R For Improving Consumer Engagement and Health Outcomes

Ken Yale[1] [*], Frank Norman[2], Na'im Tyson[2]

1. ActiveHealth Management, Inc.
2. Knowledgent Group, Inc.
*Contact author: kyale@activehealth.net

Oral Talk Abstract for the The R User Conference 2014

Keywords: R, K-Means, CART, Segmentation, Mining

Background

Health reform shifts our focus from wholesale populations to retail individual consumers. Our care management platform and services has peer-reviewed and published documentation of improved patient care quality and affordability over the past decade, but the new ecosystem requires us to go much further in improving patient engagement, health outcomes and lower costs.

Working with Knowledgent, we utilized an advanced analytical approach using *R* combined with A-B testing and segmentation techniques to increase engagement rates and lower the cost of each engagement.

Methods

Combining internal member data and claims history with externally-purchased lifestyle & behavioral data, we built a segmentation model using K-Means Clustering and CART classification trees with a 1.3 million training population. Clustering showed clearly identifiable micro-segments with common behavioral characteristics based on input factors such as age, health status, socioeconomic factors, purchasing behaviors, technical savvy, education level, geography, etc.

We then conducted live A-B tests, varying messaging & channel usage, and measured response rates. Findings showed differing response rates among segments to the use of digital communication channels vs. traditional communications, for example. In addition, communication design alterations affected response rates for segments in varying degrees.

Results

By optimizing message & channel selection for micro-segments, response rates increased 74 percent.

Mining and visualization of operational data, including time & duration & outcome of calls, produced insights which enabled the cost of each individual engagement to be reduced as well.

R packages used for the segmentation modeling included k-means for the cluster analysis and rpart for analyzing the classification and regression trees.

Models were originally built using an underlying MySQL data store, but have been subsequently ported to access Hadoop data via HiveQL as well.

Conclusion

The improvements identified client savings of $6 million in avoidable health costs and labor cost savings producing a 900% internal ROI. We shall also cover future plans for additional work.