Lumiata creates feature vectors including the new projection/blackout period claims information. (5) Lumiata applies the updated member- and group-level models to the claims data to produce group-level allowed trend predictions. (6) Using Delphi's paid trend predictions from 1b and the "stoplight" principle (Figure 3), Lumiata recommends whether to drop the rate by 5% for each group. (7) Lumiata sends its non-prediction attributes and the allowed trend predictions to Delphi with one recommendation per group up for renewal. All non-prediction fields must agree within $\leq 5\%$ between Delphi and Lumiata's calculations. (8) Delphi verifies receipt of the data and the results are consumed by their actuarial and underwriting teams. All files are transferred to and from Lumiata's platform (hosted on Google Cloud Platform) using sFTP.

Transparent Rate Setting Actuarial models have an essential property: they are "explainable" because a prediction can be decomposed into discrete multiplicative factors with an inherent interpretation. For example, "geographic area factor" = 0.9 means people in a particular zip code cost 10% less than the mean, so the base rate is adjusted (multiplied) by 0.9 for members from this zip code. This degree of explainability is crucial because actuaries need to file rates annually for individual/small group markets with the state insurance commissioner to ensure the factors used to produce the rate are compliant with legal guidelines. Furthermore, an underwriter needs to be able to explain how she arrived at a particular rate to a customer. A critique of ML models is that they lack explainability in terms of what input variables may have contributed to a particular prediction (**?**). However, explainable ML in healthcare is must-have, touching upon fundamental issues of bias, transparency, and reasonableness of ML model predictions.

Shapely values, a game theoretic algorithm, enable common ML algorithms to output feature weights specific to a prediction (**?**). We applied the Python package SHAP to our LightGBM models to yield member-level explanations. According to the algorithm, a model f and a member feature vector x admit an additive decomposition in terms of the mean value of the model $\mathbf{E}[f(x)]$ and SHAP values ϕ_i (interpreted as dollar value pmpm amounts) for the ith feature. For our member- and group-level models, f_1 and f_2 are:

$$f_1(x) = \mathbf{E}[f_1(x)] + \sum_{i=1}^{3996} \phi_{i,1} \tag{2}$$

$$f_2(x) = \mathbf{E}[f_2(x)] + \sum_{i=1}^{7} \phi_{i,2} \tag{3}$$

Given our member and group model sequence, our group model (3) admits an additive decomposition in terms of member-level $\phi_{i,1}$. The mechanics in (3) are similar to an actuarial formula, in that SHAP values for a group's members' features additively adjust the mean pmpm cost of the model over all members and groups, while the actuarial factors for a particular group multiplicatively adjust the mean pmpm cost of their entire patient population. We often found that $\leq$500 features' SHAP values account for 95% of a cost prediction, for each group. However, the specific features involved varied by group.

The transparency afforded by the group-level SHAP values provides the opportunity to explain a rate adjustment to a customer in dollar pmpm amounts using the specific risk drivers for that group and modify a rate given by the ML model by the expected change in cost for specific drugs and services. For instance, if the price of Glipizide, a drug to treat type-2 diabetes, will drop for an insurer next year by 20%, imagine a world where the insurer can multiply the SHAP values corresponding to Glipizide-related pharmacy costs by 0.80 for all the members on Glipizide, thus lowering the projected rate. These mechanics would be similar to current actuarial methods, making them easy to implement. Furthermore, greater model transparency could increase patient adherence to prescribed medications. As drug prices rise and more patients purchase high-deductible plans, patients have to pay higher out of pocket costs for drug treatment, and patient medication adherence declines (**?**). Insurers could use the SHAP values from the patient-level model to identify drugs driving up projected cost for that group, and suggest the prescribing doctor offer a lower cost alternative drug with similar efficacy. This provides a win-win opportunity, lowering drug cost for the payer, and improving patient adherence to the cheaper drug through increased affordability.

Discussion

Here, we demonstrate that: (1) ML approaches can significantly improve the accuracy and efficiency of group health insurance underwriting and (2) ML models can offer comparable interpretability to traditional actuarial methods. Our contributions provide clear direction for how to improve the efficiency and predictive performance of underwriting for employer-based insurance and how to lower the cost for members in groups of any size. Our ML-based approach improved MAE over actuarial models across the book of business: $>$500-member groups showed an improved performance.

Our model shows the most improvement over actuarial models in situations where the group size is $\leq$500 and/or the group claims experience is relatively short ($<$8 months). In these situations, the groups are not yet credible, so actuarial models perform sub-optimally. We believe the success of our model was due to modeling costs: (1) at the individual level; by contrast, actuarial methods aggregate medical history across group members, (2) with models that perform well with skewed distributions; the cost of care in an insurance population is often gamma distributed, making the ML method of gradient boosting trees, like LightGBM, highly effective (**?**), (3) using a model-agnostic approach to select features relevant for predictions, and (4) by combining individual- and group-level models to produce the final predictions.

We treated all members in the training set as if their group's renewal was 01/01/2017, with a four month blackout period starting on 08/31/2016. However, patients often try to "fit in" healthcare services before their next renewal

[4]https://github.com/slundberg/shap

period because they are likely near or above their plans deductible and hence will have these services fully covered by the insurance company. When patient records were "sliced" uniformly on 08/31/2016, there was a chance that this useful information would be lost and negatively impact the model performance. We found that dynamically "slicing" patient records according to their groups renewal date (instead of all patients on 08/31/2016), and training the models with this feature setup, the overall results on the holdout set remained roughly the same in terms of MAE and R^2. Model performance likely didn't improve because claims data are inherently fuzzy with dates; claims can take variable amounts of time to get paid, and hence a model that tries to learn precise date information will add some but not meaningful value. To account for fuzzy date information, we aggregated features over 3 months, 6 months, etc (see Methods).

As ML models are increasingly compared to more traditional statistical techniques, the most appropriate study design and model evaluation metrics should be examined. For example, the holdout set data was "out of sample" (i.e. using patients/groups unseen before by the model) but not "out of time" (i.e. projecting costs for time periods subsequent to 2017). Furthermore, claims data are not time-stationary (e.g. new drugs and treatments will be developed), so the expected model performance may not be perfectly realized in practice, but the relative difference between the models should hold. Also, we obtained a slightly worse Gini index than Delphi, despite our much better R^2, MAE, and lift plot (Table 3 and Figure 4). This discrepancy can occur because the Gini index is a ranking-based metric, whereas a regression model minimizes the prediction errors. One difficulty is the Gini index is not a differentiable quantity. Future work should develop algorithms to address this problem.

Data quality was crucial to our success. Alignment on non-prediction fields between Lumiata and Delphi ruled out errors in the data, pipeline, or output, improving communication across teams and increasing efficiency. These calculations must be automated for developing and productionalizing a medical underwriting ML application, due to the large size of data sets and rapid turnaround of results.

Due to its highly applied nature, some operational realities limit our study's evaluation. A challenge for validating our predictions is the long feedback cycle (20 months). Also, not all of Lumiata's concession recommendations could be granted due to a variety of quantitative and judgement factors under the underwriter and insurer's discretion.

Additionally, we could not determine if our individual-level model was racially biased, because we did not receive patient ethnicity data. Avoiding racial bias is important as previous studies have found evidence of racial bias in commercial cost prediction models used for clinical management (**?**). Historically, poorer minorities under-utilized healthcare services due to mistrust of the system and confusion about how to navigate it (**?**). However, because our response variable is not a clinical outcome but a financial one, we think this effect on pricing may be less significant. More work will be needed to better understand the effect of pricing insurance more affordably for minority patients, predicated on their less frequent utilization of the healthcare system.

In practice, ML approaches can help insurers be more competitive, avoiding adverse risk. It can result in the design of more "exotic" funding arrangements due to better predictive power of patient health, following the industry trend towards capitated payments (**?**). Unlike previous ML models in healthcare (**?**), our model output is interpretable by a non-technical user, simplifying operationalization (Figure 3). A user does not need to understand the inner workings of our algorithm to apply our output as a multiplicative adjustment factor to their existing actuarial models and can output the most important group-specific risk factors.

Conclusions

Machine learning on insurance claims data provides a powerful tool to improve the efficiency and affordability of plans and care offered to patients enrolled in employer-sponsored health plans. With more accurate rate-setting, health insurance companies can design nuanced plan attributes, reducing the cost of care for their members. Our ML model achieved 20% improved accuracy in absolute predictive performance over traditional actuarial methods and was able to identify over 80% of new concession opportunities available to Delphi. This allows underwriters to better price and retain $<$500 employer group customers. This study can be used by payers to give underwriters improved pricing guidance, retaining business and giving a better and more affordable experience to members.

Acknowledgements

We thank our counterparts at Delphi for collaboratively working with us to validate our model against a production-grade model with real data. Additionally, we thank the following people for their contributions: Kim Branson, Dilawar Syed, Laika Kayani, Wil Yu, Shahab Hassani, Alexandra Pettet, Derek Gordon, Ash Damle, Thomas Watson, Leon Barovic, Diana Rypkema and our investors at Blue Cross Blue Shield Venture Fund, and Khosla Ventures.

Quasi officia doloribus nisi debitis perspiciatis a modi quibusdam expedita consequuntur mollitia, aliquid itaque ab nostrum autem tempora odit velit rerum, repudiandae architecto neque ratione sapiente ducimus eum quis voluptatibus porro autem?Incidunt quo earum minus laudantium quis quasi rem, quo incidunt nihil?Iure sunt repellendus incidunt sint explicabo eos quos recusandae odio, corrupti odio doloribus fuga officiis, ipsam culpa inventore consequatur quae id velit asperiores ex quidem quam eos, distinctio vitae porro eum doloribus animi reprehenderit veritatis nulla delectus recusandae, earum quibusdam odit ex ratione?Eveniet minus quas eos quo libero asperiores recusandae sapiente, ullam adipisci veniam exercitationem nam doloremque consectetur repellat unde quaerat odit laboriosam, distinctio autem laboriosam, veritatis aliquam ex odio cumque nemo nesciunt non aliquid id consequuntur, aspernatur facilis fugiat obcaecati vitae eius minima?Labore animi doloribus fuga minima vel aliquid quod, dicta voluptas ipsam temporibus asperiores nulla id provident excepturi quaerat odit eveniet, doloremque dolorum ipsam excepturi magnam sapiente rerum.Dolorem explicabo placeat similique error, saepe aperiam facilis magni aliquam assumenda porro officia asperiores, earum delectus voluptatum voluptates ab