**MU Data Science Capstone – Spring 2018**

**Author(s):** David Hedrick

**Team-member specific contributions:**

I am working alone on this project, and therefore have made 100% of the contribution. I do not anticipate this to change for the duration of this project.

**Progress to Date:**

I proposed an analysis of the effectiveness of the Partners For Kids care coordination program on avoidable emergency utilization for my capstone project. To conduct the analysis, I fitted a logistic regression model to match care-coordinated Medicaid Managed Care members to non-care-coordinated members with similar demographic and clinical attributes based on propensity scores. I chose this method to eliminate selection bias in the care coordination program, which attempts to provide services to high-risk pediatric patients.

The data for this capstone project are data that have routine access to in my job.This has afforded me the unique advantage of knowing exactly what data I have access to, and the scope and limitations of that data. Most of the time spent to date has been spent on data carpentry. This was expected, as health care utilization data is large and complex requiring that it be condensed into usable format appropriate for analysis. For this project, a combination of Medicaid claims, eligibility data, and clinical documentation in the Nationwide Children’s Hospital (NCH hereafter) electronic medical record is used to derive each member-level attribute. My final flat file dataset is comprised of one line per Medicaid Managed Care patient who met continuous eligibility criteria during calendar year 2016.

Once the data carpentry step was completed, the next step was data analysis. Additionally, the project was organized in a deliverable format. The initial propensity scoring and matching of the treated and non-treated populations was completed, and the “closeness” of the match was evaluated. During this time, the code and output within the markdown document was developed. JavaScript and CSS code were inserted into the background of the markdown document to allow the user to toggle visibility of the R code, output and plots. Additionally, the project was split into two separate HTML output files to facilitate the inclusion of an interactive billboard.js chart. This allows users to explore the balance of the covariates within an interactive R Shiny application.

After completing the propensity match and visually inspecting its accuracy, the balance of the covariates between the enrolled and non-enrolled populations was tested by calculating the standardized differences of the covariates. According to Austin and Mamdani (as cited in Faries, Leon, Haro & Obenchain, 2010), “standardized differences less than 0.1 (10%) likely denote negligible imbalance between treated and untreated populations” (p. 57). As seen in figure 1 below, the all covariates had negligible imbalance between the treated and untreated populations.

**Figure 1: Standardized Differences of Covariates**

|  |  |
| --- | --- |
| **Covariate** | **Standardized difference between treated and untreated populations** |
| FRANKLIN\_COUNTY | 0.007 |
| PREVIOUS\_ATTEMPTS | 0 |
| PREVIOUSLY\_ENROLLED | 0.017 |
| NCH\_Previous | -0.016 |
| ABD\_FLAG | 0.013 |
| PREV\_ASTHMA\_ED | -0.003 |
| PREV\_DIABETES\_ED | -0.015 |
| abd\_months | 0 |
| homehealth\_flag | -0.006 |
| USES\_PCP | 0 |
| OVER\_10K\_PAID | 0.042 |
| IP\_LAST\_12\_FLAG | 0.035 |
| ED\_VISIT\_FLAG\_5 | -0.002 |
| NCH\_Inpatient | 0.048 |
| NCH\_ED | -0.013 |
| PRIM\_DX\_HCUP\_CHRONIC | -0.054 |
| PRIM\_DX\_BH | -0.015 |
| NCH\_ATTRIBUTED\_ASSIGNED | -0.008 |
| ABUSE\_NEGLECT | 0.003 |
| NOT\_WITH\_PARENTS | 0.056 |
| ZIP\_PERCENT\_URBAN | 0 |
| ZIP\_PERCENT\_WHITE | 0 |
| AVOIDABLE\_ED\_LAST\_12\_FLAG | -0.036 |

After evaluating the balance, I moved onto evaluation of the results between the enrolled and non-enrolled populations. I also tested the significance of the differences between the two populations in the measurement data. A significant difference was found between the rate of avoidable emergency utilization between the enrolled and non-enrolled populations. Based on the results, I concluded that members in the care coordination do not have fewer avoidable emergency visits in the post-enrollment period. In fact, the enrolled population utilized the emergency department at a rate of 51 visits per 1,000 member months, compared to 45 visits per 1,000 member months for the non-enrolled population.

**Issues Encountered:**

Healthcare organizations must safeguard patient data due to regulations and HIPAA compliance standards. There are legal considerations that limit the use of the claims and eligibility data. The five Ohio Medicaid Managed Care Plans own the claims and eligibility data for covered lives in their plans. However, claims and eligibility data are provided to us as decision support and quality improvement resource. I was required to submit an amendment to a data use agreement I submitted in the summer of 2017 to use data for this analysis. While this caused some delay in getting the project started, I do not anticipate further issues.

Additionally, selecting the covariates to fit the regression model for propensity scoring was challenging. I knew that I should include covariates that affect enrollment, but I did not know if that was sufficient, so I consulted additional research on observational health care studies. According to Austin (as cited in Faries, et al., 2010), four criteria can be used for inclusion of covariates:

* Covariates that affect treatment assignment.
* Covariates that affect both treatment assignment and outcome.
* Covariates that affect outcome, but not treatment.
* Covariates that are related to neither treatment, nor outcome.

Austin also states including only covariates affecting treatment assignment can result in generating fewer propensity score matched pairs and thus less precision in measuring treatment effects. Therefore, I selected additional covariates based on these criteria. Covariates that affect enrollment are in two groups. First, there are utilization-based measures used to target members for potential enrollment. These include greater than $10,000 in paid claims, greater than or equal to five emergency visits in the previous 12 months, an inpatient stay in the previous 12 months, presence of a behavioral health condition and presence of a chronic condition. Second, some covariates are related to treatment assignment due to systemic effects in the program’s outreach process. For example, members who were previously a patient at NCH have a systemic advantage for care coordination outreach over non-NCH patients because a medical record already exists for them. A medical record must be manually created for members we are unable to match from a Medicaid eligibility record to an NCH medical record, while care coordination referral for existing NCH patients is automated. Consequently, over 99% of enrolled members were previously an NCH patient. Other similar covariates that systemically affect treatment assignment include whether the patient was previously contacted for enrollment, whether the patient was previously enrolled, and whether the patient lives in Franklin County.

Demographic information about the member’s zip code is an example of a covariate that affects outcome, but not treatment assignment. For example, members that live in rural areas generally have less access to primary care providers, which increases the incentive to visit an emergency department at a regional hospital to receive immediate care for conditions more appropriate for a lower level of care. Previous emergency utilization with an asthma or diabetes diagnosis is an example of a covariate that affects both treatment assignment and outcome. Members with chronic conditions are specifically targeted for treatment assignment, but a history of using the emergency department for asthma or diabetes could also affect the outcome measurement.

References

Faries, D. E., Leon, A. C., Haro, J. M., & Obenchain, R. L. (2010). *Analysis of observational health care*

*data using SAS* (1st ed.). Cary, NC: SAS Institute.