Discharge Pathways and Readmission Risk

Robert P. Yerex, Ph.D
June, 2015

UVA Medical Center Accountable Care Organization

Objectives: To (1) identify the various post discharge pathways experienced by patients; (2) enumerator those pathways for encoding into statistical learning algorithms, (3) quantify differences in risk for 30 day readmission between pathways at the facility type level, specifically for Skilled Nursing Facilities (SNF).
Data Sources: Claim and Claim Line Feed (CCLF) files provided by the Centers for Medicare & Medicaid Services (CMS).
Analysis Model: Logistic regression was used to test the impact of a number of demographic and health condition variables as well as several variables created to describe the post discharge transition path.
Principal Findings: Age, race, gender, admission type, and Charlson scores all had statistically significant impacts on the probability of readmission within 30 days. Of the transition path variables tested, only the presence of a Skilled Nursing Facility (SNF) was statistically significant with regards to probability of 30 day readmission for patients who transitioned through them. The effect size can be interpreted as patients with an SNF on their discharge pathway being 0.97% more likely to be readmitted. It is critical to note that this is not a cause and effect relationship, rather patients routed through an SNF have a higher CharlsonScore, and as previously shown, a higher CharlsonScore has a very significant relationship to likelihood of readmission.

BACKGROUND

At the time of discharge, a patient may be transferred to a number of different types of health care facilities or may, ideally, go home. The number of transitions through which a patient goes has been shown to have a measureable impact the quality of patient care [1]. Information as to the relationship between the length, complexity, and pathway of post discharge transitions and the risk of readmission within 30 days can be used in making decisions and providing follow-up care to at risk patients in an attempt to reduce the readmission rate.

DATA

The data used for in this study were extracted from the Claim and Claim Line Feed (CCLF) files provided by the Centers for Medicare & Medicaid Services (CMS), to participating Accountable Care Organization (ACO) [2]. Because they are transaction oriented, considerable data manipulation was required in order to transform the CCLF records into a set of event history records each patient describing one or more spells of care that involved one or more episodes of inpatient hospital care. As defined here, an episode of inpatient care can be described as a contiguous period of time when a patient was treated in a specific location for a specified condition. A spell of care is a chain of related episodes in chronological order. The final data set

included 5,364 patients with a total of 10,167 inpatient care episodes encompassed within 9,461 care spells. The earliest inpatient event occurred on 2013/01/01 and the last on 2014/09/17.

ANALYSIS

A logistic regression model was developed with the binary variable; *readmit within 30 days*, regressed on the following independent variables:

Demographic: *PatientAge*, *PatientGender* and *PatientRace*.

Medical Condition: Charlson Comorbidity Indicators (*CharlsonScore)*: These were derived using the entire set of unique diagnoses removing those that were considered to be principle diagnoses. While the scale itself takes on whole number values, it is treated here as a continuous random variable rather than a discrete random variable. This allows for imputation of missing values[1]

Length of Spell (*LOSp*): This was calculated as the total time elapsed (in days) from the beginning of the patients care spell to the end. In addition, the length of stay for each episode was tracked and the ratio of *LOSp* to average episode duration was calculated and used as a feature. When a spell includes only one episode, this value is 1. When there were multiple episodes/claims within a single spell and that spell is long in duration, this value is greater than one. For the patient with 14 episodes within a single spell, the overall *LOSp* was 87 days and the average length of within spell episode was 4.3 resulting in a ratio of 20.2.

Care Pathway:

- Admission Type (*AdmitType*): Emergency, Urgent, Elective, Trauma Center, Unknown. This is for the initial admission that defines the start of the care spell.
- Admission Source (*AdmitSrc*): Physician referral, Clinic referral, HMO referral, Transfer from Other Hospital, Transfer SNF, Transfer Emergency room, Transfer ASC, Transfer Hospice.
- Transition / Discharge Patterns (*Route*): Coding from the original CCLF data files provides 33 codes for patient disposition upon discharge [2]. When there are multiple claims within a spell of care, the sequence of discharge codes describes a sequence of transitions. Nineteen distinct sequences were observed with 67% limited to a single transition, 18% included two transitions, 9% involved three transitions, and 6% involved four or more transitions. The largest number of transitions for a single patient spell was 14. It should be noted that many of these transitions were not care related and could be attributed to billing cycles and other administrative procedural issues. In addition to coding the patterns of transition, a simple count was also included in the model.

[1] For more details on imputation, please see Appendix B

RESULTS

What follows is a brief description and interpretation of the results. Estimates obtained directly from logistic regression are in terms of log odds ratios. For convenience they have been converted back to a probability scale. Effect sizes are described in terms of an increase or decrease in the probability of being readmitted within 30 days, with all other effects held constant. Of the 9,357 spells in the final data set, 1,398 (14.9%) resulted in readmission within 30 days.

Demographic:

- ***PatientAge***: The impact of age was statistically significant ($p < 2.68e-11$). The size of the effect was small, such that a one year increase in *Age* reduced the likelihood of 30 day readmission by 0.98%. This may initially seem counter intuitive but can be explained through investigation of the interaction between *PatientAge* and discharge pathway (represented n the variable *Route)*. As a patient's age increases, they are more likely to be discharged to a secondary care facility rather than home. Readmission from these facilities occurs less frequently as patient's age and die. There are two competing risks; the first of being readmitted within 30 days, the second is death. The cumulative effect of this competing risk (death) has the effect of biasing estimates of regressors that are not independent of a competing risk (death) when the model is formulated to address only the primary risk (readmission within 30 days) [3]. Age and the competing risk of death are highly correlated and this causes a bias in the estimation of the regression coefficient for age.
- ***PatientGender***: Women were 1.1% less likely to be readmitted within 30 days than men. This effect was statistically significant ($p = 0.073$).
- ***PatientRace*****:** This effect of race was statistically significant. When testing hypotheses based on combinations and levels within a categorical variable such as race, within the framework of logistic regression, the Wald Test approximates the Likelihood Ratio Test, but with the advantage that it only requires estimating one model. The Wald test works by testing the null hypothesis that a set of parameters is equal to some value (here 0). In the model being tested, the null hypotheses is that the each race is not different from the base case (white) in terms of likelihood of readmission within 30 day, i.e. that the two coefficients of interest (say Hispanic vs. White) are simultaneously equal to zero. The Wald Test Statistic for across group contrasts (i.e. White = 0, Black = 0, Other = 0, Asian = 0, Hispanic = 0, and North American Native = 0) had an associated p value of 0.057, indicating that at least one of the categories within race was different 0. Individual within category contrasts were used to determine that only the category of North American Native was statistically significantly different from Whites and were 6.5% more likely to be readmitted within 30 days than White patients. For a description of testing hypotheses under the likelihood model see [4]

Medical Condition:

- ***CharlsonScore*****:** Was statistically significant ($p < 8.87e-09$). A 1 point increase in

CharlsonScore was associated with a 1.04% increase in the probability of being readmitted within 30 days.

- The empirical probability density function (PDF)[2], $\hat{f}(x)$ of the variable *CharlsonScore* is shown in Figure 1, and is obviously bimodal. This is an indication of a mixture of populations, with one having a higher median (dashed lines) and mean (dotted lines) *CharlsonScore* than the other. One can also see the shift in the density curve between patients who were and were not readmitted within 30 days which is reflective of the relationship between *CharlsonScore* and the likelihood of 30 day readmission.

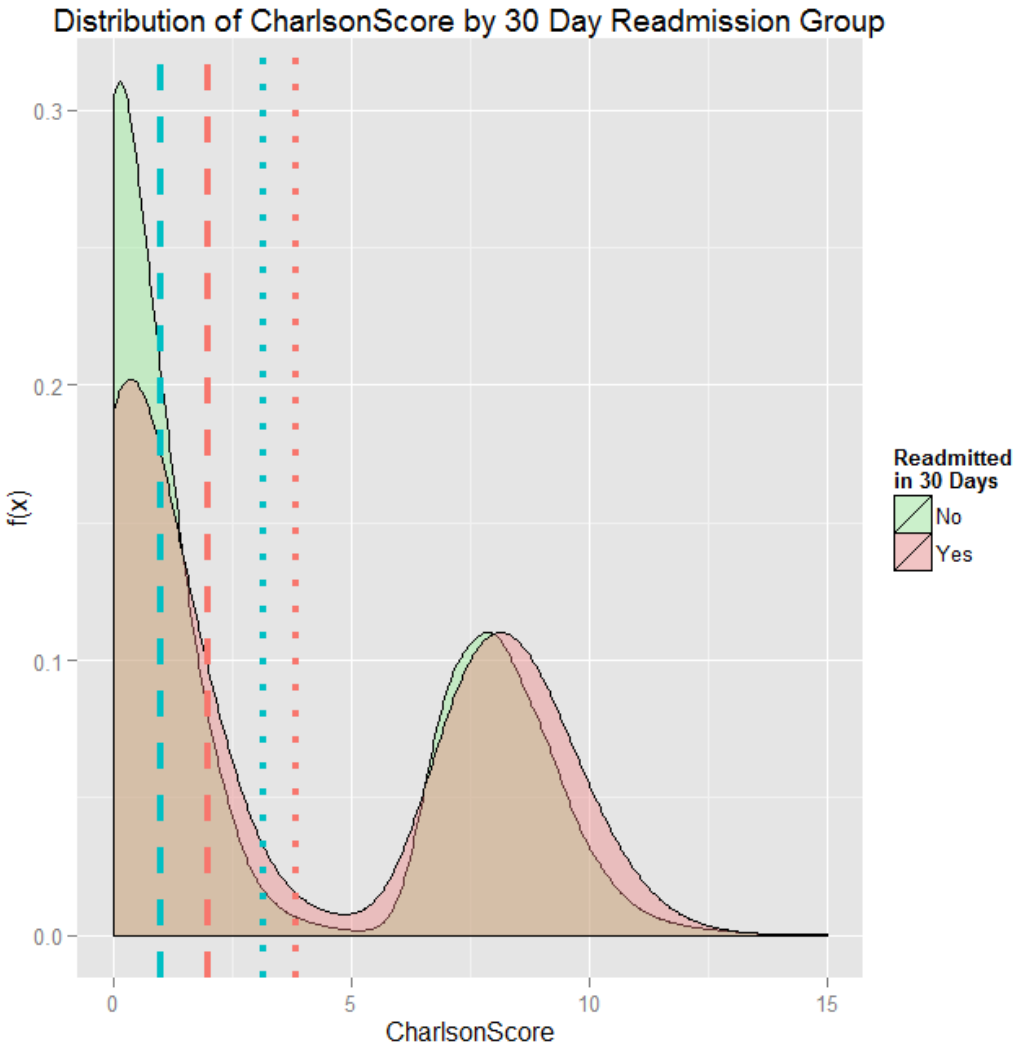

Figure 1: Distribution of CharlsonScore

[2] For those unfamiliar with density functions, please see Appendix A for a brief description.

- Using a Violin Plot, Figure 2 presents an alternative way of visualizing the difference in the distribution of *CharlsonScore* by outcome (readmitted within 30 days or not)

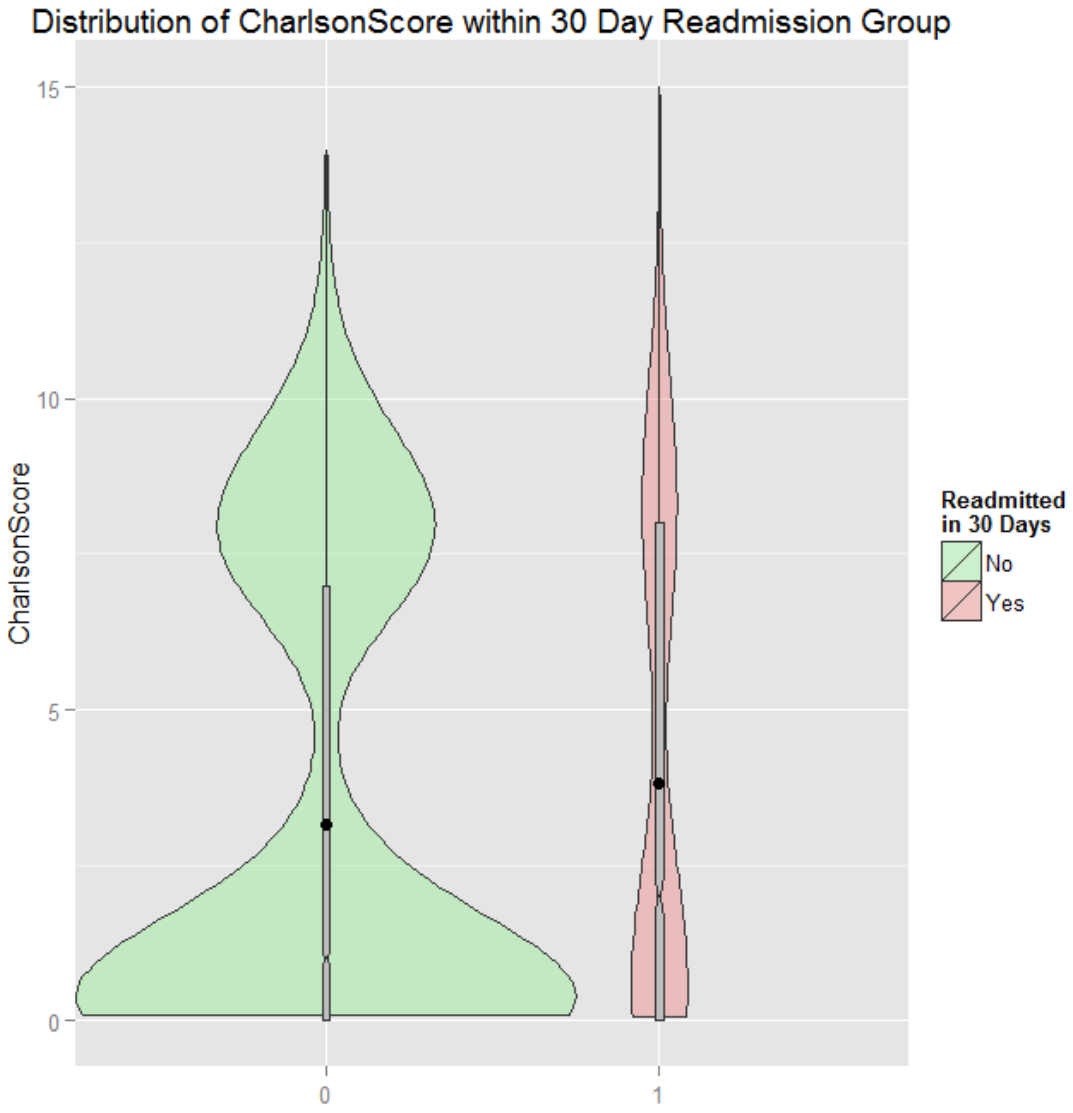

Figure 2: CharlsonScore by Readmission[3]

Length of Spell (*LOSp*):

- The length of the entire care spell was statistically significant with a p value of 0.016, with each addition day being associated with a 1.01% increase in the probability of

[3] The shape of each area is a symmetrical reflection of the shape of the distribution scaled by count. The grey bars within the violin boxes represent the *1QR*, the range between the first and third quartiles (of *CharlsonScore*). The notch in the grey bar represents the median and the black dot, the mean.

being readmitted within 30 days.

Care Pathway:

- ***AdmitType*:** Statistically significant at the across group contrast level with an associated p value of 7.1e-07. As might be expected, the within effect contrasts indicate that patients categorized as *Elective* at admission were 0.65 % less likely to be readmitted within 30 days than patients initially admitted as Emergency. No other between level contrasts was significant.
- ***AdmitSrc*:** Was not statistically significant as an effect
- ***Route*:** The number of unique transition sequences was quite large (103) and it became necessary to collapse these into smaller groups for the model to converge. A categorical variable *Route*, indicating the type of facility to which a patient was transferred was constructed with levels for Skilled Nursing Facility (*SNF)*, Inpatient Rehabilitation Facility (*IRF*), and Intermediate Care Facility respectively (*ICF*). Additional category of *Complex* and *Simple* were coded, where *Complex* indicates a discharge pathway that passes through more than one of *SNF, IRF,* and/or *IRF*, and *Simple* indicating a pathway that passes through none of these.
    - **Interaction Effects (*Route*CharlsonScore*):** There exists a complex interaction between *CharlsonScore* and *Route*. Adding the second order term *CharlsonScore*Route* to the model increased the quality of fit (from an R^2 of 0.11 to an R^2 of 0.14) The *CharlsonScore*Route* term itself was not significant but its addition allocates variance appropriately so that the effect of *CharlsonScore* can be accounted for when examining the pathways represented by the class variable *Route.* Figure 3 presents the distribution of *CharlsonScore* within *Route*. At first examination there appears to be differences between *Simple* and *IRF* as well as between *Simple* and *SNF*, and between *Simple* and *Complex*. However only the between Simple and *SNF* was significant. The area of each column is relative to the counts within those routes. The numbers within the *ICF*, *IRF*, and *Complex* route categories are too small to provide enough statistical power.

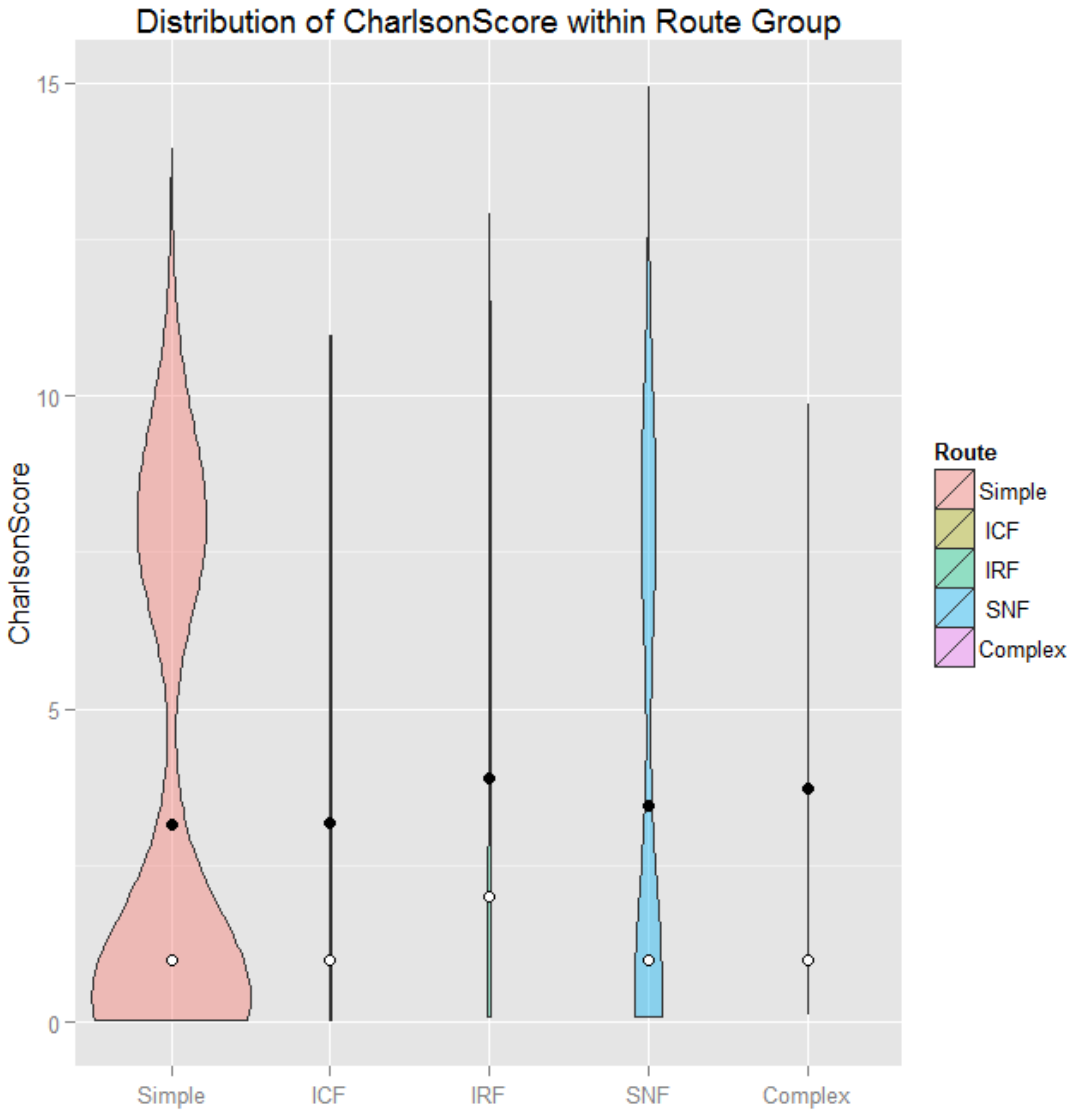

Figure 3: CharlsonScore within Route

- ***Route SNF*:** With the *CharlsonScore*Route* term included in the model, *Route*, itself becomes significant (p = 0.019). Contrasts comparing each of the categories of *Route* (*SNF, IRF, IRC, Complex*) to the base category of *Simple* indicate that only the *SNF Route* is statistically different from *Simple Route* (p = 0.087). The effect size can be interpreted as patients with an *SNF* on their discharge pathway being 0.97% more likely to be readmitted. It is critical to note that this is not a cause and effect relationship, rather patients routed through an SNF have a higher *CharlsonScore*, and as previously shown, a higher *CharlsonScore* has a very significant relationship to likelihood of readmission.

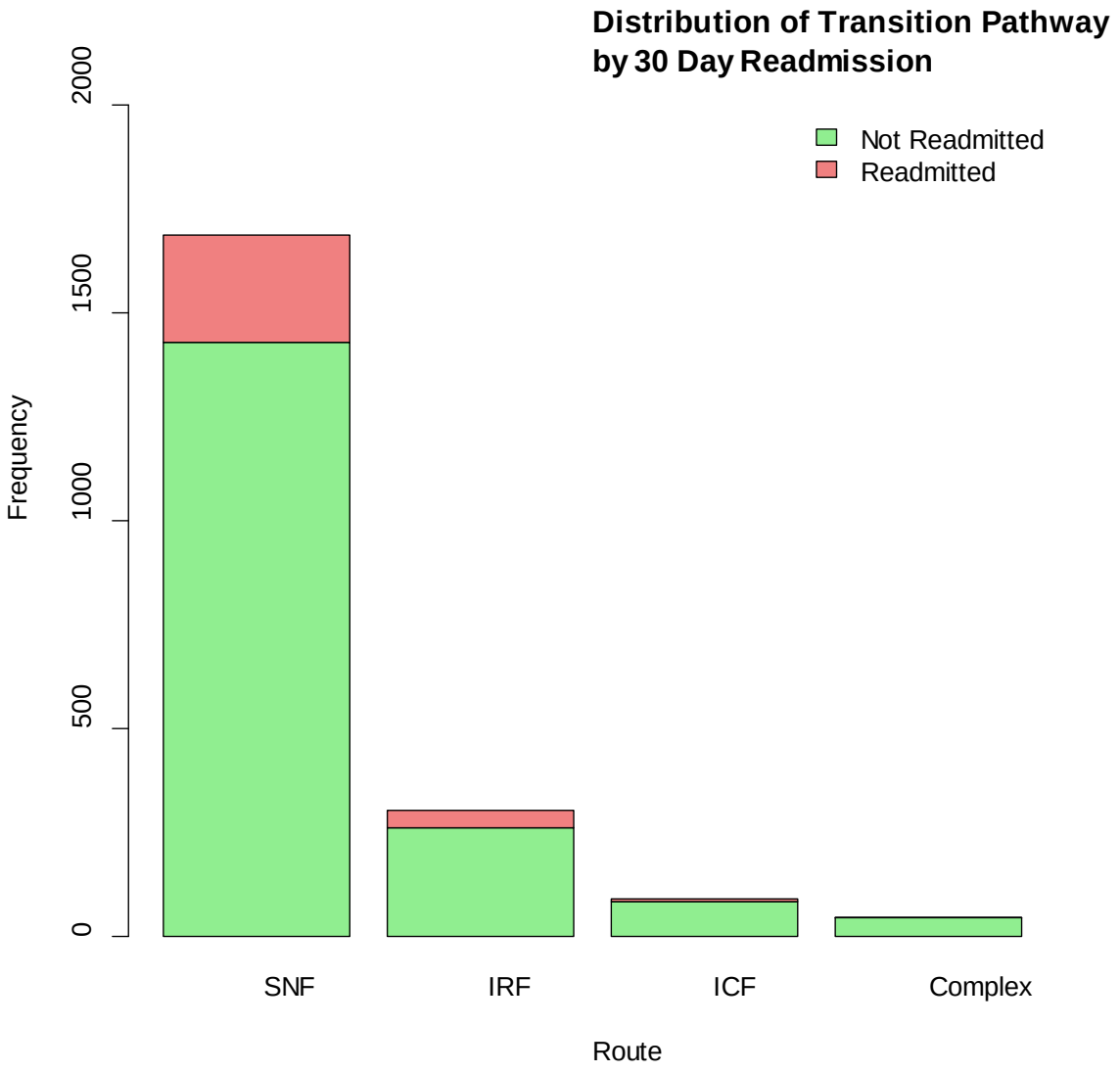

Figure 4: Readmission and Route

The relationship between *Route* and 30 Day Readmission is presented in Figure 4 and Figure 5.

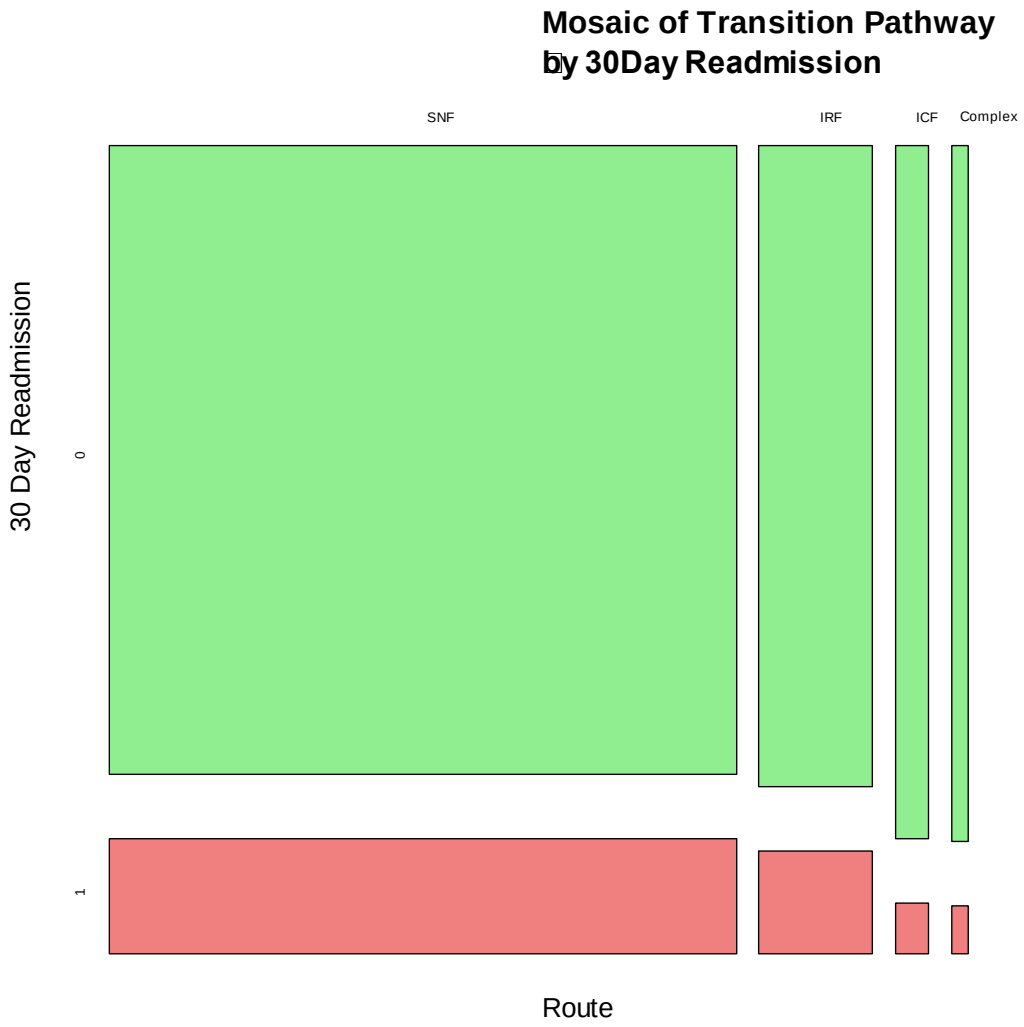

Figure 5: Readmission and Route, Alternate Visualization

UVA Patients routed through an SNF

Of the patients in the data set, 95 were discharged from UVA Health Sciences Center (NPI: 1780630608) and who transitioned through an SNF that could be identified. If a patient transitions to an SNF at which a CMS claim is not filed, the particular facility could not be identified. Table 1 shows a breakdown of those patients and the SNF facilities through which they transitioned.

		Patients Transitioning Through		
Provider ID	**Name**	**Total**	**Readmit(30)**	**%**
1568461705	HOPEWELL HEALTH CARE CENTER	1	0	0%
1003892662	SHENANDOAH NURSING HOME	2	0	0%
1104079243	ENVOY AT THE MEADOWS	2	0	0%
1205888625	GOLDEN LIVINGCENTER - SHENANDOAH VALLEY	2	0	0%
1457375099	AUGUSTA NURSING & REHAB CENTER	2	0	0%
1861416414	KINGS DAUGHTERS COMMUNITY HEALTH & REHAB	2	0	0%
1215010277	AVANTE AT WAYNESBORO, INC	3	1	33%
1043464357	ENVOY AT THE VILLAGE	4	1	25%
1558329532	TRINITY MISSION HEALTH & REHAB OF CHARLOTTESVILLE	6	2	33%
1154378347	GOLDEN LIVINGCENTER - CHARLOTTESVILLE	7	3	43%
1770589079	WESTMINSTER-CANTERBURY OF THE BLUE RIDGE	8	0	0%
1801900170	LOVINGSTON HEALTH & REHABILITATION CENTER	7	1	14%
1932216546	ORANGE COUNTY NURSING HOME & HOME FOR ADULTS	8	2	25%
1023121175	LOUISA HEALTH & REHABILITATION CENTER	15	0	0%
1578519500	THE LAURELS OF CHARLOTTESVILLE	26	2	8%
	Total	**95**	**12**	**13%**

Table 1: UVA Patients Transitioned to SNF

APPENDIX A: PROBABILITY DENSITY FUNCTIONS THEORETICALLY AND EMPIRICALLY DERIVED

In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value. The function does not describe the point probabilities of given values, rather the probability of the random variable X falling within a particular range of values (say between *a* and *b*) is given by the integral of this variable's density over that range, that is, it is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one. This can be more formally stated as:

$$\Pr[a \le X \le b] = \int_a^b f_x(x)dx \tag{1.1}$$

The function $f(x)$ is typically theoretically derived through mathematical proof [5] and presented in a closed form equation, as in the case of the commonly referred to *Normal* distribution which can be stated, in terms of its first and second moments (μ, σ) as:

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.2}$$

Figure 6 presents the shape of the function in (1.2) in the middle and lower graphs with useful properties of distribution expressed in terms, in terms of its first and second moments (μ, σ). The upper graph in Figure 6 presents the associated box plot.

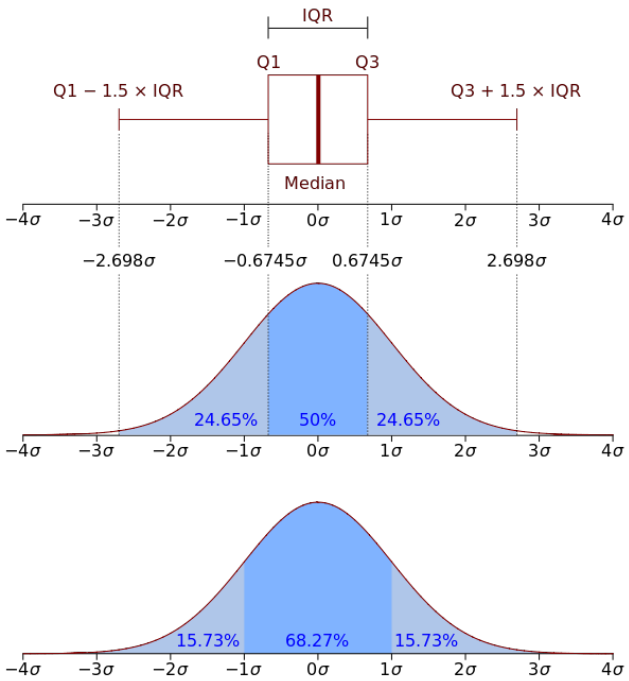

Figure 6: Normal Density Function and Boxplot[6]

An empirical distribution is one derived from observed data with no theoretical constraints. For example, if we measured the heights of several thousand people across a range of ages, ethnicities, etc., the resulting empirical distribution would appear to be quite close to the *Normal* distribution, though no natural phenomenon is an exact match to the *Normal* distribution, in many cases it is close enough. By assuming that the *Normal* distribution is the actual underlying distribution, one can then take advantage of many of the known properties of the *Normal* distribution such as using its moment generating function:

$$e^{t\mu+\frac{1}{2}\sigma^2t^2} \tag{1.3}$$

to derive closed forms for the various moments such as the mean μ, and variance σ^2. Unfortunately analysis of complex phenomenon seldom allows assumption of an underlying theoretical distribution, which is the case with the variable *CharlsonScore*. Without an underlying closed form equation, empirically estimated moments such as the sample mean $\hat{\mu}_x$ and sample variance $\hat{\sigma}_X^2$ are used in making statements regarding the characteristics of a random variable.

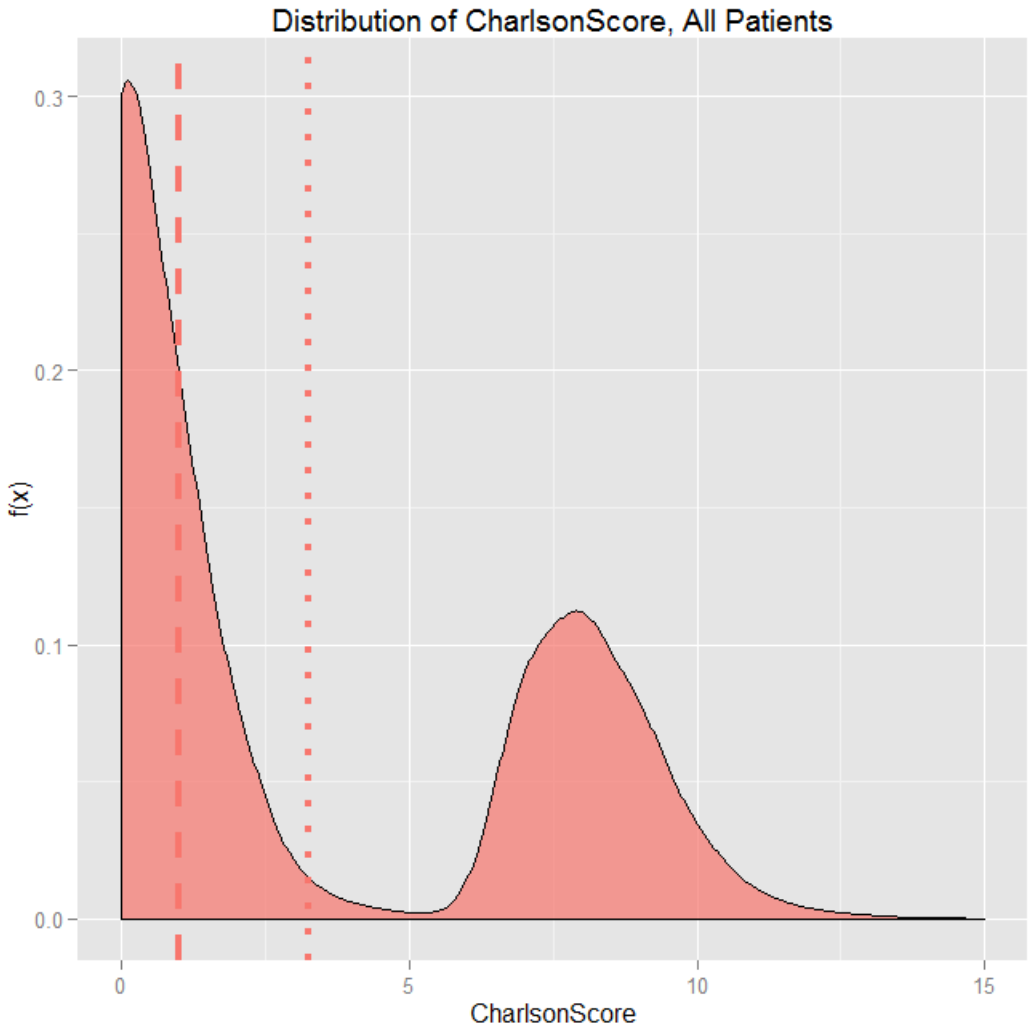

Figure 7: Empirical Distribution of CharlsonScore

The empirical distribution of *CharlsonScore*, across all patients, is shown in Figure 7, where the dashed vertical line represents the estimated sample median and the dotted line represents the estimated sample mean.

APPENDIX B: IMPUTATION OF MISSING VALUES

As with many data sources, there are a number of missing data points in this data set. When an observation is missing a few values, it is possible to impute those values. There are many ways in which missing values can be imputed. For the analysis here missing values were imputed using a technique known as *Imputation by FCS*[7]. This method uses the estimated multivariate distribution of the random variables in the model formula to essentially "fill in the blank" using what is known of the non-missing values. This is one of the more sophisticated methods of imputation available. The reader is referred to Van Buuren et. al. (2006) for details.

REFERENCES

[1] E. A. Coleman, S. Min, A. Chomiak, and A. M. Kramer, “Posthospital Care Transitions: Patterns, Complications, and Risk Identification,” *Health Serv. Res.*, vol. 39, no. 5, pp. 1449–1466, Oct. 2004.

[2] Northrup Grumman Corporation Information Systems, “Accountable Care Organization – Operational System (ACO-OS) Claim and Claim Line Feed (CCLF) Information Packet (IP) – June Release,” Centers for Medicare & Medicaid Services, NGC.ICDA.0301.04.0.0713, Jul. 2013.

[3] R. J. Gray, “A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk,” *Ann. Stat.*, vol. 16, no. 3, pp. 1141–1154, Sep. 1988.

[4] J. J. Faraway, “Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models,” in *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 1 edition., Boca Raton: Chapman and Hall/CRC, 2005, pp. 277–281.

[5] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics, 3rd Edition*, 3rd edition. McGraw Hill, 1974.

[6] J. H., Guch, “Boxplot and probability density function of a normal distribution N(0, σ2).,” *http://commons.wikimedia.org/wiki/File:Boxplot_vs_PDF.svg#/media/File:Boxplot_vs_PDF.svg*, 03-Apr-2015. .

[7] S. Van Buuren, J. P. L., Brand, C. G. M., Groothuis-Oudshoorn, and D. B., Rubin, “Fully conditional specification in multivariate imputation,” *J. Stat. Comput. Simul.*, vol. 76, no. 12, pp. 1049–1064, Dec. 2006.