Using latent class analysis to identify the association between latent class and inpatient length of stay: a case study

Christina Scolieri, ORCID 0000-0001-8416-291X
Omnicell, Inc. 500 Cranberry Woods Drive, Pittsburgh PA 16066
E-mail: Christina.scolieri@omnicell.com
Telephone: (724) 741-6957

Winston W.H. Eng, ORCID 0000-0002-2542-2301
Omnicell, Inc. 500 Cranberry Woods Drive, Pittsburgh PA 16066
E-mail: winston.heng@omnicell.com

Joshua Foley
Omnicell, Inc. 500 Cranberry Woods Drive, Pittsburgh PA 16066
E-mail: Joshua.foley@omnicell.com

Benjamin Cassidy
Omnicell, Inc. 500 Cranberry Woods Drive, Pittsburgh PA 16066
E-mail: benjamin.cassidy@omnicell.com

Daniel Edmonds
Omnicell, Inc. 500 Cranberry Woods Drive, Pittsburgh PA 16066
E-mail: daniel.edmonds@omnicell.com

Conflict of Interest: All contributing authors are employed by and own stock in Omnicell, Inc.

Table of Contents

1 Abstract 3
2 Introduction 3
3 Methodology 3
3.1 Data Description 3
3.2 Elastic Net Regularized Regression 4
3.3 Latent Class Analysis 5
3.4 Zero-Truncated Negative Binomial Regression 5
4 Results 5
4.1 Elastic Net Results 5
4.2 Latent Class Analysis Results 6
4.3 Summary Statistics 9
4.4 Zero-Truncated Negative Binomial Regression Results 9
5 Discussion 10
6 Acknowledgements 11
7 Compliance with Ethical Standards 11
References 11

1 Abstract

The primary objective of this case study was to identify differences in length of stay (LOS) in patients undergoing total hip arthroplasty (THA) that are associated with medications administered within a latent class assignment (LCA). The patient population included all patients who underwent total hip arthroplasty (THA) from August 2017 through June 2018 (n=248) in a 430-bed acute care hospital located in northeastern Ohio. Utilizing Latent Class Analysis to identify medication groupings by patient, and a subsequent zero-truncated negative binomial regression model (ZTNB) analysis, we were able to assess the differences in mean length of stay between latent class assignments. With LC1 as an initial reference point, patients assigned to LC2 had an increase in length of stay of 4.5 days ($P < 0.001$). Setting the reference point at LC2, a decrease of 0.68 days ($P < 0.001$) length of stay was associated with patients assigned to latent class 3 (LC3) instead of LC2. A second iteration of the ZTNB model including eight additional covariates also yielded statistically significant values between latent class assignment and length of stay. The results of each analysis iteration, whether including or excluding covariates, yielded consistent results in statistically significant value differences in the LC1 vs. LC2 length of stay values and LC2 vs LC3 length of stay.

Keywords: length of stay, healthcare outcomes, latent class analysis, diagnostic related group, predictive analytics

2 Introduction

Total Hip Arthroplasty (THA) has historically been reported internationally as burdening patients with an estimated mean length of stay of 6+ days, though national averages vary significantly. [1–3]. In the US, THA accounts for $9 billion in healthcare spending annually [4]. Additionally, demand for THA is expected to grow by 174% by 2030. Recent proposals from CMS which reclassify total knee arthroplasty (TKA) as an outpatient procedure further underscore the need for improvements in operating efficiency. In an effort to reduce costs and improve care quality, recent studies have demonstrated the impact perioperative practice and protocol changes can have in reducing LOS in patients admitted for THA [5]. In the patient population studied, the mean LOS was 2.80. In an effort to optimize patient care, we developed a new approach for identifying areas for improvement using LCA of medications administered to patients during their hospital stay. LCA has been employed in the past to classify patient groups based on behavioral patterns. [6] The large number of cases and the relative homogeneity of the patient population admitted for THA minimized patient specific confounding factors for identifying medication variability. [3,7]. Assessing length of stay outcomes via Diagnosis Related Groups (DRGs) has been tackled previously using generalized linear models or mixture modeling; however, medication usage has never been included [8,9]. Furthermore, existing literature is bereft of predictive analytics and focuses more on a retrospective analysis methodology, and unfortunate problem that prohibits proactive approaches to managing and improving patient outcomes.

3 Methodology

3.1 Data Description

Table 1: Subject Filtering Overview

Filtering Criteria	Retained	Excluded
Starting Count	19,254	NA
DRG = 301	248	19,006

Table 2: Data Dictionary

Variable	Description
Rate of Mortality (RoM)	The likelihood of death [MediLexicon 2018]
Severity of Illness (SoI)	The degree of illness and risk of disease manifested by patients, based either on clinical data from the medical records or on hospital discharge/billing data. Outcome comparisons usually are interpreted in terms of severity of illness to ensure meaningful data interpretations are made [MediLexicon 2018]
Day of the Week (DoW)	The specific day of the week when the patient was in the hospital
Terrestrial Season	Winter: Dec - Mar 21, Spring: Mar 22 - Jun 21, Summer: Jun 22 - Sep 20, Autumn: Sep 21 - Dec 20
Wind Speed	Average daily wind speed, in meters / seconds
Snowfall	Amount of snowfall, in millimeter
Latent Class	
Precipitation	Amount of precipitation, in tenths of millimeter
Temperature	Average temperature for each day, in tenths of degrees Celsius

Initially, the dataset included 19,254 different individuals; when looking specifically at those who had an APR-DRG of 301, there were 248 different subjects (Table). Aside from the medication prescription, other metrics were gathered from the patients and their hospital encounters; more specific measurements (Rate of Mortality, Severity of Illness) as well as environmental factors surrounding their admission (Day of the week, Seasonality, Weather/Temperature) were recorded and subsequently used as covariates within the final association analysis (Table). Protected Health Information was obfuscated prior to any analysis conducted on this dataset. Informed consent was not required in this study as patients were not identified. The environmental factors were generated using the RNOAA and Lunar R packages [10]. Moreover, a set of definitions associated with the weather variables from this package can be found at the official NOAA website and was part of the Global Historical Climatology Network-Daily (GHCN-Daily), Version 3.24 [12].

3.2 Elastic Net Regularized Regression

For the feature selection portion of the data pipeline, an elastic net regularized regression (ENRR) was applied. When dealing with high dimensional data such as situations where the number of independent variables (p) significantly outweighs the number of samples (n), the ENRR is able to narrow down the list of predictors while still accounting for any grouping effect [13]. In essence, by applying penalities to the regression coefficients, ENRR can produce the most parsimonious model with respect to the outcome while still maintaining consistency and computational efficiency [14]. The 'glmnet' R package described in Hastie *et al* 2010 provided the tools necessary to select the medications (Table) utilized within the subsequent latent class analysis [15].

Table 3: Variable Selection Filtering Overview

Filtering Criteria	Retained	Excluded
Starting Count	1906	NA
DRG = 301	549	1357
Missing HIC3 Description	540	9
Maintenance Medication Status	258	282
Elastic Net Regularized Regression	44	214

3.3 Latent Class Analysis

Developed in the 1960s, the latent class analysis (LCA) model aimed to model attitudinal variables with categorical responses [16]. Since its inception, LCA has been a standard and widely utilized tool for analyzing various data sets in the psychological, social, and biomedical sciences. [17]. For the subjects included in this study, latent class analysis (LCA) was performed on their common drug usage as defined in Table . Each medication administered to subjects was subdivided into drug classes utilizing hierarchical classification 3 (HIC3) codes. The vast number of categories in the HIC3 standard versus the approximately 250 different class categorization in the American Hospital Formulary Service (AHFS) list allowed for a much more accurate and concise output. All were chosen from the initial 549 total options found for this specific DRG and were verified by the clinical strategists as being potentially related to the outcome. Additionally, each were dichotomous; patients did ("YES") or did not ("NO") receive the drug as part of their treatment plan.

LCA initially maximizes the log-likelihood function of each specified latent class model via the expectation-maximization (EM) algorithm with a Newton-Ralphson step and subsequently estimates the latent class regression. Initially, the LCA regression was run with up to a max of 15 possible latent classes; the algorithm additionally calculated each model's associated Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). The 'poLCA' R package developed by Linzer and Lewis was used to generate the different latent class models [18]. While LCA models with different number of classes can technically be considered nested models, a likelihood ratio test is not the appropriate tool to distinguish statistically significant differences due to the overarching failure of meeting the regularity conditions [19]. Computing the difference in likelihood between a model with K classes and K-1 classes does not, in fact, follow a χ^2 distribution [20]. Moving forward, the corrected Akaike Information Criterion (cAIC) and adjusted Bayesian Information Criterion (aBIC) were calculated and utilized to the narrow down our selection list from the original 14 options. Moreover, clinical experience and familiarity with the subject proved to be the final decision-maker in terms of determining the appropriate number of latent classes with which to move forward.

3.4 Zero-Truncated Negative Binomial Regression

The outcome variable of interest, *Length of Stay*, included only positive integers within its range of possible values, and the variance of this "count data" variable was larger than its mean. In order to assess the relationship between *Length of Stay* and Latent Class Assignment after adjusing for covariates, the autological Zero-Truncated Negative Binomial Regression (ZTNB) model was performed using tools from the 'VGAM' R package developed by Yee [21]. The additional factors considered included the rate of mortality, severity of illness, day of the week, seasonality, and various weather metrics. Summary statistics for these covariates can be found in Table 5.

4 Results

4.1 Elastic Net Results

Table 4: Medications Retained from Elastic Net Regularized Regression

GENERIC DRUG DESCRIPTION	HIC3 CODE/DESCRIPTION
FERROUS SULFATE 325 MG TABLET	C3B- IRON REPLACEMENT
DEXAMETHASONE NA PHOS INJ 1MG	P5A- GLUCOCORTICOIDS
ADRENALIN EPINEPHRIN INJ 0.1MG	J5A- ADRENERGIC AGENTS, CATHECHOLAMINES
MIDAZOLAM HCL INJ PER 1 MG	H2C- GENERAL ANESTHETICS, INJECTABLE
ROPIVACAINE HYDROCHLORIDE 1 MG	H0A-LOCAL ANESTHETIC
RINGERS LACTATE INF TO 1000CC	C1W- ELECTROLYTE MAINTENANCE
ACETAMINOPHEN 325MG TAB	H3E- ANALGESIC/ANTIPYRETICS, NON-SALICYLATE
AZITHROMYCIN INJ 500 MG	W1D- MACROLIDES
ENOXAPARIN SODIUM INJ 10 MG	M9K- HEPARIN AND RELATED PREPARATIONS
MELATONIN 3 MG TABLET	P6A- PINEAL HORMONE AGENTS
DEXTROSE 5CEFAZOLIN NA INJECT 500 MG	WIW- CEPHALOSPORINS- 1ST GENERATION
FUROSEMIDE INJECTION TO 20 MG	R1M- LOOP DIURETICS
BUPIVACAINE LIPOSOME INJ 1 MG	H0A- LOCAL ANESTHEICS
HALOPERIDOL INJ UP TO 5MG	H70- ANTIPSYCHOTICS, DOPAMINE ANTAGONISTS, BUTYROPHENONES
CALCIUM-VIT D 500MG-200 UNITS	C1F- CALCIUM REPLACEMENT
TRAMADOL 50 MG TABLET	H3A- ANALGESICS, NARCOTICS
METOCLOPRAMIDE 10 MG TABLET	J9A- INTESTINAL MOTILITY STIMULANTS
HEPARIN NA INJ PER 1000 UNITS	M9K- HEPARIN AND RELATED PREPARATIONS
TRANEXAMIC ACID 1000MG/10ML VI	M9D- ANTIFIBRINOLYTIC AGENTS
POLYETHYLENE GLYCOL 3350-REC	D6S- LAXAITVES AND CATHARTICS
PETROLATUM TOPICAL 100D5W INFUSION 1000CC	C5J- IV SOLUTIONS: DEXTROSE WATER
LORAZEPAM INJECTION 2 MG	H2E- SEDATIVE- HYPNOTICS, NON-BARBITURATE
AZITHROMYCIN ORAL, CAP/POW 1GM	W1D- MACROLIDES
PHYTONADIONE (VIT K) PER 1MG	C6K- VITAMIN K PREPARATIONS
CHLORHEXIDINE TOP .12BISACODYL 10 MG SUPPOSITORY	Q3S- LAXAITVES, LOCAL/RECTAL
FUROSEMIDE INJ UP TO 20 MG	R1M- LOOP DIURETICS
CIPROFLOXACIN IVINF INJ 200MG	W1Q- QUINOLONES
HEPARIN SODIUM INJ PER 1000UN	M9K- HEPARIN AND RELATED PREPARATIONS
CALCIUM CHLORIDE 1000MG/10ML S	C1F- CALCIUM REPLACEMENT
SENNA 8.6 MG TABLET	D6S- LAXAITVES AND CATHARTICS
MEROPENEM INJECTION 100 MG	W1S- CARBAPENEMS (THIENAMYCINS)
DOPAMINE HCL INJECTION 40 MG	J5A- ADRENERGIC AGENDS, CATHECHOLAMINES
GUAIFENESIN 600 MG ER	B3J- EXPECTORANTS
AMOXICIL CLAV 875MG-125MG TAB	W1A- PENICILLINS
NOREPINEPHRINE 4MG/4ML SOLUT	J5A- ADRENERGIC AGENDS, CATHECHOLAMINES
SOD BICARBONATE 8.4LIDOCAINE/EPI 1:100,000 1CLOPIDOGREL 300MG TABLET	M9P- PLATELET AGGREGATION INHIBITORS
BENZONATATE 100 MG CAPSULE	H6C- ANTITUSSIVES, NON-NARCOTIC
HYDROCORTISONE TOPICAL 0.5CLINDAMYCIN 150 MG CAPSULE	W1K- LINCOSAMIDES

4.2 Latent Class Analysis Results

Figure 1: Information Criteria for Latent Class Models

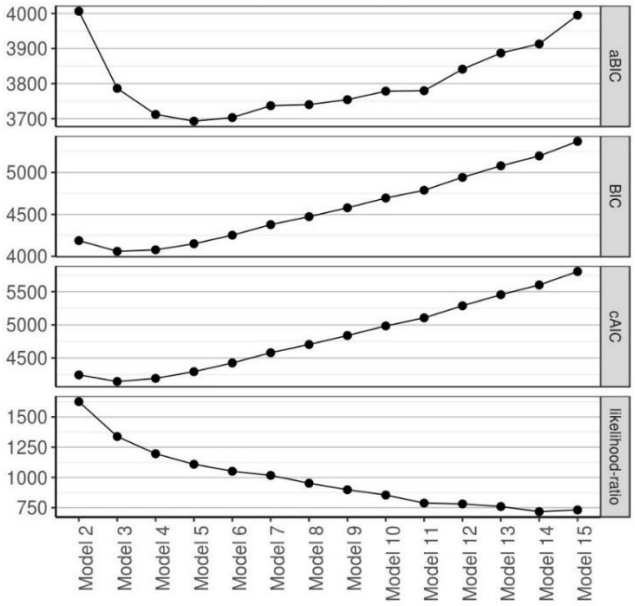


Fig. 2: Drug Usage as Separated by Medication

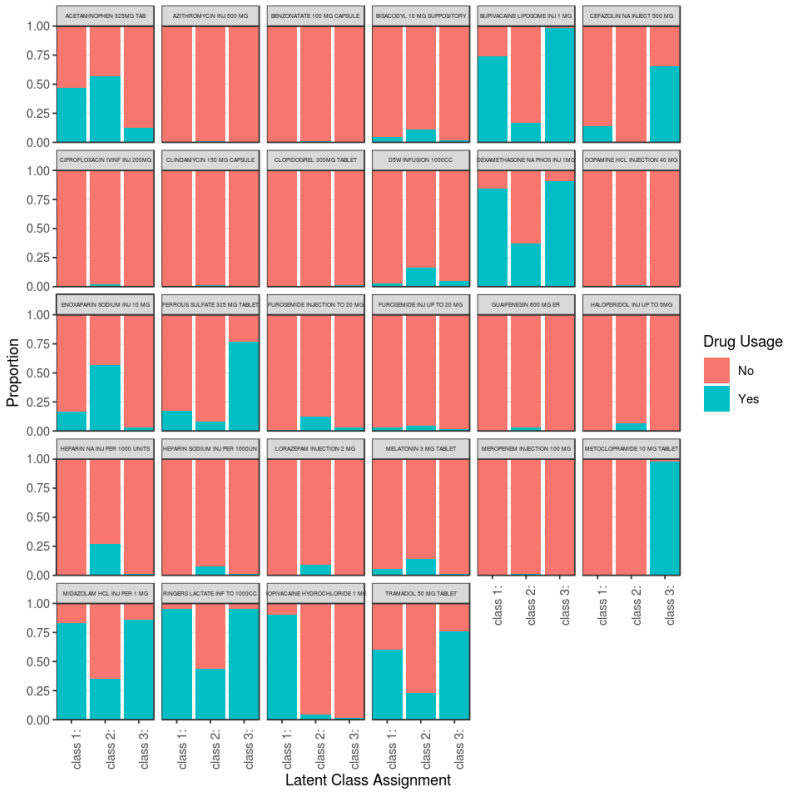


Figure 3: Drug Usage as Separated by Latent Class Assignment

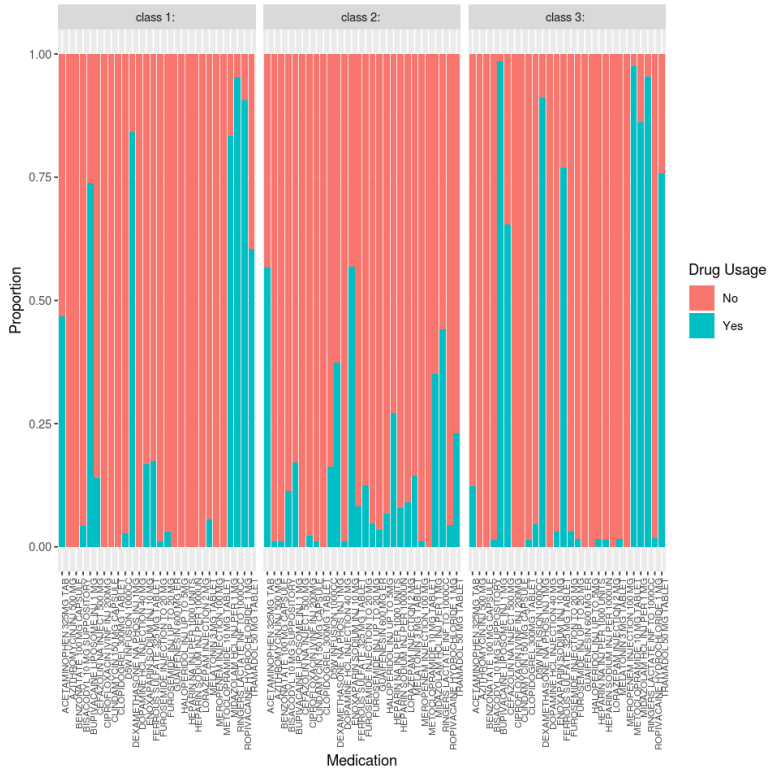
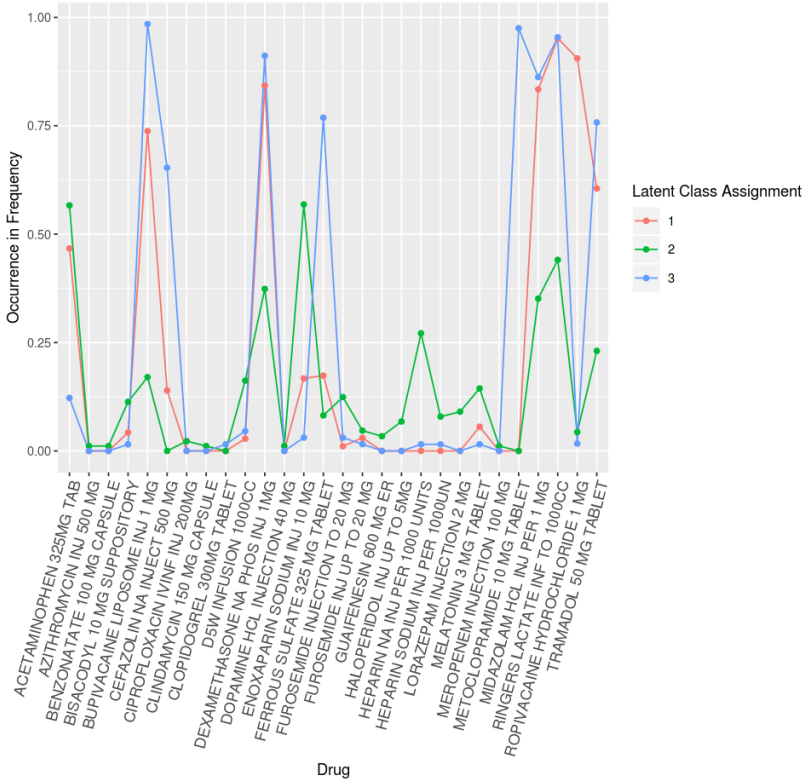


Figure 4: Frequency of Drug Occurrence for DRG 301



4.3 Summary Statistics

Table 5: Summary Statistics

	Latent Class I (n = 94)	Latent Class II (n = 88)	Latent Class III (n = 66)
Rate of Mortality = 1	72 (48.65%)	25 (16.89%)	51 (34.46%)
Rate of Mortality = 2	20 (29.41%)	36 (52.94%)	12 (17.65%)
Rate of Mortality = 3	2 (8.33%)	20 (83.33%)	2 (8.33%)
Rate of Mortality = 4	0 (0%)	7 (87.5%)	1 (12.5%)
Severity of Illness = 1	39 (33.91%)	53 (46.09%)	23 (20%)
Severity of Illness = 2	50 (42.74%)	29 (24.79%)	38 (32.48%)
Severity of Illness = 3	5 (50%)	1 (10%)	4 (40%)
Severity of Illness = 4	0 (0%)	5 (83.33%)	1 (16.67%)
Day of the Week = Monday	26 (32.91%)	14 (17.72%)	39 (49.37%)
Day of the Week = Tuesday	38 (67.86%)	14 (25%)	4 (7.14%)
Day of the Week = Wednesday	6(17.14%)	18 (51.43%)	11 (31.43%)
Day of the Week = Thursday	17 (48.57%)	6 (17.14%)	12 (34.29%)
Day of the Week = Friday	4 (28.57%)	10 (71.43%)	0 (0%)
Day of the Week = Saturday	1 (6.25%)	15 (93.75%)	0 (0%)
Day of the Week = Sunday	2 (15.38%)	11 (84.62%)	0 (0%)
Terrestrial Season = Spring	36 (38.71%)	31 (33.33%)	26 (27.96%)
Terrestrial Season = Summer	11 (40.74%)	9 (33.33%)	7 (25.93%)
Terrestrial Season = Autumn	17 (32.69%)	22 (42.31%)	13 (25%)
Terrestrial Season = Winter	30 (39.47%)	26 (34.21%)	20 (26.32%)
Average Daily Wind Speed	40 (32.52%)	40.5 (32.93%)	42.5 (34.55%)
Snowfall (mm)	0 (NA)	0 (NA)	0 (NA)
Precipitation (tenths of mm)	0(0%)	3 (66.67%)	1.5 (33.33%)
Average Daily Temperature (C)	64 (33.95%)	65 (34.48%)	59.5 (31.56%)

4.4 Zero-Truncated Negative Binomial Regression Results

Table 6: Zero-Truncated Negative Binomial Model (w/ Covariates) Results Given LC1 as Reference

term	estimate	std.error	statistic	p.value
(Intercept):1	-0.2027	0.3401	-0.5961	0.5511
(Intercept):2	1.952	0.4256	4.586	< 0.001
LC2	1.248	0.1905	6.551	< 0.001
LC3	-0.473	0.2612	-1.811	0.0702
Rate of Mortality:2	0.3822	0.1583	2.414	0.0158
Rate of Mortality:3	0.6102	0.2146	2.843	0.0045
Rate of Mortality:4	0.5598	0.4375	1.279	0.2008
Severity of Illness:2	0.04485	0.1631	0.275	0.7833
Severity of Illness:3	0.8114	0.3147	2.578	0.0099
Severity of Illness:4	0.9541	0.5007	1.906	0.0567
Terrestrial Season:Spring	-0.03886	0.1989	-0.1954	0.8451
Terrestrial Season:Summer	-0.005861	0.3015	-0.01944	0.9845
Terrestrial Season:Winter	0.0304	0.1674	0.1817	0.8559
Day of the Week:Monday	-0.4011	0.2524	-1.589	0.112
Day of the Week:Saturday	-0.1165	0.2804	-0.4153	0.6779
Day of the Week:Sunday	-0.2839	0.2951	-0.9622	0.336
Day of the Week:Thursday	0.02212	0.2794	0.07917	0.9369

Day of the Week:Tuesday	-0.328	0.2554	-1.284	0.199
Day of the Week:Wednesday	0.1602	0.2382	0.6726	0.5012
Avg Temperature (C)	0.0006644	0.0009954	0.6674	0.5045
Precipitation	-0.0008073	0.0007479	-1.079	0.2804
Avg Wind Speed	0.00396	0.004249	0.932	0.3514
Snowfall	-0.00107	0.005724	-0.1869	0.8517

Table 7: Analysis of Deviance Table for ZTNB (w/ Covariates)

	Df	**2 * LogLik Diff.**	**Pr(>Chi)**
LC	2	72.71	1.623e-16
Rate of Mortality:	3	10.7	0.01348
Severity of Illness:	3	9.302	0.02553
Terrestrial Season:	3	0.1189	0.9895
Day of the Week:	6	10.65	0.09986
Avg Temperature (C)	1	0.4359	0.5091
Precipitation	1	1.189	0.2755
Avg Wind Speed	1	0.8681	0.3515
Snowfall	1	0.03318	0.8555

5 Discussion

Consideration of the clinical setting and potential patient case mix contextualizes the applicability of the outputs of the models seen in the elastic net regression and latent class model groupings.

For the output of the Elastic Net regression, the retained medications output was reviewed for clinical utility in the patient care setting for our potential patient cohort. As a test for inclusion applicability, the medications in the retained list were evaluated for their association with treatments that could be ordered for patients having this DRG in the inpatient, perioperative, or post-operative settings. The retained list was inclusive of medications that met reasonable use inclusion evaluation by clinical team members. The list did not contain of medications that would have been aberrant to observe in a possible patient regiment during an admission for the DRG of interest.

Evaluation of risk factors for orthopedic surgical patient populations has been the subject of considerable inquiry and presents some common clinical profile modalities that can be evaluated against the outputs of latent class grouping. These include factors such as age, anticoagulation status, choice of anesthetic approach, excessive body weight, pain management needs, and comorbidities such as inflammatory arthritis [23]. Outputs with a five class grouping and three class grouping contained the most discernable medication treatment regimens that could be potential fits for patient population risk modalities.

After considering both groupings, the three class model had the most distinct medication groupings that fit to discernable population groups that would have distinct regimens during their treatment course. Latent Class Group I shows a predominant proportion of ropivacaine utilization, which indicates a specific anesthetic block approach for these patients. Latent Class Group II shows a higher proportion of both heparin injection and enoxaparin indicating anticoagulation needs for this group. Latent Class Group III shows the highest proportion of long acting regional anesthesia, tramadol, and pro-motility agent use indicating pain management needs for this final grouping.

When setting the LC1 as the reference and adjusting for *Rate of Mortality*, *Severity of Illness*, *Terrestrial Season*, *Day of the Week*, *Average Temperature*, *Average Precipitation*, *Average Wind Speed*, and *Snowfall*, we found that

there was a statistically significant difference in *Length of Stay* if an individual had alternatively received a LC2 treatment plan. In fact, being assigned to *LC2* instead of *LC1* increased *Length of Stay* by 3.483 days ($p < 0.001$). When setting the LC2 as the reference, we found that there was a statistically significant difference in *Length of Stay* if an individual had alternatively received a LC3 treatment plan. In fact, being assigned to *LC3* instead of *LC2* decreased length of stay by 0.179 days ($p < 0.001$). There was no statistically significant difference in the length of stay outcome between *LC1* and *LC3* individuals ($p < 0.0702$).

The findings of this study corroborated our original hypothesis that differences in length of stay amongst similar patient groupings created via the LCA model were associated with medication selection practices. Though this study focused specifically on DRG 301, we believe that utilization of the LCA model to identify difference in length of stay and medication selection groups is applicable across various diagnostic related groups. In its current state, the model does not include patient protected health information (PHI) as additional covariates and only encompasses obfuscated billing and claims data. This limitation of our current dataset requires various assumptions to be made regarding patient specifics upon result interpretation. Expansion of this study would be dependent on access to PHI from the hospital's electronic health record (EHR). The addition of PHI factors including but not limited to patient age, gender, weight, height, and lab results would allow for adjustment in our model and more generalizable, evidence-based outputs. Internal steps have recently been taken to ensure Omnicell's maintenance of HIPAA compliance when utilizing PHI data sets. We intend to conduct a follow-up iteration of this study's methodology with the inclusion of various PHI covariates once those policies and procedures are finalized within the organization.

6 Acknowledgements

The authors would like to thank Omnicell for their support of this project. We also extend our sincere gratitude to Aultman Hospital for their partnership and collaboration in this study.

7 Compliance with Ethical Standards

Funding: No funding was obtained to conduct this study.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent: Not applicable, Protected Health Information (PHI) was obfuscated prior to any analysis conducted on this dataset.

References

1. Husted H, Hansen HC, Holm G, Bach-Dal C, Rud K, Andersen KL, et al. What determines length of stay after total hip and knee arthroplasty? A nationwide study in denmark. Archives of orthopaedic and trauma surgery. Springer; 2010;130:263.

2. Jones CA, Voaklander DC, Johnston D, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. The Journal of rheumatology. 2000;27:1745–52.

3. Vorhies JS, Wang Y, Herndon J, Maloney WJ, Huddleston JI. Readmission and length of stay after total hip arthroplasty in a national medicare sample. The Journal of arthroplasty. Elsevier; 2011;26:119–23.

4. Bhattacharyya T, Mehta P, Freiberg AA. Hospital characteristics associated with success in a pay-for-performance program in orthopaedic surgery. JBJS. LWW; 2008;90:1240–3.

5. Stambough JB, Nunley RM, Curry MC, Steger-May K, Clohisy JC. Rapid recovery protocols for primary total hip arthroplasty can safely reduce length of stay without increasing readmissions. The Journal of arthroplasty. Elsevier; 2015;30:521–6.

6. Ahn J, McCombs JS, Jung C, Croudace TJ, McDonnell D, Ascher-Svanum H, et al. Classifying patients by antipsychotic adherence patterns using latent class analysis: Characteristics of nonadherent groups in the california medicaid (medi-cal) program. Value in Health. Wiley Online Library; 2008;11:48–56.

7. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the united states from 2005 to 2030. JBJS. LWW; 2007;89:780–5.

8. Wang K, Yau KK, Lee AH. A zero-inflated poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays. Computer methods and programs in biomedicine. Elsevier; 2002;68:195–203.

9. Singh CH, Ladusingh L. Inpatient length of stay: A finite mixture modeling analysis. The European Journal of Health Economics. Springer; 2010;11:119–26.

10. Chamberlain S. Rnoaa: 'NOAA' weather data from r [Internet]. 2017. Available from: https://CRAN.R-project.org/package=rnoaa

11. Lazaridis E. Lunar: Lunar phase & distance, seasons and other environmental factors [Internet]. 2014. Available from: http://statistics.lazaridis.eu

12. Menne MJ, Durre I, Korzeniewski B, McNeal S, Thomas K, Yin X, et al. Global historical climatology network-daily (ghcn-daily), version 3.24. NOAA National Climatic Data Center. 2012;

13. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). Wiley Online Library; 2005;67:301–20.

14. De Mol C, De Vito E, Rosasco L. Elastic-net regularization in learning theory. Journal of Complexity. Elsevier; 2009;25:201–30.

15. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model via coordinate descent. Journal of Statistical Software [Internet]. 2011;39:1–13. Available from: http://www.jstatsoft.org/v39/i05/

16. Lazarsfeld PF. Latent structure analysis. Psychology: A study of a science. McGraw-Hill New York; 1959;3:476–543.

17. McCutcheon AL. Latent class analysis. Sage; 1987.

18. Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. Journal of Statistical Software [Internet]. 2011;42:1–29. Available from: http://www.jstatsoft.org/v42/i10/

19. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. Structural equation modeling. Taylor & Francis; 2007;14:535–69.

20. McLachlan GJ, Baek J. Clustering of high-dimensional data via finite mixture models. Advances in data analysis, data handling and business intelligence. Springer; 2009. pp. 33–44.

21. Yee TW. Vector generalized linear and additive models: With an implementation in r. New York, USA: Springer; 2015.

22. Yee TW, Wild CJ. Vector generalized additive models. Journal of Royal Statistical Society, Series B. 1996;58:481–93.

23. Memtsoudis SG, Poeran J, Zubizarreta N, Rasul R, Opperer M, Mazumdar M. Anesthetic care for orthopedic patientsis there a potential for differences in care? Anesthesiology: The Journal of the American Society of Anesthesiologists. The American Society of Anesthesiologists; 2016;124:608–23.

24. Prokopetz JJ, Losina E, Bliss RL, Wright J, Baron JA, Katz JN. Risk factors for revision of primary total hip arthroplasty: A systematic review. BMC musculoskeletal disorders. BioMed Central; 2012;13:251.

25. MediLexicon. Mortality rate. medilexicon: Stedman's medical dictionary. wolters kluwer health. [Internet]. 2004–2018 [cited 2019 Sep 19]. Available from: https://www.medilexicon.com/dictionary/75887

26. MediLexicon. Severity of illness. medilexicon: Stedman's medical dictionary. wolters kluwer health. [Internet]. 2004–2018 [cited 2019 Sep 19]. Available from: https://www.medilexicon.com/dictionary/43516