Analyzing the Composition of Diabetes Patients and Impact of Seasonal and Climate Trends on Emergency Room Utilization in Central Virginia

1st Bradley Katcher
School of Data Science
University of Virginia
Charlottesville, USA
bk5pu@virginia.edu

2nd Elizabeth Driskill
School of Data Science
University of Virginia
Charlottesville, USA
ekd6bx@virginia.edu

3rd Jiaxing Qiu
School of Data Science
University of Virginia
Charlottesville, USA
jq2uw@virginia.edu

4th Wendy Novicoff, Ph.D.
Professor of Public Health Sciences
University of Virginia School of Medicine
Charlottesville, USA
WMN2V@hscmail.mcc.virginia.edu

***Abstract*—Diabetes is an epidemic both nationally and in the Commonwealth of Virginia, and there are gaps in understanding of what demographic groups are most impacted by diabetes and how these patients utilize the emergency room. It is also known that diabetes patients are more likely to experience dehydration at high temperatures, which could potentially lead to heat exhaustion or heat stroke. However, there is limited research on the effect of climate on the number or proportion of diabetes patients presenting to the emergency room. The main objective of this project will be to examine trends in emergency room utilization for patients with diabetes in Virginia, specifically targeting seasonal and climate trends, giving emphasis to exploring trends during heat and cold waves.**

I. Introduction

More than 30.3 million Americans are currently living with diabetes, and another 84.1 million are living with prediabetes [1]. Nearly one in every ten Virginians (approximately 631,000 individuals) have diabetes [2]. While research recognizes the negative health impacts that environmental stressors have on patients with diabetes [3], there is a gap in the literature with regards to direct impacts of climate on climate-related ED visits for patients with diabetes. Heat waves have been associated with increases in acute renal failure, cardiovascular disease, diabetes, electrolyte imbalance, and nephritis, including significant morbidity factors surrounding these diseases [4]. This analysis adds to this research by exploring the composition of diabetic patients who go to the ER and exploring the impact of climate on ER visits.

This study used administrative emergency room records from seven hospitals within two hospital systems in the Commonwealth of Virginia to analyze the composition of patients that presented to the emergency room with diabetes-related diagnoses (based on ICD-9 and ICD-10 codes) and studied the relationship between their ER visit and climate conditions. The data span from 2010 to 2017 and include 1,856,886 observations. Using patient zip codes, American Community Survey data was linked [5] to explore the demographics and social determinants of health associated with diabetic patients who present to the ER for care. Additionally, this study analyzed how various climatological events impact diabetes patients' utilization of the emergency department. A predictive model was employed to estimate the number of diabetic patients presenting at the emergency room given certain climate conditions. Inferential statistics were used to gain further insights about how different climate variables may affect diabetes patients' usage of the emergency room. Using this information, this research provides recommendations to hospitals about how their patient population may change during certain climatological events so that they can be better prepared to provide care in the future, as well as suggests areas for future research.

II. Data

The raw administrative emergency room records consist of 1,856,886 observations. Any observation with a zip-code outside of the 2XXXX range was removed (i.e. below 20000 or above 29999), as this represents a region geographically close to Virginia. This leaves 1,791,918 observations, roughly a three-and-a-half percent decline. Of this pool of patients, approximately 1.2% came to the ER for diabetes-related reasons, or 21,902 observations. It is important to note that these patients could be repeat visitors to the ER, but the data did not identify that.

Patients were identified as diabetic if their primary diagnosis had specific ICD-9 or ICD-10 codes. The ICD-9 codes associated with diabetic-related illness include the prefixes of 249.XX or 250.XX or certain other ICD-9 codes or prefixes.[1] For ICD-10 codes, the following prefixes were included: V180, V771, V653 and a variety of additional ICD-10 codes, mostly of group E.

After accounting for missingness and deaseasonalizing the data (specific procedure contained in the methods section), 15-year bins for age and one-hot encoded indicator variables for types of insurance by hospital system were created, and then aggregated the number of patients by diabetic-status by day and hospital to obtain the count of diabetic and non-diabetic patients and average of the four climate variables. One, two, and three-day lags for each of the climate variables by day and hospital were also constructed. The final, aggregated dataset used for regression analysis, therefore consisted of 20,395 daily observations from seven hospitals (2,922 individual days), with a maximum of 289 non-diabetic patients and 11 diabetic patients at any given hospital on a single day.

III. Methods

A. Data Engineering

1) Imputation

There were a significant number of missing values within the administrative data, most prevalent for climate-related variables. The missing values were imputed with SARIMA(p,

[1]The following prefixes or codes: 775.1, 648.0, 253.5, 588.1, 790.2, 751.7, 357.2, 362.0, 775.6, 337.1, 353.5, 536.3, 775.0, 648.8, 271.4, 731.8, 275.0, 251.0, 251.2, 251.1, 707.1, 362.0, 775.0, 780.0, 251.0, 276.2, 251.1, 713.0, 713.5, 366.41, 443.81, 581.81, 583.81, 790.29, 362.01, 362.07, 362.02, 362.03, 362.04, 362.05, 362.06

d, q, P, D, Q, s) models. 'p' and seasonal 'P' indicated the order of autoregressive terms, 'd' and seasonal 'D' indicated how many differentiations the model takes, 'q' and seasonal 'Q' indicated the order of moving average terms, and 's' indicated the seasonal pattern's length. The models for temperature, apparent temperature, vapor pressure, and wet bulb temperature were sarima(1,2,3,1,0,0,7), sarima(1,2,3,1,0,0,7), sarima(1,2,3,0,0,0,0), and sarima(1,2,3,1,1,1,7), respectively.

Some observations within the data lacked indicators for demographic characteristics such as race and age. In these cases, race was left missing, and thus was not included in aggregate levels, making the assumption that the race data was missing at random. For age data, missing age values were assigned to the mean age across the data, 42 years.

The scope was narrowed to only include patients with zip codes within the range 20000-29999. A significant portion of the missingness within the dataset came from patients whose zip codes were outside of this range. The main population of interest was patients within central Virginia, so this further justified the decision to remove patients not located in this area. In the climate analysis, the scope was limited to patients with the following weather stations: Bluefield, Charlottesville, Louisa, Lynchburg, Martinsburg, Pulaski, and Shenandoah Valley.

2) Seasonality

The scope of analysis was limited to the relationship between climate variables and the number/proportion of diabetes patients presenting to the ER, and not other seasonal factors that could potentially have acted as confounding variables. To remove seasonality from the dataset, the average of each climate variable for each day within a year across all years included in the dataset was calculated. For example, the average temperature on January 1st for each hospital between the years of 2010-2017 was computed. Then, this average value was subtracted from each individual day's climate features to appropriately scale the data and remove seasonality. This process was applied for the following variables: temperature, apparent temperature, vapor pressure, wet bulb temperature, as well as the count of diabetes and non-diabetes patients.

Additionally, one day and three day lags were incorporated into the analysis for each of the four climate variables. To understand the impact of heat and cold waves, indicators for absolute and relative heat and cold waves were created. Absolute heat waves were defined as at least three consecutive days with temperatures greater than 27 degrees Celsius, and cold waves were defined as at least three consecutive days with temperatures less than 0 degrees Celsius. Relative heat waves were defined as an instance of any of the previous three days having a temperature in the top 5% of all temperatures for that hospital system, whereas relative cold waves were the same for the bottom 5% of temperatures.

B. Exploratory Data Analysis

In addition to basic data visualization techniques, several logistic regression models were employed to conduct initial exploratory data analysis, specifically in regards to demographic features. Indicator variables were created via one-hot encoding to characterize different races and genders as well as to classify whether or not a patient presented with diabetes-related symptoms based on ICD-9 and ICD-10 codes. To test whether there was a difference in the proportion of a given race or gender between populations of diabetic and non-diabetic patients, separate logistic regression models were run for each of the following races: White, Black, Hispanic, Asian, Biracial, Pacific Islander, American Indian, Unknown, Other, and Patient Refused. In each regression, the given race was the predictor variable in the following equation:

$$\log\left(\frac{\Pr(Diabetic)}{1-\Pr(Diabetic)}\right) = \beta_0 + \beta_1 Race_i + \epsilon_i \qquad (1)$$

For sex, the following logistic regression was run to determine if males or females represent the disproportionate amount of diabetic-related visitors to the ER:

$$\log\left(\frac{\Pr(Diabetic)}{1-\Pr(Diabetic)}\right) = \beta_0 + \beta_1 Sex_i + \epsilon_i \qquad (2)$$

In addition to basic demographic data analysis, temporal differences were explored, utilizing t-tests to explore differences in the proportion of diabetes-related patients presenting to the ER on different days of the week. The temporal analysis was also expanded to year and month. Other descriptive trends were employed visually and statistically in exploring driving distance to the UVA emergency room from the patient's home zip code by utilizing Google Maps Distance Matrix API (implemented using the gmapsdistance package in R) to the University of Virginia Emergency Room (1215 Lee St, Charlottesville, VA 22908).

A number of variables from American Community Survey data regarding various social determinants of health were linked on patient's zip codes in order to explore the socioeconomic conditions from where the patients were coming. These included the proportion of individuals falling within certain income levels (as a percentage of the federal poverty line), insurance type/status, access to a vehicle, individuals on SNAP (food stamps), as well as median income levels. The means of each of these variables were compared for individuals who presented with diabetes-related illnesses to those who did not.

Finally, climate data measured throughout central Virginia was analyzed. Overall, the goal was to explore the relationship between the number and/or proportion of diabetes patients presenting to the ER and these other features in order to narrow the scope and identify potential variables of interest that could be utilized in future model building and statistical analysis.

C. Inferential Statistics

Inferential statistics were used to find whether the proportion of diabetes patients presenting during extreme weather conditions (defined as apparent temperature greater than 27 degrees Celsius or less than 0 degrees Celsius) were significantly different from the overall population proportion. The entire dataset was assumed to be the population. The null hypothesis, H_0, was that patients observed at extreme weather groups would follow the same distribution as the population.

P-values were computed and compared to a significance level of $\alpha = 0.05$. The population parameter's value was given by the overall diabetic proportion, and the standard error of each weather group was given by the equation:

$$se_i = \frac{\sigma}{\sqrt{n_i}} \tag{3}$$

where i was the indicator of weather group i, σ was the population standard deviation, and n_i was the number of observations in group i.

Considering the time effect, the same test was run on 1, 2, and 3-day lagged weather features separately.

D. Regression Analysis

In order to predict the number of diabetic patients that will be seen at a given hospital on a given day, the analysis employed a variety of count models. Using an 80-20 train-test split, each of the various regression models were run (multivariate linear, negative binomial, and Poisson) to predict the number of diabetic-related patients that a given hospital can expect to see, as well as to identify what factors influence the number of patients that a hospital can expect to see. In modelling, these methods were very similar to those adopted by Lay et. al. [6] in linking morbidity and environmental data and using a negative binomial model to model the daily ER visit count. Like Lay et. al., a Bayesian updating approach was not employed. Disaggregating by hospital has a potential power and sample size issue, so instead, hospitals were pooled together for a majority of the results.

The relationship between various deseasonalized climate factors and the number of diabetic patients was explored, and then the sensitivity to inclusion of indicators for heat waves and cold waves, demographic characteristics, temporal factors, and insurance status was tested. Using the models learned from the training data, predictions were made for the number of diabetic patients that will present with the testing data, and measures of accuracy (the sum of true positive and true negative classifications divided by the number of observations) were compared to the non-informative rate for all regression models. The impact of the various factors on the outcome of interest was explored and discussed in the results section. Each model was run for all hospitals combined. One-day lags were used for each of the climate-related variables in the analysis. When a diabetic patient experiences extreme weather, it may take several hours to one day for this to result in symptoms that may require them to present to the emergency room. Because of this, it is more likely that a patient will present to the ER on the day following extreme weather rather than the day of, so it was hypothesized that 1-day lag would be a better predictor variable than the actual climate features for a specific day and this was incorporated into the analysis.

IV. Results and Discussion

A. Exploratory Data Analysis

The logistic regression results demonstrated that, compared to females, males have a higher relative risk of presenting to the ER with diabetes-related symptoms.

Based on logistic regression results, it was determined that the relative risk of presenting to the ER with diabetes-related symptoms is highest for black patients and lowest for white patients in both UVA and Carilion hospital systems. Within UVA, the difference in relative risk among races is larger compared to Carilion. It is known from the literature that diabetes disproportionately affects black patients compared to white patients, but these results showed that they have a greater risk of presenting to the ER for diabetes-related issues as well. It should also be noted that patients presenting to the ER for diabetes-related symptoms are often uninsured; this affects the black population disproportionally. Additionally, UVA is a state-supported hospital and is required to treat all patients regardless of insurance status.

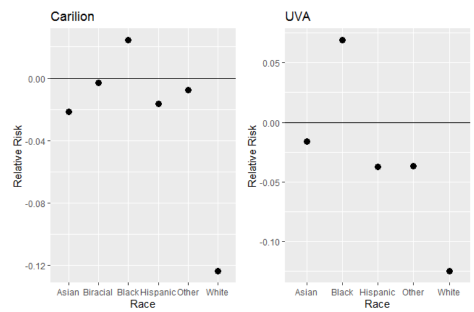

Fig. 1. Relative risk between races.

The results showed that, within the Carilion health system, there was a significantly higher proportion of diabetes patients presenting to the emergency room on Friday and a significantly lower proportion on Sunday. Within the UVA health system, there was a significantly higher proportion presenting on Tuesday and Thursday and a significantly lower proportion on Saturday and Sunday.

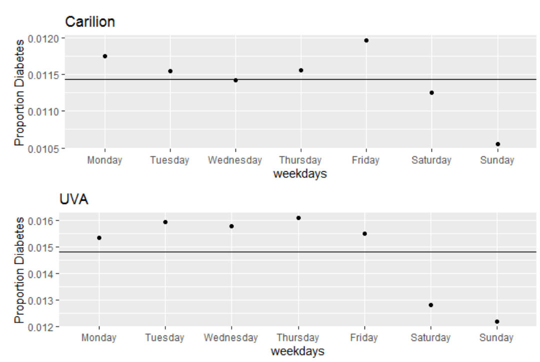

Fig. 2. Difference between weekdays.

The age distribution for patients with diabetes-related symptoms peaks at an older age compared to the distribution for patients without diabetes-related symptoms. This was expected, as type 2 diabetes tends to affect the older population. These results were similar between the two hospital systems.

Another variable explored was the distance that diabetic and non-diabetic individuals travel to go to the emergency room. The results in Table I showed that patients with diabetes who

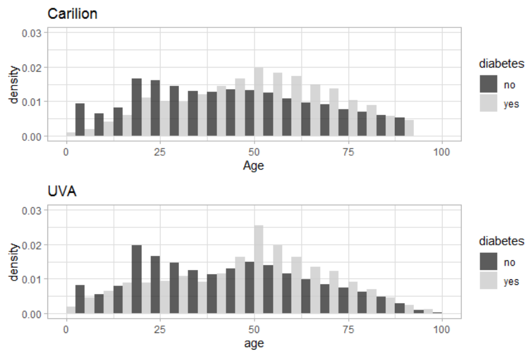

Fig. 3. Difference in age distribution.

go to the UVA emergency room travel between 3.9% and 10.1% further than their non-diabetic counterparts.

Table II shows the breakdown of a variety of ACS variables that were linked on patient zipcode [5]. Accordingly, the conclusions that are drawn represent the areas where the patients come from, not the individual patients themselves. The results demonstrated that at the Carilion health system, patients who present with diabetes are more likely to have lacked access to a vehicle, whereas the reverse is true for patients at the UVA health system. Regarding socioeconomic factors, patients who present with diabetes-related illness came from areas that have higher rates of individuals on SNAP at both UVA and Carilion, however, the base rate at Carilion was much higher (around 13% compared to around 8%) than at UVA. This pattern mimics that of median income, with patients presenting with diabetes coming from areas with lower median incomes both at UVA and Carilion, but the baseline median incomes were much lower for Carilion (around \$48,000 compared to around \$60,000 at UVA). When income was decomposed among low income levels as a percentage of the Federal Poverty Level (FPL) (<50% FPL, <100% FPL, <125% FPL, <150% FPL, <200% FPL), stark trends appear between the two hospital systems. In the UVA health system, non-diabetic patients are more likely to be in lower income groups than diabetic patients, but the reverse is true for the Carilion health system. Breaking this trend down further by hospital within the Carilion health system, the results indicate that the Carilion income patterns hold true for Carilion Franklin Memorial Hopsital (CFMH) and Carilion Roanoke Memorial Hopsital (CRMH), but little to no difference is observed between diabetic and non-diabetic patients for Carilion Giles Community Hopsital (CGCH), Carilion New River Valley Medical Center (CNRV), Carilion Stonewall Jackson Hopsital (CSJH), and Carilion Tazell Community Hopsital (CTCH) (with CSJH and CTCH occasionally having a higher proportion of individuals from low-income areas than diabetic patients) seeing diabetic patients coming from the area. This seems to suggest large heterogeneity between hospitals within the Carilion system and supports the regression findings that hospital location plays an important role in determining the impact of diabetic-related illnesses.

With regards to insurance, for the UVA health system, diabetic patients were more likely to come from areas with higher rates of coverage from Medicare, Medicaid, or no insurance, whereas non-diabetic patients were more likely to come from areas with higher rates of private insurance (employer, direct purchase), or Tricare. Within the Carilion health system, diabetic patients were more likely to come from areas with higher rates of Medicaid, but lower rates of direct purchase, Medicare, and Tricare, and it appears there is no statistically significant difference between diabetic and non-diabetic patients among uninsurance rates. Given that disadvantaged populations are more likely to present to the ER, Medicaid coverage and uninsurance rates were explored by hospital within the Carilion system. Among the six hospitals, only one (CRMH) had diabetic patients coming from a region with statistically higher rates of Medicaid coverage than non-diabetic patients. The same is true for uninsurance, with the only hospital being CGCH. The remainder of hospitals had no significant difference in area insurance rates from which diabetic and non-diabetic patients came. This once again supports the conclusion that, in future modelling, hospitals should be treated separately, rather than pooled, in order to account for differential effects.

In plotting the percent difference in ER visits between diabetes patients compared to non-diabetes patients as it relates to apparent temperature, two significant peaks are observed around the 0th and 100th percentiles of apparent temperature. Smaller peaks are also observed near relatively extreme temperatures below the 20th percentile and above the 80th percentile. Diabetes patients were more likely to present to the ER during extreme hot and cold apparent temperatures, and additional analysis was conducted to explore this idea further.

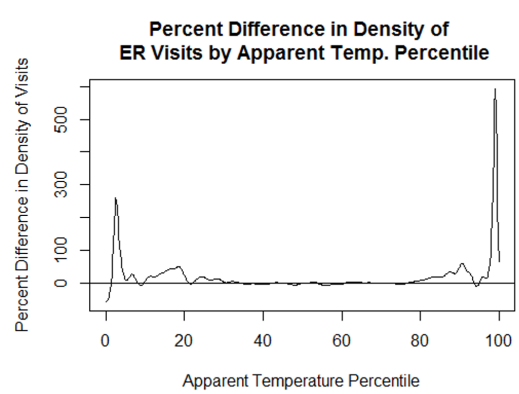

Fig. 4. Difference in apparent temperature.

B. Regression Analysis

All emergency rooms were pooled together and the number of diabetes-related emergency room visits was explored as a function of the lagged climate features and indicators for relative and absolute heat waves. These model predictions (negative binomial, poisson, and linear) were not significantly different than the 'no information' rate, but interesting conclusions could be drawn from the results. For all three

TABLE I
DRIVING DISTANCES BY DIABETIC STATUS: UVA PATIENTS

	Driving Distance				
	Distance (mi.): Diabetic Patients	Distance (mi.): Non-Diabetic Patients	Diff. (mi.)	Pct. Diff	P-Val
All Patients	31.78	30.59	1.19	3.9%	0.003
Within 100 mi.	26.92	25.42	1.50	5.9%	<.001
Within 50 mi.	21.19	19.24	1.95	10.1%	<.001
Within 25 mi.	11.12	11.77	0.65	5.8%	<.001

models, there were statistically significant increases in visits when there were absolute heat waves at the 0.001 level, and statistically significant decreases in visits when there were absolute cold waves at the 0.05 level, but the same pattern did not hold true for relative heat and cold waves. When the the model was re-run and the indicators for heat and cold waves were omitted, the negative binomial and linear models showed a statistically significant increase in visits associated with an increase in the lagged temperature. Lagged temperature remained significant if indicators were removed for heat and cold waves for most of the specifications.

Across model types, when demographic variables were included as controls, the climate variables made much less of an impact, while gender, race, and age were significant. Being female made an individual less likely to present to the ER for a diabetes related incident, as did being white, and younger (under the age of 30). These results were robust across the three different classes of models evaluated. Upon introducing temporal factors, the same patterns were observed with the relative importance of demographic factors, and it was observed that on weekends (Saturday and Sunday), fewer individuals presented to the ER with diabetes-related illnesses.

When fixed effects were incorporated for each location, significantly different baseline likelihoods of presenting to the ER existed, suggesting that there were differential effects for various hospitals within these systems. However, the relative importance of race and age remained (the more Hispanics increased the predicted number of individuals presenting, with whites still presenting less, but losing statistical significance), and younger individuals remaining less likely to present. The temporal trends also held true, but the models indicated a slight uptick in the number of individuals presenting on Mondays. Future research should explore possible differential treatment effects for each hospital system. Due to time and resource constraints, such analysis was outside the scope of this project. Weather factors still had little influence on the outcome, with the addition of these controls. However, it appeared that an increase in lagged temperature slightly increased the number of individuals presenting, and an increase in lag wet bulb temperature slightly decreased the likelihood of presenting.

Lastly, the inclusion of health insurance was explored as a feature in the analysis. Given the structure of the data and the ways that different hospitals code payer status, separate models were run for UVA and Carilion systems. At UVA, the likelihood of diabetic-related illnesses presenting was increasing in the proportion of individuals with Medicare, and individuals without insurance (no insurance specified or marked as self-pay). For the Carilion system, the likelihood of presenting was decreasing in the proportion of individuals with Medicaid and increasing in the proportion of individuals with commercial insurance. The same weather trends appeared, i.e temperature associated with an increase and wet bulb temperature associated with a decrease, and increased vapor pressure also caused a slight increase. However, upon the inclusion of facility fixed effects for the Carilion health system, there was a lack of significance for insurance types, suggesting heterogeneous effects across hospitals within the system, and this was recommended as an area for further analysis. Detailed regression results can be found in the online appendix via: https://bit.ly/2RGOJXc

The main takeaways from the regression analysis were three-fold. First, temperature had a slight positive impact on the number presenting and increased wet bulb temperature had a slight negative impact, but these results were sensitive to the inclusion of other controls. Absent demographic, temporal, or insurance controls, heat and cold waves were strong predictors, but this effect seemed to disappear with controls. Weather had a slight impact, but the relative impact of that compared to other controls was hard to discern. Second, demographic controls seemed to play a significant role, especially gender and age, and this seemed relatively robust to the inclusion of other controls and model forms. Insurance also sometimes had an impact, depending on the inclusion of facility as a control. Temporal factors had some impact, i.e. on weekends, fewer diabetic-related illnesses were seen, but factors such as the year did not make an impact (as expected). Finally, varying results were observed when accounting for different baseline characteristics of facilities. This suggested that the hospitals within the system matter in terms of how each of these factors analyzed impacted the presenting of diabetic-related illnesses, and more research should be devoted to individual hospital analysis, exploring heterogeneous treatment effects.

C. *Inferential Statistics*

The H_0 hypothesis that there were no significant differences between the diabetes proportion within each extreme weather group and the population proportion was rejected at the 0.05 significance level. i.e. The proportion of diabetes patients at extreme temperatures, which was defined as temperatures above 27 and below 0 degrees Celsius, was significantly higher than the overall proportion of diabetes patients, which was 0.01217. The test results are shown in the table below. For each climate feature, the sample range, number of patients, percentage of diabetes-related patients and p-values were listed.

V. CONCLUSION

While predictive models produced inconclusive results, a significant relationship was found between extreme temperatures, defined as temperatures above 27 degrees Celsius and below 0 degrees Celsius, and the proportion of diabetes-related ER visits. This suggests that diabetes patients may

TABLE II
SOCIAL DETERMINANTS OF HEALTH BY SYSTEM AND DIABETIC STATUS: LINKED BY ZIP CODE

	UVA			Carilion		
Condition	Non-Diabetic	Diabetic	P-val	Non-Diabetic	Diabetic	P-val
Lack access to a vehicle	3.53%	2.89%	<.001	2.90%	3.14%	<.001
On SNAP (food stamps)	8.37%	8.82%	<.001	12.99%	13.51%	<.001
Median Income	$60,813.28	$60,195.63	<.001	$48,400.98	$47,985.17	<.001
<50% FPL	3.87%	3.58%	<.001	4.31%	4.48%	<.001
<100% FPL	9.05%	8.44%	<.001	11.09%	11.46%	<.001
<125% FPL	12.40%	11.98%	<.001	14.68%	15.09%	<.001
<150% FPL	15.88%	15.56%	<0.001	19.07%	19.47%	<.001
<200% FPL	23.08%	23.05%	0.835	28.83%	29.16%	<.001
Employer	49.23%	48.35%	<.001	44.70%	44.71%	0.893
Direct Purchase	9.53%	9.27%	<.001	6.31%	6.15%	<.001
Medicare	5.07%	5.21%	<.001	7.20%	7.16%	0.130
Medicaid	8.16%	8.48%	<.001	10.66%	11.00%	<.001
Tricare	1.28%	1.19%	<.001	0.94%	0.91%	<.001
VA	0.29%	0.32%	<.001	0.42%	0.45%	<.001
Two Types of Insurance	16.15%	16.47%	<.001	18.96%	18.74%	<.001
Uninsured	10.28%	10.71%	<.001	10.81%	10.88%	0.690

TABLE III
INFERENTIAL STATISTICS TEST RESULTS

	Temperature(C)				Apparent Temperature(C)				Vapor Pressure(hPa)				Wet Bulb Temperature(C)			
Lag	Range	No.	Pct	P.val	Range	No.	Pct	P.val	Range	No.	Pct	P.val	Range	No.	Pct	P.val
0	<0	126083	1.3023%	0.003	<0	144710	1.2977%	0.003	<3	109169	1.3447%	<0.001	<0	275508	1.2911%	<0.001
0	>27	6119	1.3024%	0.03	>27	108609	1.2927%	0.011	>23	159666	1.2651%	0.004	>27	54	1.8519%	0.335
1	<0	126083	1.2499%	0.146	<0	144710	1.2314%	0.313	<3	109169	1.2174%	0.499	<0	275508	1.2573%	0.028
1	>27	6119	1.3503%	0.002	>27	108609	1.3093%	0.002	>23	159666	1.2614%	0.054	>27	54	0%	0.207
2	<0	126083	1.1770%	0.095	<0	144710	1.1813%	0.116	<3	109169	1.2036%	0.338	<0	275508	1.2130%	0.416
2	>27	6119	1.3058%	0.025	>27	108609	1.3130%	0.002	>23	159666	1.2445%	0.162	>27	54	0%	0.207
3	<0	126083	1.1833%	0.134	<0	144710	1.1810%	0.103	<3	109169	1.2082%	0.391	<0	275508	1.2352%	0.197
3	>27	6119	1.3298%	0.006	>27	108609	1.2955%	0.009	>23	159666	1.3121%	<0.001	>27	54	1.8519%	0.335

be more likely to present to the emergency room for care when temperatures are significantly above or below average. This finding could potentially be useful to hospitals, as they can expect to see an increased proportion of diabetes patients when temperatures become extreme. Knowing this information will help hospital staff to better care for this possible increase in the proportion of diabetes patients by allocating additional resources and staff in preparation. The findings might also be helpful to diabetes patients within central Virginia, as it might encourage them to stay indoors or take more extreme measures to avoid contracting symptoms that may require them to present to the emergency room for care.

With linear constraints in order to preserve interpretability, this research was unable to build a model that predicts accurately, but this is still useful information. Although the relationship between climate variables and the proportion of diabetes patients presenting to the ER was significant based on inferential statistics, this relationship may not be strong enough to be able to produce accurate predictions for the number of diabetes patients that will present to the ER on a given day. Therefore, while hospitals may expect an increase in diabetes patients during extreme temperatures, this increase may not be large enough to warrant many additional resources or other preparations. Regardless, the findings can still help hospitals gain a better understanding of the patient population presenting to the emergency room with diabetes-related symptoms as well as the types of climate factors that may exacerbate their symptoms, which will ultimately make them better suited to provide care.

This research could be extended by examining areas beyond the scope of central Virginia as well as by looking into other diseases or comorbidities that are commonly seen in the emergency room and how they may relate to climate. Additionally, more complex models could have been useful in prediction; however, even though more complicated models may predict slightly better than these simpler models, they might be difficult for many people to understand. For this reason, the research focused on more interpretable models. A significant portion of the time was spent on data cleaning and dealing with the missingness within the dataset, and researchers working with this dataset in the future will not have to do quite as much of this, so they can hopefully spend more time performing statistical analysis, building different types of predictive models, and exploring other interesting relationships beyond the scope of this analysis.

REFERENCES

[1] National Center for Chronic Disease Prevention and Health Promotion. (2017). National Diabetes Statistics Report, 2017 (p. 20). Retrieved from Center for Disease Control and Prevention: Division of Diabetes Translation website: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

[2] Virginia Department of Health. (2019). Diabetes and Prediabetes. Retrieved September 18, 2019, from Data website: http://www.vdh.virginia.gov/diabetes/data/

[3] Cook, C. B., Wellik, K. E., Fowke, M. (2011). Geoenvironmental Diabetology. Journal of Diabetes Science and Technology, 5(4), 834–842. https://doi.org/10.1177/193229681100500402

[4] Knowlton Kim, Rotkin-Ellman Miriam, King Galatea, Margolis Helene G., Smith Daniel, Solomon Gina, … English Paul. (2009). The 2006 California Heat Wave: Impacts on Hospitalizations and Emergency Department Visits. Environmental Health Perspectives, 117(1), 61–67. https://doi.org/10.1289/ehp.11594

[5] U.S. Census Bureau; American Community Survey, 2017 American Community Survey 5-Year Estimates, Tables B27010, B17026, S0802, S2201, S1901; generated by Bradley Katcher; using data.census.gov; ¡https://data.census.gov/cedsci/¿; 29 December 2019).

[6] Lay, C. R., Mills, D., Belova, A., Sarofim, M. C., Kinney, P. L., Vaidyanathan, A., et al. (2018). Emergency department visits and ambient temperature: Evaluating the connection and projecting future outcomes. GeoHealth, 2, 182–194. https://doi.org/10.1002/2018GH000129

[7] Rene, S. (n.d.). Change in Temperature Can Affect Blood Sugar Levels. Retrieved September 17, 2019, from Piedmont Healthcare website: https://www.piedmont.org/living-better/change-in-temperature-can-affect-blood-sugar-levels?ga=2.34494754.201619031.1568745883-1785235297.1568745883