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