

# Predicting High-Cost Members in the HCCI Database





# Predicting High-Cost Members in the HCCI Database

## **AUTHOR**

Brian Hartman, Ph.D., ASA  
Actuarial Program Director  
Brigham Young University

Rebecca Owen, FSA, FCA, MAAA  
HealthCare Analytical Solutions

Zoe Gibbs  
Actuarial Science Major  
Brigham Young University

## **Caveat and Disclaimer**

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information

Copyright © 2018 by the Society of Actuaries. All rights reserved.

## CONTENTS

<b>1 Introduction</b> .....	<b>4</b>
<b>2 Data</b> .....	<b>5</b>
<b>3 Methods</b> .....	<b>6</b>
<b>4 Results</b> .....	<b>9</b>
4.1 Inference.....	9
4.2 Prediction.....	15
<b>5 Conclusion</b> .....	<b>16</b>
<b>6 Acknowledgments</b> .....	<b>20</b>
<b>References</b> .....	<b>21</b>
<b>About The Society of Actuaries</b> .....	<b>23</b>

# Predicting High-Cost Members in the HCCI Database

Using the Health Care Cost Institute (HCCI) database, which contains claim information on approximately 47 million members annually over a seven-year time period, we examined which characteristics best predict and describe high-cost members. We found that cost history, age, gender and prescription drug coverage are all predictors of future high costs, with cost history being the most predictive. We also compared the predictive accuracy of logistic regression to extreme gradient boosting and found that the added flexibility of this method improved the predictive power. Finally, we discovered that, with our extremely unbalanced classes, oversampling the minority class provides a better predictive model than undersampling the majority class or using the training data as is.

## 1 Introduction and Literature Review

A small proportion of members are responsible for a large part of the total health care costs. While most people use very few services—mostly preventive care or minor acute care—and some are regular consumers at a moderate level, nearly 75% of all health care expenditures are made by only 17% of the users (McWilliams and Schwartz, 2017). Some of this care is completely unexpected—burns or serious car accidents or the transition of a normally mild disease into a crisis due to an unexpected and unavoidable situation, such as the development of encephalitis from a case of West Nile virus. Some is due to lack of due care and caution, as well as terrible luck, such as the development of septicemia. Some of it is expected but not necessarily predictable, like cancer care. And some is the result of a chronic disease that has worsened in severity, so people are fighting for their lives after years of debility.

Due to the size of the contribution to costs of this small segment of the population, there is considerable interest in understanding what portion of it can be predicted. The portion that is entirely unpredictable (and rare) may be estimated by a distribution based on large population studies. It is the portion that could be estimated using a predictive model based on the characteristics of the population that is of interest; it would allow for some predictions in the future costs of a specific population, as well as identify people for interventions and additional care. While risk adjustment models are good at predicting average costs of care for a specific category of people, they are still not effective at identifying individuals who may be at risk for a very high claim in the near future.

This paper explores the types of models that will identify individuals who are most likely to exceed a high cost threshold based on a number of characteristics available in the most common information source: administrative claims data. While it is true that the addition of information from chart review and clinical recommendations are essential for really identifying a person at risk, this data are often difficult to incorporate into the actuarial studies for trend and pricing work. This work seeks to add another tool to the risk quantification process for members whose costs form a large part of the overall costs of care and significantly contribute to the force of trend.

Many authors have examined the issue of high-cost claimants from different directions and explored the common characteristics of high-cost patients. Zook and Moore (1980) were among the first to thoroughly examine high-cost members. They looked at 2,238 patients (of whom 13% were high-cost) and found that smoking and alcohol consumption were much more prevalent in the high-cost group than in the low-cost

one. Schroeder et al. (1979) examined high-cost members in San Francisco. They found that very few (17%) of the high-cost claimants suffered from an actual medical catastrophe; most had chronic conditions. The authors concluded that catastrophic insurance would have covered most of the costs of these patients. Joynt et al. (2013) found that only a small portion of the total spending for high-cost members was due to preventable acute care. Zulman et al. (2015) showed that multimorbidity is common among high-cost members of the U.S. Veterans Affairs Health Care System. They suggest that interventions are needed to help those members better manage multiple conditions.

Other studies have focused on how to predict which members within a population will be high-cost. Bates et al. (2014) encouraged the use of predictive models with the proliferation of electronic medical records. Garfinkel et al. (1988) used the National Medical Care Utilization and Expenditure Survey to look at predictors of high-cost patients. They found that health status, followed by economic factors, best predict high-cost members. Meenan et al. (2003) compared many contemporary risk-adjustment models to determine which were the best at predicting high-cost patients. Moturu et al. (2007, 2009) used health care data from Arizona to better predict high-risk members. Fleishman and Cohen (2010) analyzed the Medical Expenditure Panel Survey (MEPS) and compared the ability of a risk score (diagnostic cost group) with that of a count of chronic conditions to predict which members would be in the highest cost decile the following year. They also checked whether self-rated health status and functional limitations improved predictions. They found that the risk score was the best predictor. After controlling for the risk score, the number of chronic conditions, self-reported health status and functional limitations were significantly associated with future high costs.

Additionally, Billings and Mijanovich (2007) focused on ways to assist high-cost members once they are identified. Hong et al. (2014) compared 18 different care management systems for high-cost, high-need members. The authors defined best practices for care management programs going forward. Blumenthal et al. (2016) continued to discuss how to better care for the high-cost population, focusing on changes health care providers could make. Hayes et al. (2016) used the 2009–2011 MEPS to examine high-cost patients. They determined that targeted and tailored interventions are necessary for high-need individuals.

## 2 Data

Our data were gathered by the Health Care Cost Institute and comprise member information from three of the largest health insurers in the United States. When we performed our analysis, the data were from the years 2009 through 2015. The number of members in each year are listed in Table 1.

**Table 1**  
Number of Members in Each Dataset

Year	Number of Members
2009	48,511,544
2010	47,539,751
2011	46,193,435
2012	46,544,359
2013	47,351,996
2014	48,087,209
2015	47,782,320

The variables we are interested in for our analysis are described in Table 2.

**Table 2**  
Variable Names and Descriptions

Variable Name	Description
Z_PATID	Member ID number
RX_CVG_IND	Prescription drug coverage indicator (1 if the member has coverage). If 1, the pharmacy costs for the year are included in the total allowed costs (CAT).
GDR	Gender (1 for male, 2 for female)
AGE	Age in years
MKT_SGMNT_CD	Market segment code (I for individual market, G for individual group conversion, L for large, S for small, O for other) For inference, we focus only on the individual market, (INDV_FLAG), but in prediction, we will use all segments.
CAT	Total allowed, adjudicated cost for the year, divided into five groups (<\$100, \$100K–\$250K, \$250K–\$500K, \$500K–\$1M, >\$1M)

For each year, the vast majority of the members had less than \$100,000 in total claims. To understand how rare the group we are exploring is, Table 3 shows the number of members in each high-cost group and their percentage of the full dataset.

**Table 3**  
Number of High-Cost Members in Each Dataset

Year	\$100K–\$250K	%	\$250K–\$500K	%	\$500K–\$1M	%	>\$1M	%
2009	96,554	0.1990	17,738	0.0366	4,162	0.0086	661	0.0014
2010	100,812	0.2121	18,162	0.0382	4,393	0.0092	706	0.0015
2011	108,965	0.2359	20,375	0.0441	4,773	0.0103	841	0.0018
2012	117,325	0.2521	22,393	0.0481	5,250	0.0113	941	0.0020
2013	126,099	0.2663	24,275	0.0513	5,458	0.0115	998	0.0021
2014	135,050	0.2808	26,018	0.0541	5,749	0.0120	1,030	0.0021
2015	147,220	0.3081	28,425	0.0595	6,517	0.0136	1,200	0.0025

As is readily apparent, there are not many members in the extremely high-cost group. The numbers are further reduced because our analysis will focus on the members who are in the dataset for three consecutive years. For this reason we are grouping the members rather than predicting their costs directly. We are taking the role of an intervention manager and trying to find those members who are most likely to be high risk. The proportion of all members in each of the high-cost groups has also increased every year.

### 3 Methods

In our methods, we simulate common analyses for a health insurer. We will use data from the previous two years to predict whether the member will be high-cost in the following year. For example, to predict whether the member will be high-cost in 2012, we will use data from 2010 and 2011. Because we have data from 2009 through 2015, we will predict each member in 2011 through 2015. For each prediction year, we only use those members for whom we have data for the year in question and the previous two. That reduces our sample sizes to those shown in Table 4.

**Table 4**  
Number of Members in Each Three-Year Dataset

Year	Sample Size
2011	25,954,734
2012	26,539,732
2013	27,061,494
2014	26,425,810
2015	25,199,632

Reducing our dataset to only those who were members for three continuous years impacts our data (and therefore our inference) in several ways. First, there are no members under the age of two in our prediction datasets. Because of that, the median age of the members in our dataset is about nine years older than that of those not in our set (39 vs. 30). Further, much of the lifetime medical spending occurs in the final year of life, so those who were expensive and then passed away in previous years are not included in our dataset. About 25% of the high-cost members (>\$100,000) from any year are not in the dataset in the following year. Of those in our dataset, around 50% have prescription drug coverage, whereas about 60% of those not in our dataset have coverage. Additionally, there are proportionally about twice as many members in the actual full individual market compared to those in our dataset (about 8% vs. 4%). Most importantly, the proportion of high-cost members is about the same between the two groups, except for a few more people (one or two per 100,000 members) above 1 million in the group not in our dataset.

For inference, we will use the entire dataset and look at the parameter estimates and how they change over time. We fit logistic regression models to four different measures of high cost, greater than \$100,000, \$250,000, \$500,000 and \$1 million in claims in one year.

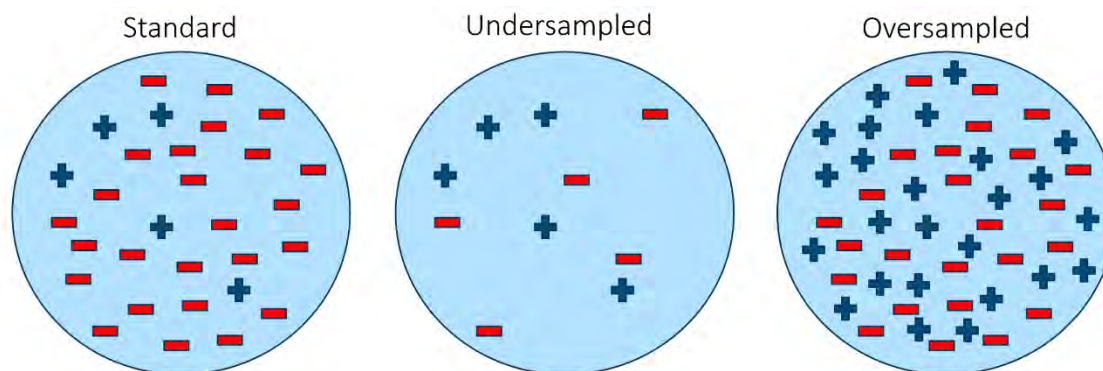
When we compare the predictive accuracy of the two models, we fit the model to one year (say 2012) and use that model to fit the following year (2013). This shows which model is superior in a realistic situation. This is better than comparing model accuracy by dividing each set into a training and test set.

Classification can be difficult when the positive class is extremely rare, as it is in our case. To help mitigate that issue, we train the 2012 models on three different datasets:

- **Standard.** The original 2011 data (say 1,000 high-cost members and 1 million low-cost members for illustration).
- **Under.** A dataset with the 2011 low-cost members undersampled, making an equal number of high- and low-cost members. We randomly select (without replacement) 1,000 of the 1 million low-cost members to be in the training set. This means that we have a training sample of 2,000 members.
- **Over.** A dataset with the high-cost members oversampled, again making an equal number of high- and low-cost members. We randomly select (with replacement) a sample of 1 million from the 1,000 high-cost members. In this case, our training sample will include 2 million members.

The three different datasets are further described in Figure 1, with the plus sign indicating high-cost members and the minus sign showing low-cost members.

**Figure 1**  
The Three Sampling Techniques



Regression trees are a machine learning method in which the data are divided into smaller and smaller groups until the remaining groups are relatively homogeneous. They are very flexible methods and can work really well when you are dealing with large datasets. Each division of the dataset is called a branch, and the final groups are called children (or leaves). You can improve the performance by making multiple trees from subsets of the data and then combining the results from those trees, a process called boosting. We will be using a specific version of boosted trees called XGBoost (Chen and Guestrin, 2016).

We can adjust a number of (hyperparameter) settings to change how flexible the model is or prevent overfitting:

- **Maximum tree depth, range = (3, 10).** Maximum number of branch levels in any tree. A higher number here makes it more likely that an individual tree is overfit, meaning that the individual children are much too small for the model to generalize well to new data.
- **Minimum child weight, range = (1, 10).** Parameter telling the tree-building process when to stop. If splitting a node would make a child have less weight than this parameter, then the process stops. The larger this value is, the simpler the trees will be.
- **Subsample, range = (0.5, 1).** Proportion of the total training set used to build each tree. A smaller value will help to prevent overfitting because each of the trees will be built on a noticeably different dataset.
- **Column sample by tree, range = (0.5, 1).** Proportion of all the possible covariates used to build each tree. Similar to subsample, a larger value will prevent the model from overfitting.
- **Eta, range = (0,1).** The learning rate. A higher eta will speed up convergence, while a lower eta may make the convergence more precise.

We created four different sets of hyperparameters. The first (untrained) uses commonly held default values for each of the hyperparameters listed. The next learner (trained1) starts with the hyperparameters from the untrained learner and then compares it to 10 different possible settings randomly drawn from the set of possible hyperparameters, shown in parentheses in the list. The settings are compared through threefold cross-validation and by choosing the set of hyperparameters that minimizes the area under the curve (AUC) in the cross-validated training set. The following learner (trained2) starts with the chosen



hyperparameters in trained1 and compares them to 10 other possible sets. The final learner (trained3) follows the same pattern. The chosen hyperparameters are shown in Table 5.

**Table 5**  
Chosen Learners

Parameter	Untrained	Trained1	Trained2	Trained3
Maximum tree depth	6	3	5	5
Minimum child weight	1	9.77	2.98	9.26
Subsample	1	0.66	0.79	0.97
Column sample by tree	1	0.76	0.6	0.69
Eta	0.3	0.54	0.52	0.63

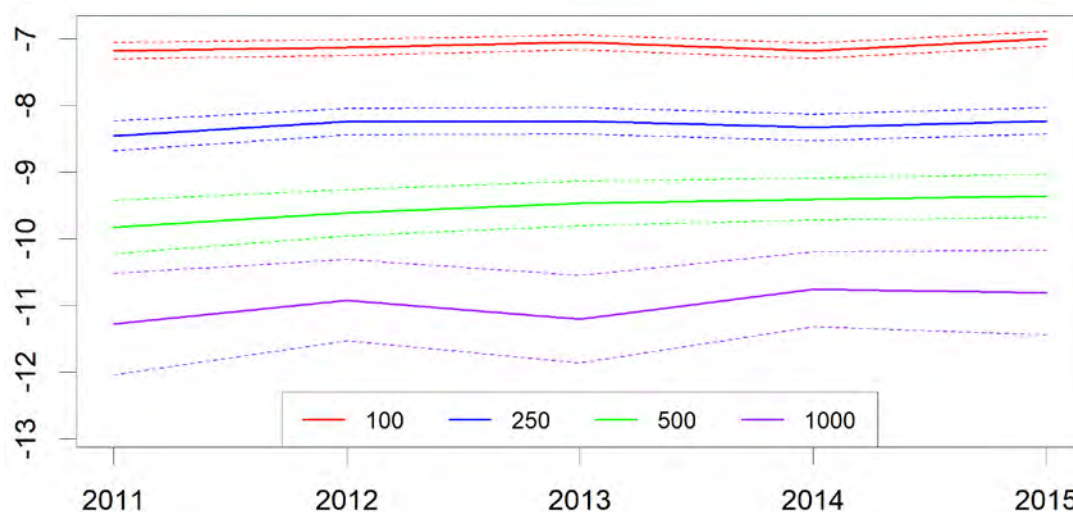
## 4 Results

We have divided our results section into two parts. First, we will examine the inference results gathered from the logistic regressions on each year from 2011 to 2015, using the two previous years to help predict current year cost. Second, we will discuss the prediction results of the various models.

### 4.1 Inference

To help visualize the impact of the various covariates on the high-cost predictions, we plot each of the regression coefficients across all the years and for each cut defining high-cost. The plots are shown in Figures 2 through 8. The solid lines are the fitted regression coefficients, while the dashed lines are the 95% confidence interval bounds. These are logistic regression coefficients, so they do not have the easy interpretability of linear regression coefficients. When looking at the plots, notice the broad patterns. Positive values mean the groups are more likely to be classified as high-cost. The magnitude matters as well: the higher the coefficient, the greater the increase in propensity to be high-cost. Also, the tighter the confidence bounds are, the more sure we are that those estimates are correct.

**Figure 2**  
Baseline Impact on Propensity to Be High-Cost



The intercepts (Figure 2) seem relatively constant over time. They are also in the expected order, with the \$1 million cutoff lowest and the \$100,000 cutoff highest. Holding all other variables constant, the probability of exceeding \$1 million must be no greater than the probability of exceeding \$100,000.

**Figure 3**  
Impact of Prescription Drug coverage on High-Cost Propensity

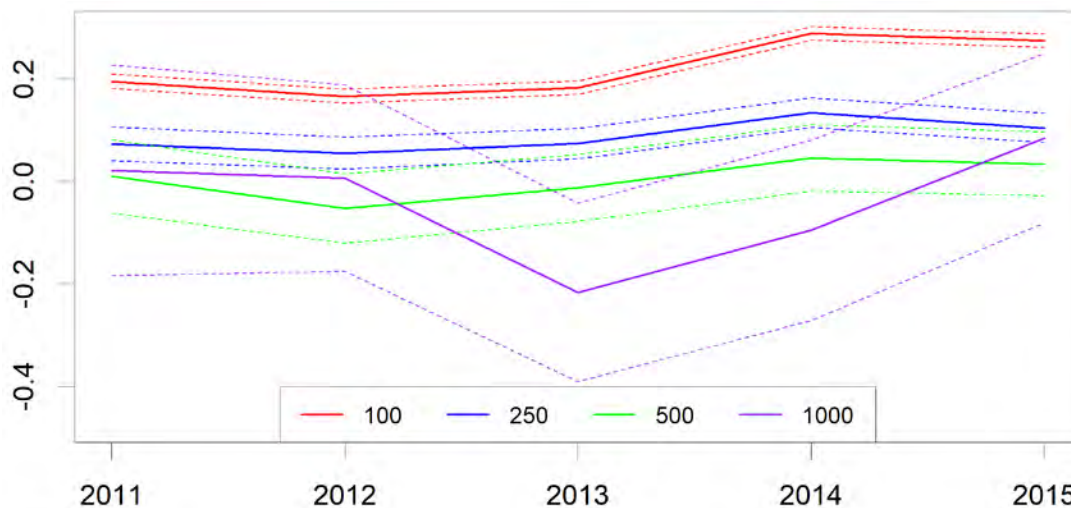
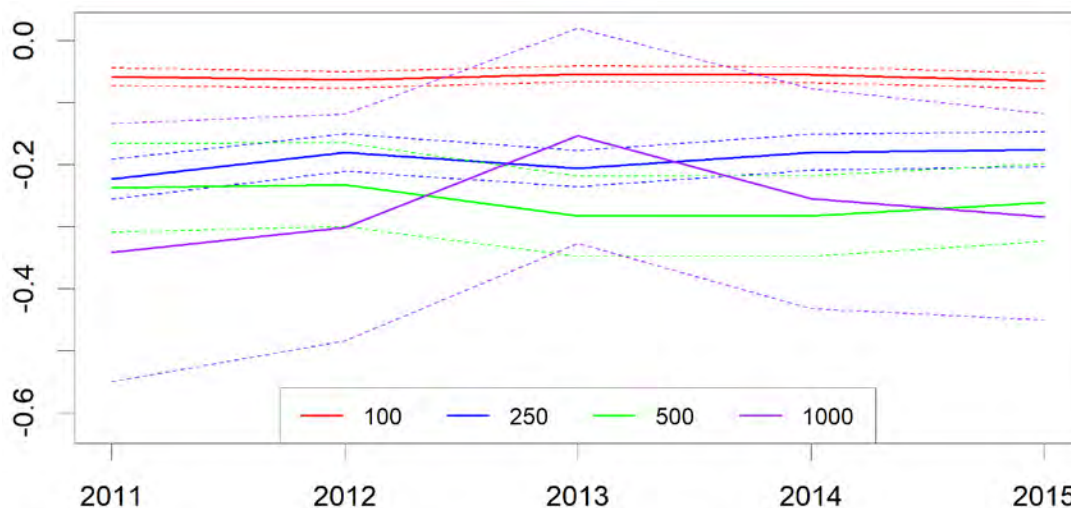


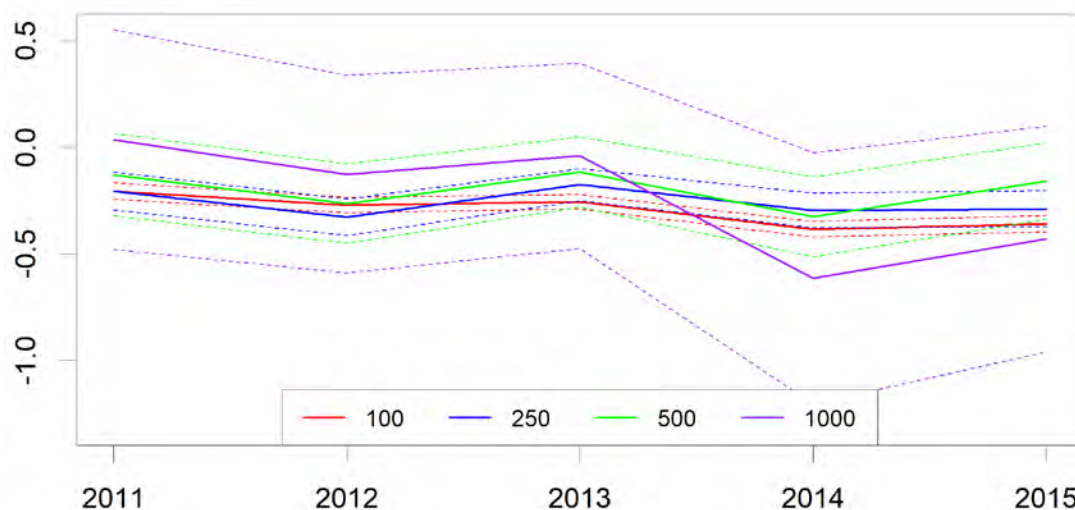
Figure 3 shows that members with prescription drug coverage are slightly more likely to have claims in excess of the cutoffs. The effect is relatively constant across years and is largest on the \$100,000 cutoff. The effect of prescription drug coverage is relatively small for all the other cutoffs.

**Figure 4**  
Impact of Being Female on High-Cost Propensity



The effect of gender (Figure 4) is relatively small at the \$100,000 cutoff, with female members being slightly less likely to exceed the limit. For the other three cutoffs, the effect is significantly larger. In all cases, female members are less likely to be high-cost. This effect is also rather consistent over time.

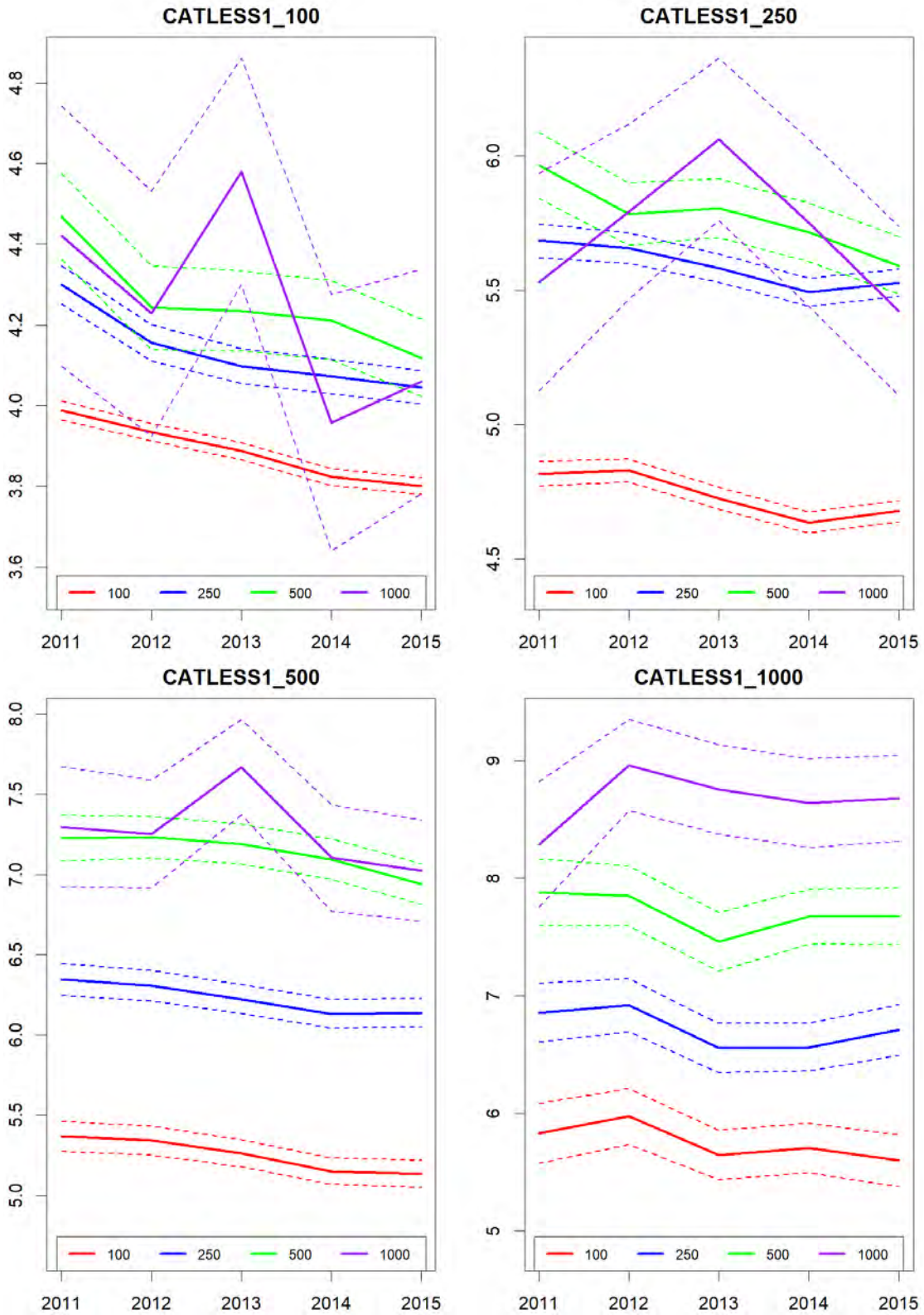
**Figure 5**  
Impact of Being in the Individual Market on High-Cost Propensity



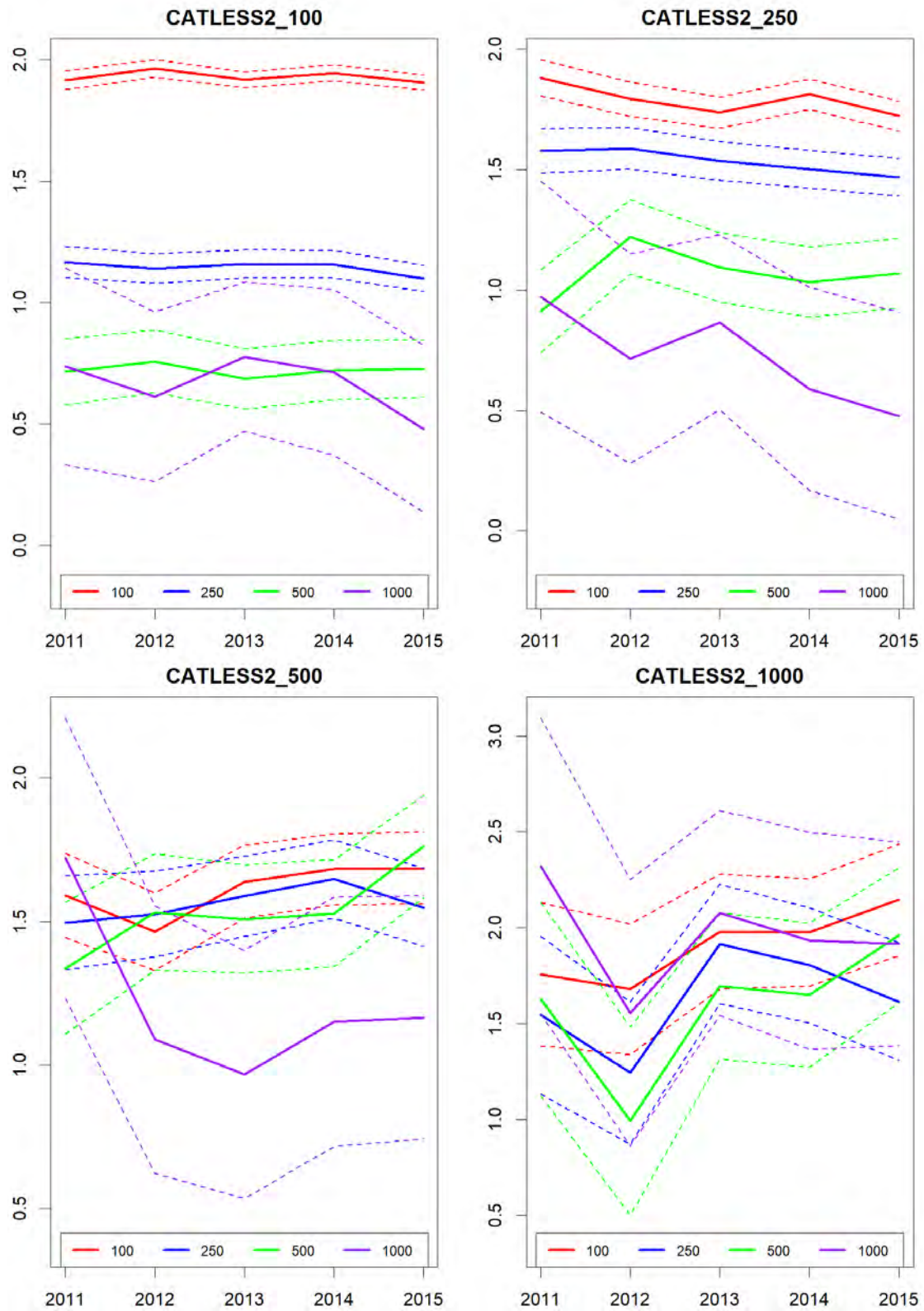
A member’s participation in the individual market (Figure 5) does not have a significant impact at the \$500,000 or \$1 million cutoffs, but it makes the member a little less likely to exceed the \$100,000 or \$250,000 cutoffs.

Member cost history (Figures 6 and 7, with CATLESS1 and CATLESS2 referring to which cost group a claimant was in 1 and 2 years ago, respectively) has the biggest impact on the probability of an individual being high-cost. When looking at the cost history, if members were high-cost last year, they are more likely to be high-cost this year. The impact increases as the prior year costs increase. Additionally, the impact increases as the cutoff increases. In all cases, the uncertainty around the \$1 million cutoff estimates are the greatest. The cost history from two years ago had a similar but more muted impact.

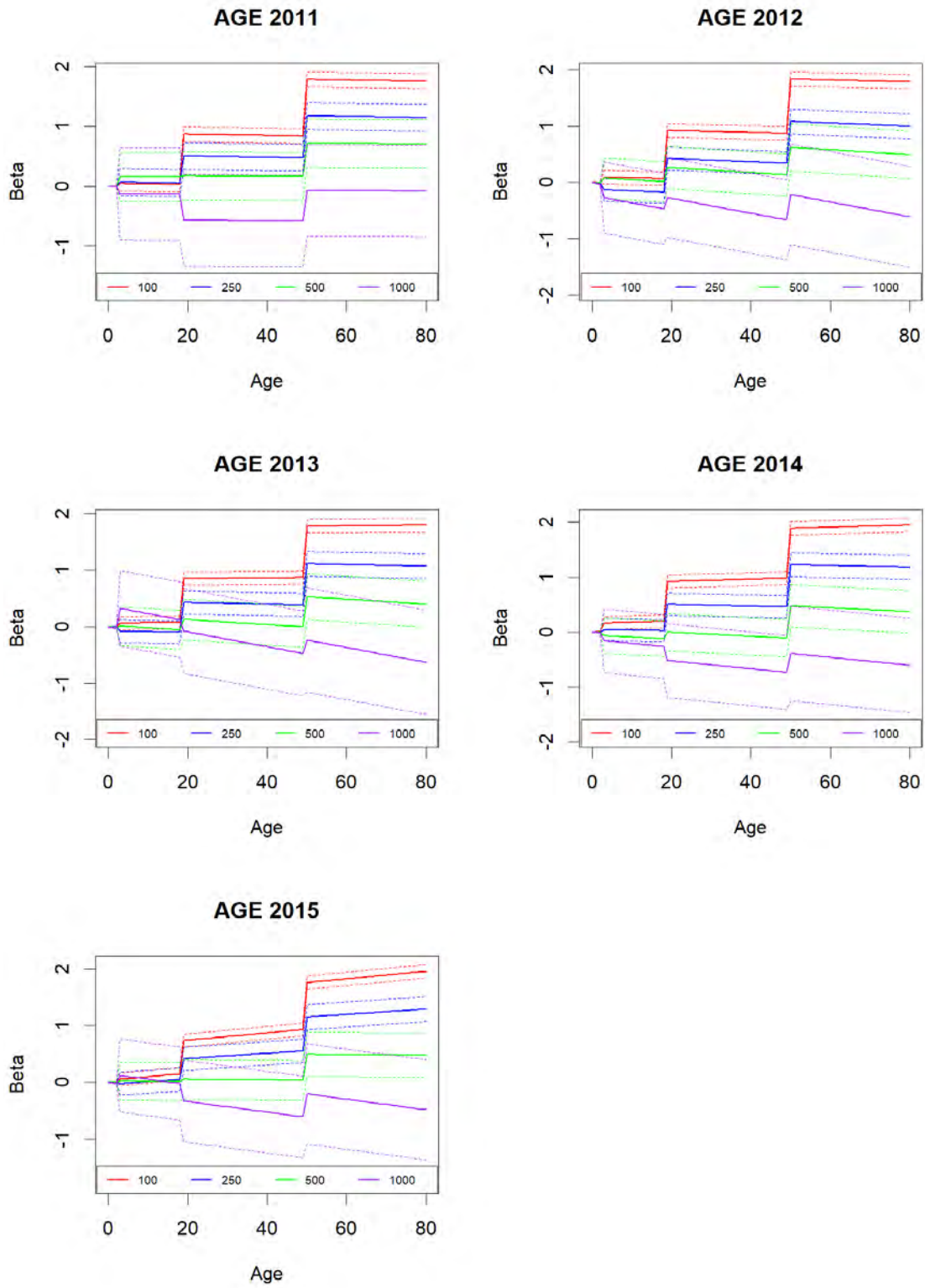
**Figure 6**  
Impact of Being High-Cost Last Year on High-Cost Propensity



**Figure 7**  
Impact of Being High-Cost Two Years Ago on High-Cost Propensity



**Figure 8**  
 Impact of Age for Each Prediction Year on High-Cost Propensity



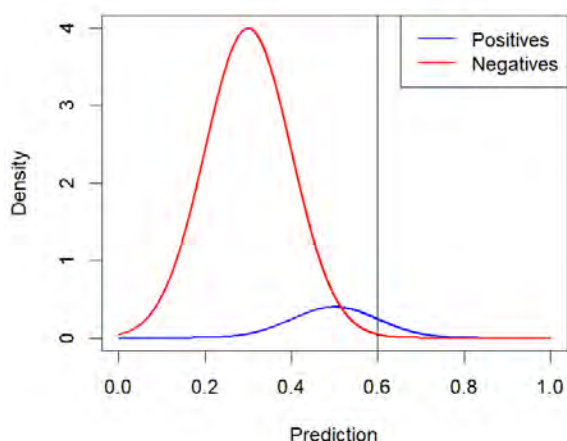
The age pattern (Figure 8) is relatively consistent across years. For the \$100,000, \$250,000 and \$500,000 cutoffs, members are more likely to be high-cost as they age. While the linear term for each additional year is relatively small, there are significant jumps between the categorical groups (at ages three, 18 and 50). The \$1 million cutoff has no significant age effect due to the large amount of uncertainty in these estimates.

### 4.2 Prediction

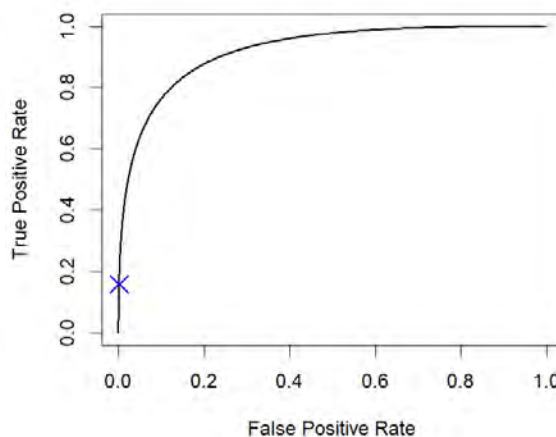
In this section we compare the predictions from a logistic regression to those from gradient boosted trees, both with default hyperparameter settings and settings optimized through cross-validation. We compare the accuracy using the area under the receiver operating characteristic (ROC) curve. The ROC curve plots the true positive rate against the false positive rate as the threshold changes. Therefore, the larger the AUC, the better the model discerns between high- and low-cost members. This allows us to compare how the models are performing over the entire dataset.

For example, using simulated data, if Figure 9a gave the predicted values of the positive (high-cost) and negative members, then a threshold of 0.6 (denoted by a vertical line) might correctly identify about 16% of the high-cost members as truly high-cost (true positives) but incorrectly identify about 0.1% of the low-cost members as high-cost (false positives). We would plot that point (0.0013, 0.1587) on a curve, as in Figure 9b (denoted by blue x). Finally, we would run through all possible thresholds and plot all of the resulting rates. The full curve is the ROC curve.

**Figure 9a**  
Simulated Distribution of High-Cost Propensities

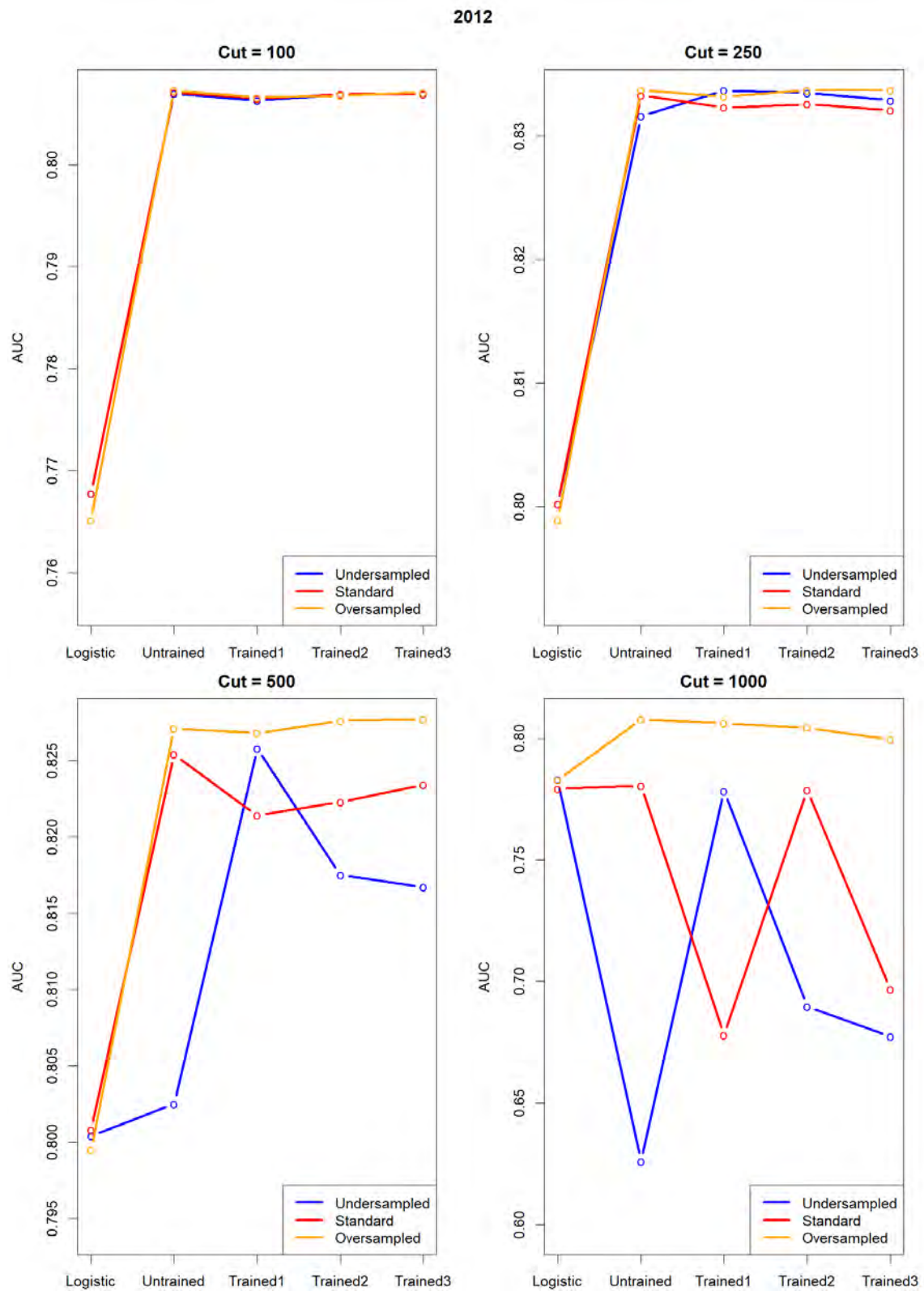


**Figure 9b**  
ROC Curve



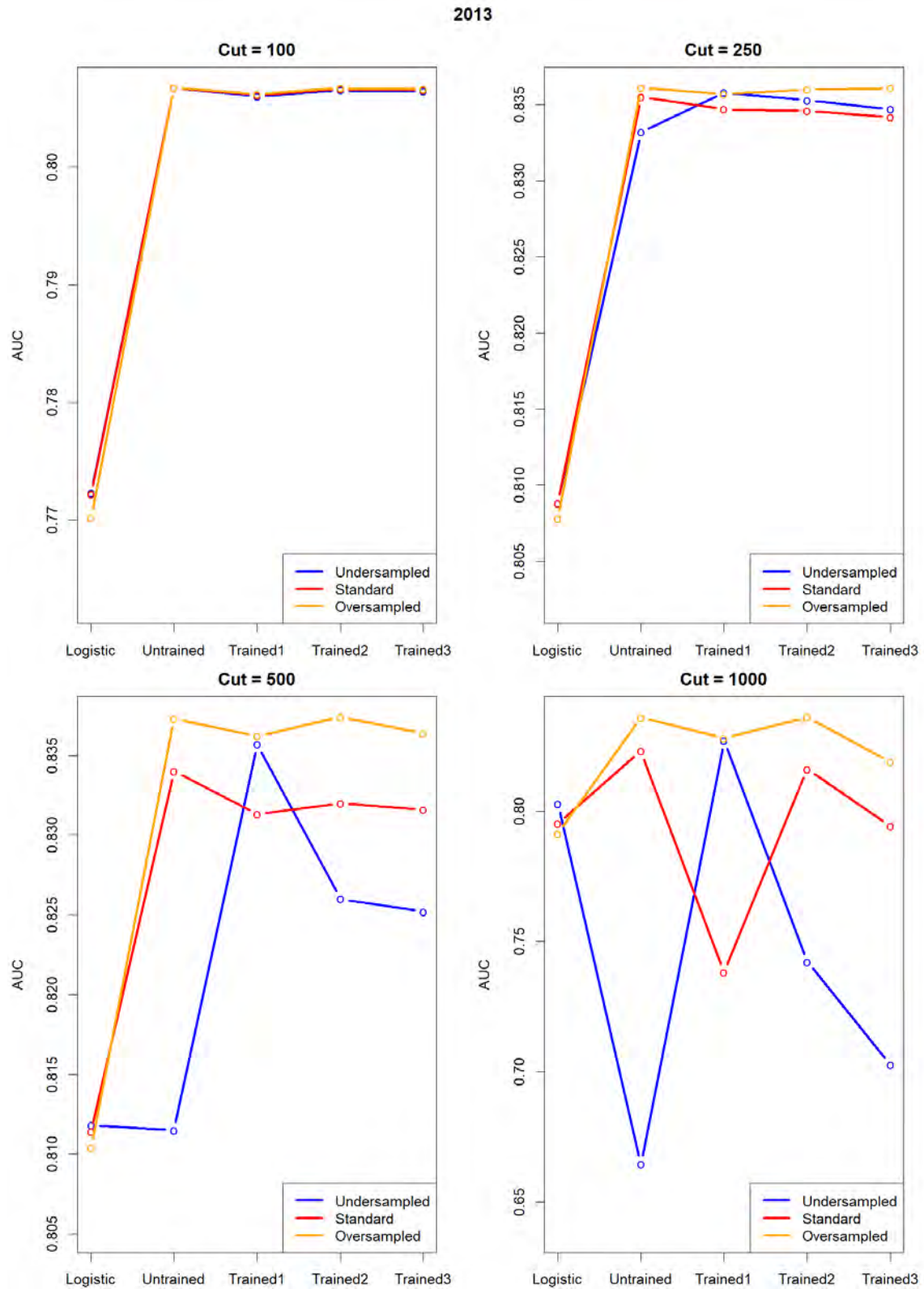
Plots of the AUC values are shown in Figures 10 through 13. The resulting patterns are rather consistent across years. For the \$100,000 and \$250,000 cutoffs, all of the xgboost learners significantly outperformed logistic regression. Also, the sampling method does not seem to have much of an impact. As the number of positive cases decreases (for the \$500,000 and \$1 million cutoffs), oversampling outperforms the other two sampling methods for the xgboost learners. This is less true for trained1, where undersampling essentially works just as well as oversampling. For the other three xgboost learners, undersampling is by far the worst method. Logistic regression performs equally well, regardless of sampling methodology.

**Figure 10**  
Area Under the ROC Curve for Prediction Year 2012

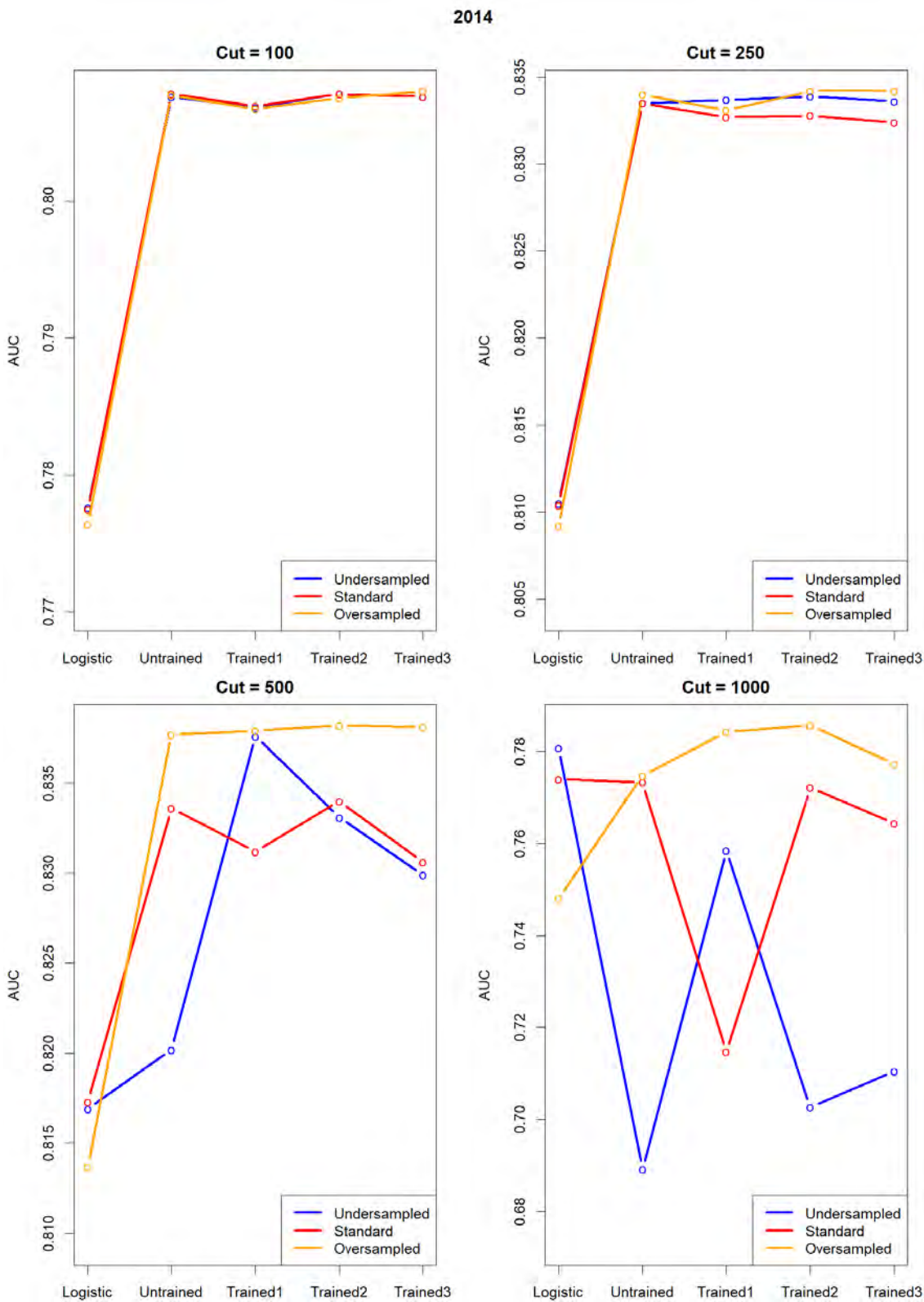




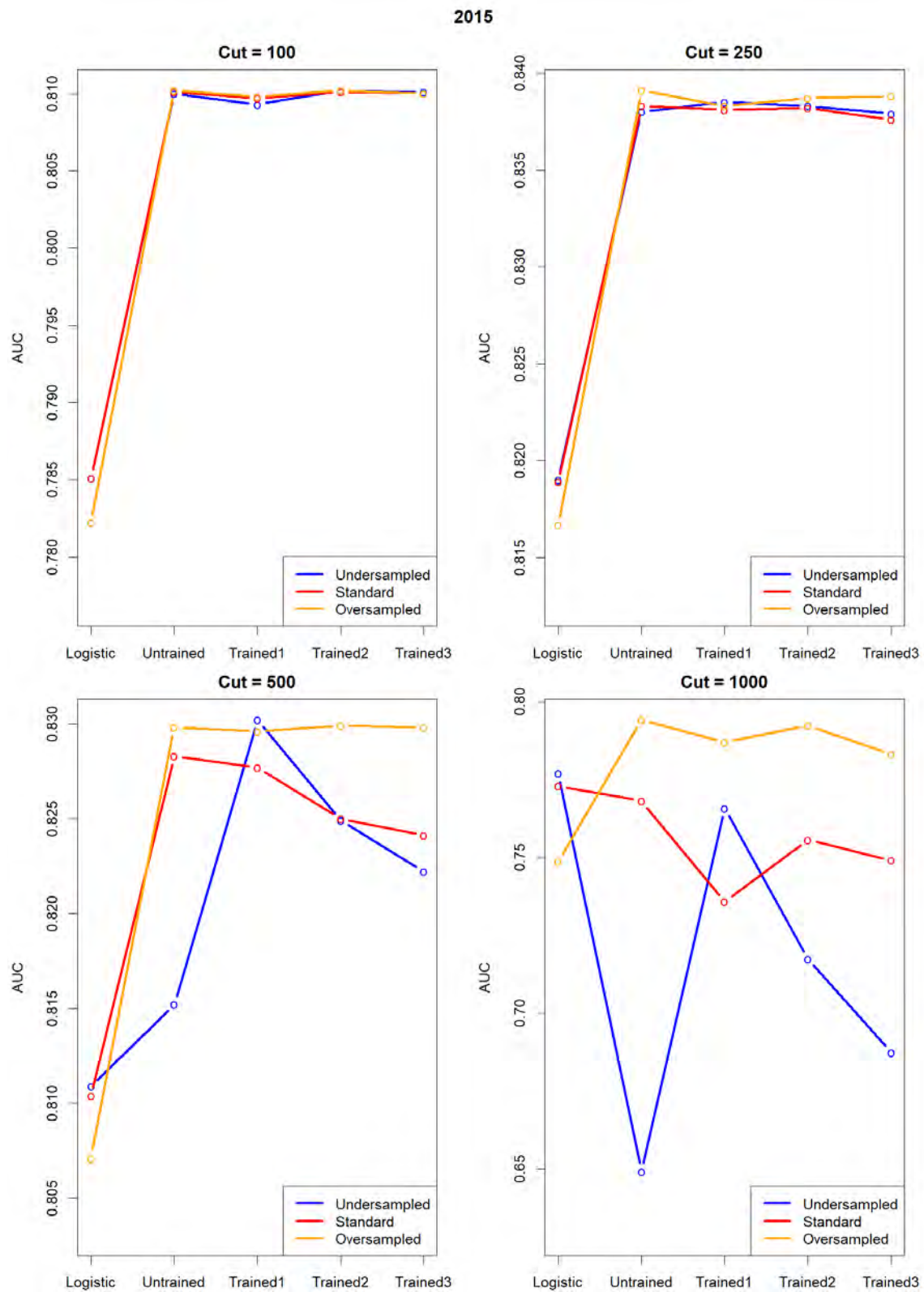
**Figure 11**  
Area Under the ROC Curve for Prediction Year 2013



**Figure 12**  
Area Under the ROC Curve for Prediction Year 2014



**Figure 13**  
Area Under the ROC Curve for Prediction Year 2015



## 5 Conclusion

In this paper, we used HCCI data to examine which characteristics best predict and describe high-cost members. Using a logistic regression, we found that cost history, age, gender and prescription drug coverage all predict high costs, but cost history is the most predictive. In addition to the logistic regression, we compared the predictive accuracy of extreme gradient boosting and found that the added flexibility of this method improved the predictive power. Finally, we showed that with our extremely unbalanced classes, oversampling the minority class provides a better predictive model than undersampling the majority class or using the training data as is.

There are many potential avenues for future work. With the HCCI data, it would be very interesting to explore the many relationships among the members (spatially, temporally and hierarchically) in more depth. It would also be interesting to try to quantify the impact of wellness programs or to explore further the impact of the Affordable Care Act.

## 6 Acknowledgments

The authors acknowledge the assistance of the Health Care Cost Institute and its data contributors—Aetna, Humana and UnitedHealthcare—in providing the claims data analyzed in this study. We are also grateful for the support of the Society of Actuaries, which funded this work. Finally, we thank Brad Barney for his suggestions and advice.

## References

- Bates, D. W., S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar. 2014. "Big data in health care: Using analytics to identify and manage high-risk and high-cost patients." *Health Affairs* 33 (7): 1123–1131.
- Billings, J., and T. Mijanovich. 2007. "Improving the management of care for high-cost Medicaid patients." *Health Affairs* 26 (6): 1643–1654.
- Blumenthal, D., B. Chernof, T. Fulmer, J. Lumpkin, and J. Selberg. 2016. "Caring for high-need, high-cost patients—An urgent priority." *New England Journal of Medicine* 375 (10): 909–911.
- Chen, T., and G. Guestrin. 2016. "XGBoost: A scalable tree boosting system." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York City: ACM, 785–794.
- Fleishman, J. A., and J. W. Cohen. 2010. "Using information on clinical conditions to predict high-cost patients." *Health Services Research* 45 (2): 532–552.
- Garfinkel, S. A., G. F. Riley, and V. G. Iannacchione. 1988. "High-cost users of medical care." *Health Care Financing Review* 9 (4), 41–52.
- Hayes, S., C. Salzberg, D. McCarthy, D. C. Radley, M. K. Abrams, T. Shah, and G. F. Anderson. 2016. "High-need, high-cost patients: Who are they and how do they use health care? A population-based comparison of demographics, health care use, and expenditures." Issue Brief. Commonwealth Fund pub. 1897, vol. 26.  
[https://www.commonwealthfund.org/sites/default/files/documents/\\_\\_\\_media\\_files\\_publications\\_issue\\_brief\\_2016\\_aug\\_1897\\_hayes\\_who\\_are\\_high\\_need\\_high\\_cost\\_patients\\_v2.pdf](https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_issue_brief_2016_aug_1897_hayes_who_are_high_need_high_cost_patients_v2.pdf).
- Hong, C. S., A. L. Siegel, and T. G. Ferris. 2014. "Caring for high-need, high-cost patients: What makes for a successful care management program." Issue Brief. Commonwealth Fund pub. 1764, vol. 19.  
[https://www.commonwealthfund.org/sites/default/files/documents/\\_\\_\\_media\\_files\\_publications\\_issue\\_brief\\_2014\\_aug\\_1764\\_hong\\_caring\\_for\\_high\\_need\\_high\\_cost\\_patients\\_ccm\\_ib.pdf](https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_issue_brief_2014_aug_1764_hong_caring_for_high_need_high_cost_patients_ccm_ib.pdf).
- Joynt, K. E., A. A. Gawande, E. J. Orav, and A. K. Jha. 2013. "Contribution of preventable acute care spending to total spending for high-cost medicare patients." *JAMA* 309 (24): 2572–2578.
- McWilliams, J. M., and A. L. Schwartz. 2017. "Focusing on high-cost patients: the key to addressing high costs?" *New England Journal of Medicine* 376 (9): 807–809.
- Meenan, R. T., M. J. Goodman, P. A. Fishman, M. C. Hornbrook, M. C. O’Keeffe-Rosetti, and D. J. Bachman. 2003. "Using risk-adjustment models to identify high-cost risks." *Medical Care* 41 (11): 1301–1312.
- Moturu, S. T., W. G. Johnson, and H. Liu. 2007. "Predicting future high-cost patients: A real-world risk modeling application." *2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007)*: 202–208.
- Moturu, S. T., W. G. Johnson, and H. Liu. 2009. "Predictive risk modelling for forecasting high-cost patients: A real-world application using Medicaid data." *International Journal of Biomedical Engineering and Technology* 3 (1-2): 114–132.

Schroeder, S. A., J. A. Showstack, and H. E. Roberts. 1979. "Frequency and clinical description of high-cost patients in 17 acute-care hospitals." *New England Journal of Medicine* 300 (23): 1306–1309.

Zook, C. J., and F. D. Moore. 1980. "High-cost users of medical care." *New England Journal of Medicine* 302 (18): 996–1002.

Zulman, D. M., C. P. Chee, T. H. Wagner, J. Yoon, D. M. Cohen, T. H. Holmes, C. Ritchie, and S. M. Asch. 2015. "Multimorbidity and healthcare utilisation among high-cost patients in the US Veterans Affairs Health Care System." *BMJ Open* 5, e007771. doi: 10.1136/bmjopen-2015-007771.

## About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving more than 30,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

**Objectivity:** The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

**Quality:** The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

**Relevance:** The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

**Quantification:** The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries  
475 N. Martingale Road, Suite 600  
Schaumburg, Illinois 60173  
[www.SOA.org](http://www.SOA.org)