A Look at Parent-Adolescent Relationships & E-Cigarette Usage in Minnesota on a County and Individual Level

Fiona Adams, Sofia Pozsonyiova, Quinn Rafferty

December 16, 2019

Abstract

In recent years, e-cigarette usage and vaping have increased at an alarming rate in the United States among adolescents. According to the 2015 National Youth Tobacco Survey (NYTS), an estimated 27.1% of U.S. adolescents, representing approximately 7,260,500 persons, had ever tried e-cigarettes (Health, Services, and others 2016). Based on the limited research on the topic, our group has identified that the parental relationship may be one of the most viable and effective areas for intervention. Thus our paper aims to address the question: Do adolescents in Minnesota who have positive parental and at-home environments have a lower risk of engaging with e-cigarettes and vaping products, as compared to adolescents who have negative parental and at-home environments? We first observe this relationship using areal spatial methods such as SAR and aggregated to a county-level using repeated measure data to inform our GEE models. With our chosen SAR model we observed that on a county level, variables such as gender, perceived risk and level of parental care are significant when predicting the proportion of the county that vapes. However, still on a county level, the effect of parental care depends on the proportion of respondents that are male in the county of interest. We then investigated this further, in an effort to see if these variables had a different impact at the individual scale compared to county scale. The individual scale, using GEE models, showed that there is a difference of impact of these variables, especially for males. Thus, we hope that our project highlights the different effects risk factors have between the county and individual level so the state can tackle through interventions and policy this ever growing public health crisis.

1. Introduction

In the past 5 years, e-cigarette use has risen sharply among teens and adults, despite significant negative health risks. According to data from the National Youth Tobacco Survey, the 2011 prevalence of e-cigarette use among high school students was 1.5%, a number that increased nearly tenfold by 2015, surpassing the rate of conventional cigarette-use among high school students (Health, Services, and others 2016). At a 16% prevalence rate, this equates to 2.4 million high school students and 620,000 middle school students have used an e-cigarette at least one time in the past 30 days (Health, Services, and others 2016). These trends have led to substantial concern and discussion within public health communities, who hope to determine and mitigate the risk factors for e-cigarette use.The most vulnerable populations are middle school and high school aged individuals due to risk factors special to their population like young age and strong peer & family pressures. Researchers have identified three key themes and risk factors among e-cigarette users: male gender, socioeconomic background, and weak social and parental support. (Hawkins, Catalano, and Miller 1992).

This paper looks to examine the third risk factor, which has had fewer devoted studies. We ask the question, do adolescents in Minnesota who have positive parental and at-home environments have a lower risk of engaging with e-cigarettes and vaping products, as compared to adolescents who have negative parental and at-home environments? In order to answer this question, we look to gain a better understanding of adolescent e-cigarette and vaping usage in Minnesota at an individual level and a county-level taking into account spatial correlation. Since there is a limited number of resources and literature that has studied these effects, we hope our research serves as an invitation for local public health departments to explore this epidemic before it reaches the magnitude of conventional cigarette smoking.

2. Methods

2.1 The Data Set

Our data is from the Minnesota Student Survey (MSS), a survey administered once every three years since 1989 (Education 2019). The survey acts as the primary source of data about the health and well-being of Minnesota's youth, and primarily focuses on students in grades five to twelve. The survey is voluntary on the part of the school districts, but all school districtrs are encouraged to participate including; public, charter, trival and nonpublic schools. Even tough the survey is voluntary for the school districts at least 81% of Minnesota school districts choose to participate in it (Education 2019). Moreover, the survey is a collaboration between local schools and four state agencies; the Minnesota Departments of Education, Health, Human Services and Public Safety. The state agencies develop the survey content, which consists of nearly 300 questions concerning the activities, opinions, behaviors, and experiences of each student.

It is important to note that the survey is voluntary on the part of the school districts. Parents or guardians may choose to opt their child out of the survey, or students opt themselves out too. Additionally, the students that do choose to participate are able to skip any question or stop at any point.

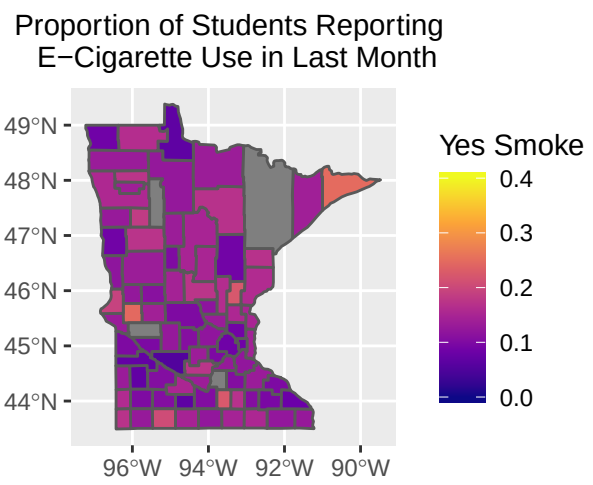
2.2 Variables of Interest

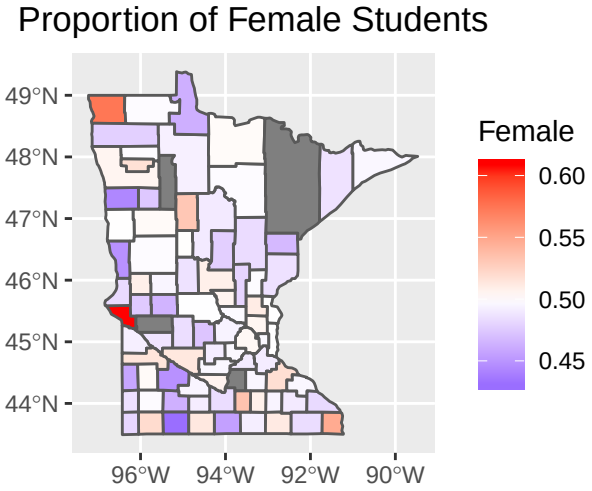
This paper looks specifically at the 2019 MSS, which contains around 170,000 entries and over 300 variables. However, we chose to omit non-answers for all of our variables of interest. Once these non-answers were omitted, there were 23,000 student surveys remaining, for which students answered every question that we were interested in.

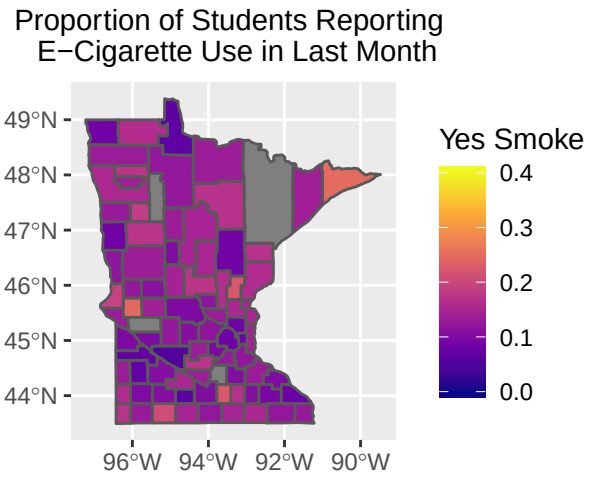
With that being said, since our question of interest focuses on the relationship between parents/guardians and adolescents in regards to e-cigarette usage, we chose to focus on seven variables: smoking status, sex, age, perceived harm of smoking, perceived parental care, parental perception of e-cigarette usage, and parental alcohol usage. Each of these variables was originally on a quantitative scale, but we transformed each into categorical levels as each variable had varying scales. Additionally, once our variables were on a categorical scale, we combined categories (i.e HighCare, SomeCare, LowCare) so that we would have more interpretable results. Table 1 within our appendix depicts our selected variables and their corresponding levels.

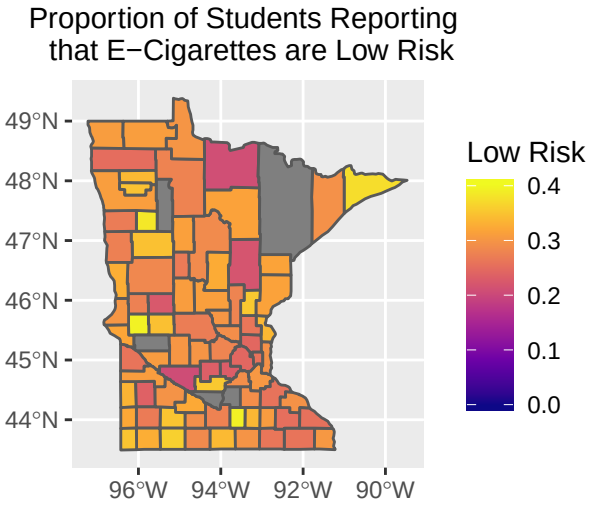
Additionally, each answer in the MSS is coded by the county the student was from, meaning we were able to create a county-level dataset with the proportion of student answers for each variable. For example, the variable "sex" shows the proportion of males and females in each county. These proportions were used to understand general spatial trends concerning our variables of interest. These proportions can be seen in the graphs below. It is important to note that the scale within each of the figures has been manually adjusted. Notably, all variables apart from sex had proportions below 40%, so they use a smaller scale. This is in an effort to preserve the visibility of the between county variation across Minnesota.

The first pair of graphs show an inverse relationship between e-cigarette use and gender. For example, Kittson County, in the top left, has the highest proportion of female students, and the lowest proportion of e-cigarette users. This indicates that sex is a risk factor for smoking.

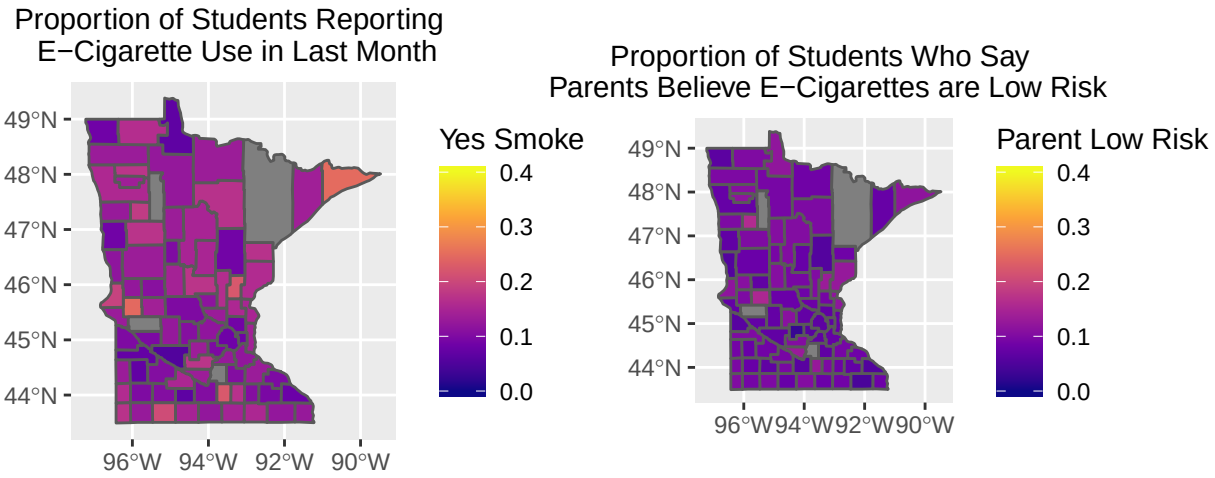








The second pair of graphs show proportion of students that smoke and the proportion of students who believe the effects of smoking are low. We chose these two because the level of perceived personal risk and danger an adolescent has towards vaping is a large indicator of their likelihood to engage/try smoking E-cigarettes (Soneji et al. 2017). When looking at the number of risk students associate with smoking as compared to county-level e-cigarette use, we see a strong positive relationship. That is, counties with more students who do not think people risk bodily harm if they use e-cigarettes have higher proportions of students using e-cigarettes. This is best represented in Cook County, in the top right corner of Minnesota, where more students believe vaping is low-risk, and more students smoke.



In the last set of graphs we examined parent perceived risk of smoking because researchers Parent et. al. identified a strong correlation between parental influence and risk of engaging with vapes and E-cigarettes (Pentz et al. 2015). From this visualization, we can see that while there is a relatively high proportion of students who believe e-cigarettes have a low risk of harm, very few students reported their parents feeling the same way. It is, therefore, harder to measure the relationship between parental risk perception of vaping and e-cigarette use per county. However, one can notice quite a few counties, particularly in the bottom right corner of Minnesota, that have both high e-cigarette use and a relatively high proportion of parents who say e-cigarettes are low-risk. This, therefore, establishes parent risk perception as a variable of interest.

When graphed, there was no clear correlation between smoking status and our three remaining variables of interest: perceived amount of parental care towards the surveyed student, whether the student lived with someone they determined to be drinking too much alcohol, and student age. Despite the lack of clear patterns on a county-level, we chose to explore these variables on a repeated measures level due to evidence from existing literature regarding these variables (Pentz et al. 2015), (Soneji et al. 2017).

If the lack of correlation persists when these variables are analyzed within a spatial regression model, we will explore each variable using this repeated measures methodology to see if there is a stronger relationship between each of these three variables and e-cigarette smoking status on an individual level.

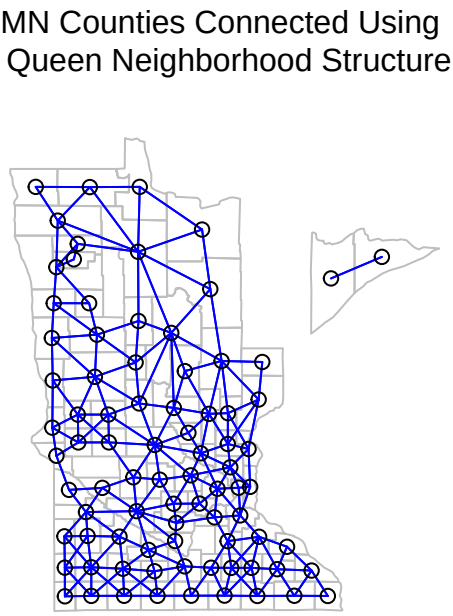
2.3 Statistical Methods and Model Selection Process

2.3.1 Spatial Methods

For our spatial data we chose to utilize a spatial autoregressive model (SAR model), instead of a point pattern/processes. We did this because the data from MSS does not provide the pinned point location of where the subjects in the survey live within the county. Also, for the purposes of our project, we want to explore the relationships between aggregate summaries within counties while specifying spatial dependence in terms of notions of neighborhoods and spatial proximity. Thus, our objective and dataset are suited for an areal data methodology. Furthermore, this data cannot assume that the occurrence or non occurrence of a point at a location is random because existing literature and evidence suggests that the networks and neighborhoods an individual lives in highly influences their risk of vaping. This is just a natural product of the networks and actions of human behavior. Therefore, we moved forward with a SAR model.

We used a SAR model to determine trends in e-cigarette use while accounting for spatial autocorrelation. The spatial dependence in the SAR model, which comes from matrix W, causes the simultaneous autoregressive of each random variable on its neighbors. For our project, we chose to utilize a Queen neighborhood-based correlation structure. After running a Moran.I test on both structures, we fail to reject the null hypothesis, there is zero spatial autocorrelation present, with the variable NoSmoke at alpha =.001. To further narrow down which structure to choose, we looked at the difference between the observed and expected values. We want the observed values to be significantly larger than the expected value because this indicates that the values of x are positively autocorrelated (Gittleman and Kot, n.d.). If observed was less than expected this

means there is a negative autocorrelation that would violate the criteria of the SAR. Thus, we choose the Queen structure because the observed values are 0.069 larger than the expected while the Rook difference was only 0.019 larger.



Then, moving forward with the Queen neighborhood structure, we fit our chosen model into a SAR model, using variables selected based on the exploratory visualizations discussed earlier which were backed by existing literature on parental and adolescent relationships and vaping.

2.3.2 Repeated Measures Methods

We also utilized repeated measures methods to better understand the effects of smoking on the individual level. These are compared to our spatial results, as a way to see whether individual and county-level trends differed.

We decided to use a marginal model over a mixed effect model because marginal models do not require distributional assumptions. This is because marginal model estimations only depend on specifying a few aspects of the observed distribution, and do not require an understanding of how individuals within the dataset are correlated (Hubbard et al. 2010). Furthermore, marginal models provide us with more informative outputs; each output includes robust errors, standard errors, and p-values, which enable us to compare two models with different correlation structures. Importantly, we are still able to obtain robust inferences even if the correlation structure is misspecified (Hubbard et al. 2010).

Thus, for our analysis and model selection process, we utilized the geeM R Package. The geeM R package is optimized for large datasets and produces easily interpretable outputs. Furthermore, due to the underlying ANOVA methods in the package, it gives us the ability to conduct model comparisons (Heggeseth 2019). Moreover, GEE models are similar to regression models except, GEE requires a specific vector that identifies the clusters, a link function, and the correlation structure (Heggeseth 2019). That being said, for our correlation structure, we went with an ar-1 as that structure is known to be the most complex and serves well in helping choose predictors. Once predictors are chosen, further investigation can be made in regard to the correlation structure. To determine the accuracy of different correlation structures, each model's Robust SE and model standard errors (Model SE) can be compared. Along with standard errors, geeM also provides Wald statistics and associated p-values for testing an initial hypothesis. These will be further highlighted within the results section.

3. Results

3.1 Spatial Results

The final output to our SAR model is:

	Estimates	Std.Error	z value	Pr(>|z|)
(Intercept)	-1.15	0.49	-2.31	0.021
Low Perceived Risk	0.36	.08	4.83	0.0
Low Parent Care	9.94	4.53	2.19	0.03
Male Biological Sex	2.18	1.00	2.18	0.03
Household Alcoholic	0.64	0.17	3.71	0.0
Low Parent Care * Male Sex	-19.38	8.96	-2.16	0.03

In equation form, the model provides the following information:

Prediction proportion of respondents in a county that smoke e-Cigs =

$$-1.1485 + 0.363(Risk.Low) + 9.942(Parent.Low.Care) + 2.18(Male)$$
$$+0.635(Yes.Household.Alcoholic) - 19.383(Male * Parent.Low.Care)$$

This model indicates that all variables of interest: sex, perceived harm of smoking, perceived parental care, parental perception of e-cigarette usage, and parental alcohol usage, had positive impacts on the proportion of respondents within a county who use e-cigarettes.

In specific, • The counties with a greater proportion of male residents had a greater proportion of residents who vape, after accounting for other factors.

• The counties with a greater proportion of respondents who indicated e-cigarette use was low risk had a greater proportion of respondents who vape, after accounting for other factors.

• The counties with a greater proportion of respondents who indicated their parents cared little about them had a greater proportion of respondents who vape, after accounting for other factors.

• The counties with a greater proportion of respondents who indicated they lived with someone who "drank too much" had a greater proportion of respondents who vape, after accounting for other factors.

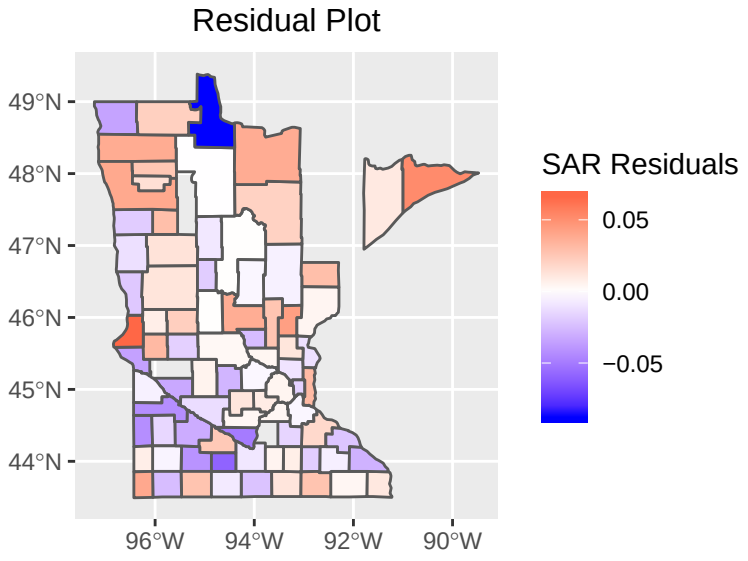
However, the effect of parental care depends on the proportion of respondents that were male in that county. For counties with a large amount of male students who indicate their parents care very little about them, we expect a lower amount of e-cigarette smokers. While this contradicts the literature (Health 2018), it does not necessarily indicate that males with low parent care are more likely to smoke, as this data is aggregated to the county-level. In addition, very few students overall indicated that their parents cared very little about them, and fewer once we look at only male students, meaning that the expected proportion of smokers is affected minimally by the interaction between sex male and low parent care.

This model provides relatively accurate results. For example, for Hennepin county, our predicted smoking proportion is as follows:

Predicted proportion of respondents within Hennepin county who smoke e-cigs =

$$-1.1485 + 0.363(0.2436515) + 9.942(0.0761456) + 2.18(0.4923746)$$
$$+0.635(0.05621898) - 19.383(0.4923746 * 0.0761456) = \mathbf{0.079}$$

As compared to the actual proportion of Hennepin county students that smoke e-cigarettes, 0.084, our model is only 0.5% away. When looking at the residuals for all counties, we see similar accuracy, and no discernible pattern, indicating a strong model.



3.2 Repeated Measures Results

We re-created the SAR model on an individual level, with the addition of age as a predictor given its value on the individual level as contrasted with average age per county. Given that the lowest age in the dataset was 10, we adjusted the age variable so that the coefficient on age describes the difference in smoking probability for every year past age 10, rather than past age 10. In addition, rather than predicting the proportion of each county that smoked, this model predicts the log odds of smoking for each individual, which we then transformed into the probability of smoking for each individual. As discussed above, we used a marginal model approach. We confirmed the model selection done through spatial methods using multiple methods, including hypothesis testing and sensitivity/specificity comparison. To determine whether each variable should be kept in the model, we calculated a Wald statistic as follows:

$$z = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$

If each variable should be kept in the model, we generally expect Wald statistics to be high; however, the Wald statistics we observed were relatively low. Despite this, we chose to keep the variables in the model due to their significance within the literature and the highly variable nature of survey data on young populations (Heggeseth 2019).

With these variables in mind, the final GEE model gave us the following results:

	Estimates	Model SE	Robust SE	Wald	p-value
(Intercept)	-1.39	0.06	0.11	-12.20	0.0
Low Perceived Risk	0.70	0.03	0.04	16.75	0.0
Low Parent Care	0.11	0.04	0.03	3.16	0.0
Male Biological Sex	-0.19	0.03	0.02	-8.34	0.0
Household Alcoholic	0.15	0.03	0.03	5.764	0.0
Age - 10	0.13	0.01	0.01	10.68	0.0
Parent Low Care * Male Sex	-0.12	0.06	0.04	-2.96	0.0

Note that the above estimates are in log odds form. When converted to probabilities and put into an equation predicting the probability of e-cigarette use, the model is:

$$\textbf{Probability of e-cigarette use} =$$

$$0.1999 + 0.669(Risk.Low) + 0.526(Parent.Low.Care) + 0.454(Male)$$

$$+0.538(Yes.Household.Alcoholic) + 0.533(Age - 10) + 0.469(Male * Parent.Low.Care)$$

These results confirm the conclusions from other studies, indicating that low parent care, age over 10, low perceived risk of vaping, male sex, living with someone who drinks too much alcohol, and the interaction between male sex and low parent care are all statistically significant risk factors for e-cigarette use, using the an α threshold of 0.001. Importantly, the interaction between male sex and low parent care has a positive effect on smoking at the individual level, in contrast to the results from county-aggregated data. Therefore, this discrepancy shows that it is an individual risk factor, and is being overlooked because it provides a different result on the county level.

To analyze this model, we calculated sensitivity and specificity, finding that with a balanced threshold of 0.5, our model achieves a sensitivity of 81% and specificity of 30%. This indicates a strong true positive rate and lower, but still relatively strong, true negative rate.

4. Conclusion

Our data does not allow us to dig deeper into the rural/urban or North/South divide when it comes to e-cigarette use. While we found that counties with more smokers tended to be more rural and more Southern, there was no clearly visible spatial pattern of e-cigarette usage in Minnesota. Despite this, we found that counties that neighbor each other demonstrate similar patterns of e-cigarette use. We also found valuable results when looking at the effects of our studied risk factors on e-cigarette use on the county-level.

Our SAR model found significant positive correlation between all studied risk factors, with the exception of the interaction between male sex and low parent care, which had a negative, but minimal, result on county-level smoking rates. Our GEE model found that vaping risk factors remained the same on the individual level, but that the interaction between male sex and low parent care led to a *positive* increase in the probability of smoking rather than a negative aggregate increase. Given this discrepancy, it is more valuable to look at the influence of this interaction at the individual level, as low parent care is a factor that influences individuals more than counties as a whole. This is an important finding for those involved in school-level interventions, and aligns with past research concerning risk factors for vaping and e-cigarette use.

It is clear that a lot of work remains to fully curtail e-cigarette usage and vaping among adolescents before it reaches a tobacco industry magnitude. However, progress is being made in research, development, policy, and practice. Our research of this MSS data is rather preliminary and observational, but it still illuminates interesting details about the risk factors the state could tackle through interventions and policy at the individual and county level.

5. Limitations

5.1 Spatial Limitations

Areal data in and of itself brings about a couple of limitations. Because areal data only provides summaries for each area studied, we cannot pinpoint where each specific instance falls within the county (Heggeseth 2019). If we had data that provided the exact longitudinal and latitudinal coordinates of each student, we would have been able to use point process data, allowing us to make more detailed and interpretable conclusions. However, our specific dataset only allowed us to utilize areal methods. In other words, because of the lack of detail in this dataset, we cannot study the specific spatial distribution of the adolescents who vape, giving us limited information about vaping patterns throughout Minnesota.

Also, the structure of areal data-limited us from being able to analyze certain predictive risk factors for e-cigarette usage. Specifically, we could not look at the effects of age on the county level as "average age" turned out to be the same for each county as each county had the same age range. Therefore, we could not draw valuable insights and thus had to implement a repeated measures method.

5.2 Repeated Measures Limitations

As a result of our data limitations, we had to implement a repeated measures method to capture the individual effects of smoking. This required us to create a new dataset that contained each instance per county rather than an aggregated county-level dataset. However, since students were able to skip questions, there was a fair share of NA's. As a result, our new dataset was limited to only 23,000 of the initial 170,000 observations. Also, being that all our subjects are children and young adults, the results are highly variable. Nonetheless, we expected that these subjects would have strongly varying perceptions and interpretations of the questions given their age.

6. Acknowledgements

We thank Amy Kinney, the Senior Research Scientist at the Minnesota Center for Health Statistics, for providing us access to this data. Also, we would like to thank Brianna Heggeseth for helping us throughout the entirety of the project. Special thanks also to Vittorio Addona and Leslie Myint for their advice on logistic GEE models.

7. Appendix

Variable Name	**Definition**	**Categorization**
Age	The age the adolescent is in at the time of the survey.	**Age range:** 10-19
Biological Sex	The binary biological sex of the adolescent.	**Male** or **Female**
Parent Care	How much the adolescent believes their parents care about them.	**1-3:** Low perceived care, **4-5:** High perceived care
Household Alcoholic	Adolescent identifies whether they live with someone who consumes too much alcohol	**Yes:** they live with someone who drinks too much, **No:** they do not live with someone who drinks too much
Smoking status	Adolescents identify how many days out of the past 30 days they vaped or smoked e-cigarettes.	**No smoke:** 0 days of e-cigarette use out of last 30 days, **Yes smoke:** 1+ days of e-cigarette use out of last 30 days
Perceived Risk	Adolescents identify how much they think people risk harming themselves physically or in other ways if they vape or use e-cigarettes	**1-2:** Low risk, **3-4:** High risk
Parent Risk	Adolescents identify how wrong their parents would feel it would be for them to vape or use e-cigarettes	**1-2:** Parent low risk, **3-4:** Parent high risk

8. References

Education, Minnesota Department of. 2019. "Minnesota Student Survey." *Minnesota Department of Education*. https://education.mn.gov/MDE/dse/health/mss/.

Gittleman, and Kot. n.d. "Moran.I." *Function | R Documentation*. https://www.rdocumentation.org/packages/ape/versions/5.3/topics/Moran.I.

Hawkins, J. David, Richard F. Catalano, and Janet Y. Miller. 1992. "Risk and Protective Factors for Alcohol and Other Drug Problems in Adolescence and Early Adulthood: Implications for Substance Abuse Prevention." *Psychological Bulletin*.

Health, Minnesota Department of. 2018. "Which Minnesota Teens Try Vaping?" *Tabacco and Nictoine Use in Minnesota: Bries Reports and Statistics*. MDH. https://www.health.state.mn.us/communities/tobacco/data/docs/0102whichteensvape.pdf.

Health, US Department of, Human Services, and others. 2016. "E-Cigarette Use Among Youth and Young Adults. A Report of the Surgeon General." *Atlanta, GA*.

Heggeseth, Brianna. 2019. "Correlated Data Notes." https://bcheggeseth.github.io/CorrelatedData/.

Hubbard, Alan E, Jennifer Ahern, Nancy L Fleischer, Mark Van der Laan, Sheri A Satariano, Nicholas Jewell, Tim Bruckner, and William A Satariano. 2010. "To Gee or Not to Gee: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health." *Epidemiology*. JSTOR, 467–74.

Pentz, Mary Ann, HeeSung Shin, Nathaniel Riggs, Jennifer B Unger, Katherine L Collison, and Chih-Ping Chou. 2015. "Parent, Peer, and Executive Function Relationships to Early Adolescent E-Cigarette Use: A Substance Use Pathway?" *Addictive Behaviors* 42. Elsevier: 73–78.

Soneji, Samir, Jessica L Barrington-Trimis, Thomas A Wills, Adam M Leventhal, Jennifer B Unger, Laura A Gibson, JaeWon Yang, et al. 2017. "Association Between Initial Use of E-Cigarettes and Subsequent Cigarette Smoking Among Adolescents and Young Adults: A Systematic Review and Meta-Analysis." *JAMA Pediatrics* 171 (8). American Medical Association: 788–97.