basis, usually at renewal, with or without the policyholder explicitly informing the company.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
C.7.g	Obtain a means to calculate the rate charged a consumer.	3	<p>The filed rating plan should contain enough information for a regulator to be able to validate policy premium. However, for a complex model or rating plan, a score or premium calculator via Excel or similar means would be ideal, but this could be elicited on a case-by-case basis. The ability to calculate the rate charged could allow the regulator to perform sensitivity testing when there are small changes to a risk characteristic/variable. Note: This information may be proprietary.</p> <p>For the rating plan, the rate order of calculation rule may be sufficient. However, it may not be feasible for a regulator to get all the input data necessary to reproduce a model's output. Credit and telematics models are examples of model types where model output would be readily available, but the input would not be readily available to the regulator.</p>
C.7.h	In the filed rating plan, be aware of any non-insurance data used as input to the model (customer-provided or other). In order to respond to consumer inquiries, it may be necessary to inquire as to how consumers can verify their data and correct errors.	1	<p>If the data is from a third-party source, the company should provide information on the source. Depending on the nature of the data, it may need to be documented with an overview of who owns it.</p> <p>The topic of consumer verification may also need to be addressed, including how consumers can verify their data and correct errors.</p>
<b>8. Accurate Translation of Model into a Rating Plan</b>			
C.8.a	Obtain sufficient information to understand how the model outputs are used within the rating system and to verify that the rating plan's manual, in fact, reflects the model output and any adjustments made to the model output.	1	The regulator can review the rating plan's manual to see that modeled output is properly reflected in the manual's rules, rates, factors, etc.
<b>9. Efficient and Effective Review of Rate Filing</b>			
C.9.a	Establish procedures to efficiently review rate filings and models contained therein.	1	<p>"Speed to market" is an important competitive concept for insurers. Although the regulator needs to understand the rate filing before accepting the rate filing, the regulator should not request information that does not increase his/her understanding of the rate filing.</p> <p>The regulator should review the state's rate filing review process and procedures to ensure that they are fair and efficient.</p>

Section	Information Element	Level of Importance to the Regulator's Review	Comments
C.9.b	Be knowledgeable of state laws and regulations in order to determine if the proposed rating plan (and models) are compliant with state laws and/or regulations.	1	This is a primary duty of state insurance regulators. The regulator should be knowledgeable of state laws and regulations and apply them to a rate filing fairly and efficiently. The regulator should pay special attention to prohibitions of unfair discrimination.
C.9.c	Be knowledgeable of state laws and regulations in order to determine if any information contained in the rate filing (and models) should be treated as confidential.	1	The regulator should be knowledgeable of state laws and regulations regarding confidentiality of rate filing information and apply them to a rate filing fairly and efficiently. Confidentiality of proprietary information is key to innovation and competitive markets.
C.9.d	Obtain complete documentation that would allow future audits of model predictions.	1	The company should provide comprehensive documentation of the rating algorithm such that a rate can be reproduced for any theoretical risk. Comprehensive documentation could be provided as one of the following: a complete set of tree diagrams, a set of if-else logic statements that represents the trees, or a table showing every possible combination of risk characteristics and the final prediction.

## RANDOM FOREST GLOSSARY OF TERMS

**Accumulated Local Effects Plots:** A type of interpretability plot. Accumulated local effects (ALE) plots calculate smaller, incremental changes in the feature effects. ALE shows the expected and centered effects of a variable.

**Bagged Trees:** An ensemble of trees where each tree is based on a “bootstrap aggregated” sample.

**Branch:** A connection on a decision tree between a parent node and a child node. A relationship based on a predictor variable is checked at each node, determining which branch applies.

**Candidate Variables:** The variables specified by the modeler to be used within the full model. The random variable selection by a Random Forest means that component trees might only use a subset of these variables in each tree.

**Child Node:** The node below a parent node. The child node is the result of a split that occurs based on a predictor variable. The node above the child node, which is where the split occurred resulting in the creation of the child nodes, is called the parent node. There is one parent node for every child node. The root node is the only node that is not a child node.

**Component Tree:** An individual tree within an approach based on an ensemble of trees, such as Random Forest or gradient boosting machine.

**Deviance:** A measure of model fit. Deviance is based on the difference between the log-likelihood of the saturated model and the log-likelihood of the proposed model being evaluated. Smaller values of deviance demonstrate that a model’s predictions fit closer to actual. Deviance on training data will always decrease as model complexity increases.

**Hyperparameter:** A model hyperparameter is a model setting specified by the modeler that is external to the model and whose value cannot be estimated from data.

**Node:** A point on a decision tree. Nodes are either root nodes (the top node), leaf nodes (a terminal node at which point no further splitting occurs), or an internal node that appears in the middle of the tree while splitting is still taking place.

**Out-of-Bag Error:** Error calculated for observations based on the trees that did not include them in the set of training observations. Out-of-Bag Error is calculable when bootstrapping is used to generate different datasets for each component tree in an ensemble tree method.

**Parent Node:** The node above a child node. The parent node is where a split occurs based on a predictor variable. The nodes below the parent node, which are a direct result of the parent node’s split, are called child nodes. There are typically two child nodes for every parent node. Terminal nodes cannot be parent nodes.

**Partial Dependence Plots:** A type of interpretability plot. The partial dependence plot computes the marginal effect of a given variable on the prediction.

**Pruning:** The process of scaling back a tree to reduce its complexity. This results in trees with fewer branches and terminal nodes appearing higher on the tree. Pruning is more common on models built on a single decision tree rather than on ensemble models such as Random Forests or gradient boosting machines.

**Random Forest:** An ensemble of trees where each tree is based on a bootstrap aggregated sample, and each split is based on a random sample of the candidate variables.

**Root Node:** The first (top) node in a decision tree. This node contains the entire set of data used by the tree as no splits have occurred yet.

**Shapley Additive Explanation Plots:** A type of interpretability plot. Shapley plots investigate the effect of including a variable in the model by the order in which it is added. The Shapley value represents the amount the variable of interest contributes to the prediction.

**Splitting:** The process of dividing a node into two or more sub-nodes, starting from the root node. Splitting occurs at every node up until the terminal (leaf) nodes when the stopping criterion is met.

**Stopping Criterion:** A criterion applied to the splitting process that informs the node when it is ineligible to split any further. Volume of data is often used as a stopping criterion, such that each leaf node is based on at least a pre-determined amount of data.

**Terminal Node:** An end node containing no child nodes because the node has met the stopping criterion. The terminal node is associated with a prediction for one of the component trees. The terminal node is also known as a “leaf” node, the resulting endpoint of a decision tree.

**Tree-Based Model:** A model that can be represented as a decision tree or a collection of decision trees.

**Tree Depth:** The maximum number of splits between the root node and a leaf node for a tree.

**Variable Importance:** A measure of how the variables (a.k.a. features) contribute to the overall model. There are multiple ways to measure variable importance.

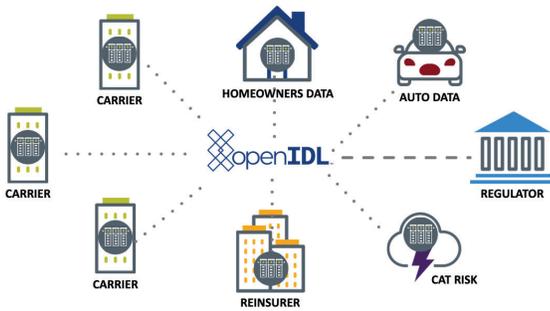
[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022\\_Summer/CASTF/CASTF%20Random%20Forest%20Appendix%20\(Final\).pdf](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022_Summer/CASTF/CASTF%20Random%20Forest%20Appendix%20(Final).pdf)



The Linux Foundation is working with the global technology community to solve the world's hardest problems through open source and **creating the largest shared technology investment in history**. Some of the 150+ gamechanging initiatives hosted by The Linux Foundation include:



openIDL is a permissioned insurance blockchain network and harmonized data store.



openIDL is for:

<p><b>Carriers</b></p> <p>Generating operational efficiency, flexibility, interoperability, and product development opportunities.</p>	<p><b>Regulators</b></p> <p>Access to timely and accurate information enabling more holistic and dynamic reporting as well as valuable and relevant insights into exposures, market activity, and trends.</p>	<p><b>Industry at Large</b></p> <p>Establishing and maintaining an agnostic network, providing efficient mechanisms for relevant participation and collaboration.</p>
--	---	---



### openIDL Members

<b>PREMIER MEMBERS</b>	
AAIS <small>Member National Auto Insurance Solutions</small>	The Hanover Insurance Group THE HARTFORD TRAVELERS
<b>GENERAL MEMBERS</b>	
SELECTIVE INSURANCE™	
<b>ASSOCIATE MEMBERS</b>	
Regulators MISSISSIPPI SCC MISSISSIPPI	Associations MOBI
<b>INFRASTRUCTURE PARTNER MEMBERS</b>	
SENOFI CHAINYARD	

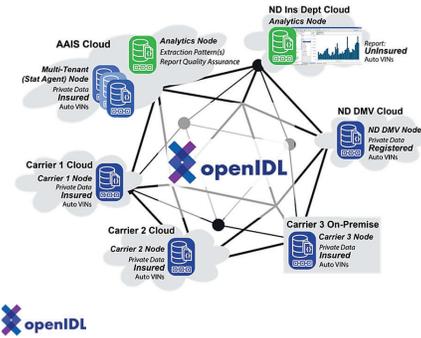


### Goals and Benefits of openIDL for the Insurance Industry

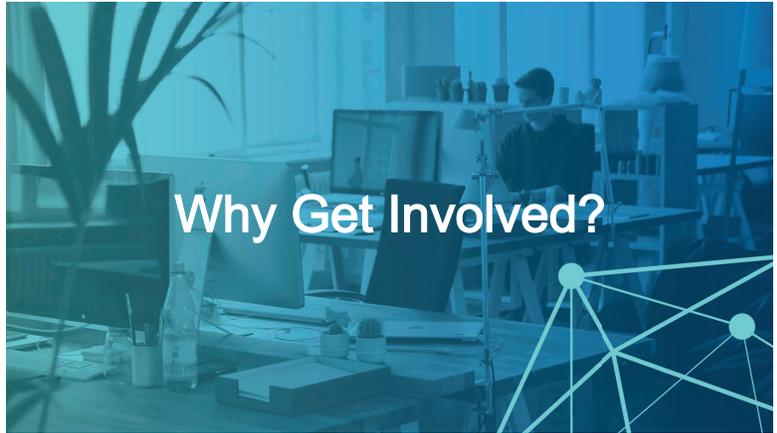
- Dramatically improve the **timeliness** and **availability** of insurance industry data for regulators
- **Reduce costs** and **improve security** of regulatory reporting by carriers
- Enable transparent and constructive **collaboration** of regulators, agencies, and carriers
- Provide a **standardized, federated** insurance data repository and an extensible network community



## North Dakota Uninsured Motorist Network



- › Establishes a role for regulators and advisory organizations on the network.
- › Demonstrates how “stat data” can be reused for other applications.
- › The network is being evaluated for procurement.



## openIDL for Regulators:

### Interdepartmental/Interjurisdictional Data

Data efficiency and integration across your state, departments and related jurisdictions.

### Officials and Legislatures

Provide legislators and leaders visualizations into trusted data for policy decisions and dashboards on impact and effectiveness.

### Community Information, Preparation and Response

Collaborate with the public and industry to prepare and protect communities from common risks.

### Licensing and Regulation

Connect trusted state and third-party data sources and licensing agencies to provide efficient experiences and solve problems unique to your community.

### Industry Audit and Compliance

Streamline and help your businesses and critical infrastructure remain compliant to the needs of the constituents and law through objective rules, transparent infrastructure and private data.



## openIDL for Technology Providers

### Products & Platforms

Create and leverage private data in the openIDL Regulatory Reporting Data Network or your own proprietary network while ensuring your assets and confidential data remain secure – within your control.

### Services & Partnerships

Design, implement and maintain solutions for your clients' businesses and operations, or your own unique vertical or horizontal partnerships to enable a unique customer experience and value.

### Data and Insights

Gather actionable insights from customer activities and experiences that can inform and improve the value of your products and services.

### Customer Data Privacy

Gain insights, efficiently and transparently, while still protecting your customers' data, identity and operations

### Quality and Accountability

Establish a framework of operations that assures integrity and objective quality, with transparent accountability and controls that establish and maintain your competitors.



## openIDL is Strategic for Member Insurers

As an **Insurer** and Member, leverage **openIDL** for you:

### Enterprise

- › Automate internal reporting/auditing, retire legacy datastores/endpoints, integrate/automate operations, reduce data sharing/shipping, dynamic capital management, internal accountability

### Partner Relationships

- › Efficient data operations, transparent/accountable performance, innovative solution delivery and new value-add opportunities in existing relationships: reinsurance, broker, TPA/MGA, agent, systems, etc.

### Programs

- › Parametric benefits, automated/objective underwriting, ML/AI knowledge pooling, responsive/specific rating, contract certainty, loss control solutions, niche and customized coverage programs



## Why companies join openIDL

### Thought Leadership

Be part of defining and maintaining the technologies that are at the forefront of the industry

### Networking & Partnerships

Network with other members of openIDL  
Support the community

### Build

Ensure the success of openIDL by building networks, products, services and solutions with and on top of openIDL's code bases that are critical to your lines of business.



*"It is strategically important for Selective to be part of industry efforts to innovate our regulatory reporting and use distributed ledgers"*



Michael H. Lanza  
Executive vice president, general counsel & chief compliance officer

## Contact us

General Information:

[info@openIDL.org](mailto:info@openIDL.org)

Jeff Braswell, Executive Director  
[jbraswell@openidl.org](mailto:jbraswell@openidl.org)



[Join openIDL](#)

Or join [openidl.slack.com](https://openidl.slack.com) and send us a message



Linux Foundation  
548 Market Street PMB 57274  
San Francisco, CA 94104  
Phone/Fax: +1 415 723-9709 [www.linuxfoundation.org](http://www.linuxfoundation.org)

14

## Legal Notices

The Linux Foundation, the Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates. Facebook and the "f" logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>, each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.

Please email [legal@linuxfoundation.org](mailto:legal@linuxfoundation.org) with any questions about The Linux Foundation's policies or the notices set forth on this slide.

THE LINUX FOUNDATION

[https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022\\_Summer/CASTF/openIDL%20NAIC%20Portland%202022%20CASTF%20slides%20.pptx](https://naiconline.sharepoint.com/sites/NAICSupportStaffHub/Member%20Meetings/C%20CMTE/2022_Summer/CASTF/openIDL%20NAIC%20Portland%202022%20CASTF%20slides%20.pptx)