United States Abortion Clinics

The Impacts of Interacting Partisanship and Need

December 15, 2019

Abstract

We researched the relationships and trends found between the 2017 number of abortion clinics per 100,000 people in each U.S. state and various political, demographical, and general state characteristics. Due to the correlation between neighboring states in our data, it was necessary to utilize a Simultaneous Autoregressive (SAR) model rather than a more standard Ordinary Least Squares (OLS) model. Given the controversy surrounding the topic of abortion, a strong focus was placed on both politics and need throughout our research. Based on our work, we found that an interaction between state partisan composition and the logged teen pregnancy rate has a significant relationship with the number of clinics per 100,000 people observed in each state. Whether a state lost abortion clinics during recent years and the area per capita in a state were also influencing factors. Our final model allowed us to specify important patterns found in the accessibility of abortion clinics across the fifty states and reinforces the importance of state-level government. We hope that our work will help to encourage readers to be more aware of state government and to be proactive in using their votes to participate in important policymaking.

Contents

1 Introduction 1

2 Methods 2
   2.1 The Data Set ................................................................. 2
   2.2 Variables of Interest ...................................................... 3
   2.3 Statistical Methods and Model Selection Process ................. 5

3 Results 6

4 Conclusion 9

5 Acknowledgements 10

References 11

1 Introduction

Abortion is a controversial topic in American politics, with ideologies oftentimes at odds when it comes to policymaking. We wanted to investigate the state-level factors that influence an individual’s access to abortion clinics. Increases in legal and financial constraints have affected women’s access to abortions, despite the fact that women have cited the same reasons to have – or not have – an abortion (Finer et al. 2005). Since 1981, the abortion rate has fallen before reaching a historic low of 13.5% in 2017 (“Induced Abortion
in the United States” 2019). Regardless of whether women’s mindset towards abortion has changed, this trend has likely been exacerbated in the U.S. with the federal government’s passing of “TRAP” laws, which are “targeted restrictions on abortion providers designed to close them down rather than make them safer for women” (“What Are Trap Laws?” 2019), the solidification of an anti-abortion majority in the Supreme Court in 2018, and the Trump-Pence administration’s Title X gag rule. Notably, this gag rule has greatly impacted funding for Planned Parenthood, as it “prevents Title X recipients from referring patients for abortion care” (“Trump Administration Finalizes Title X Domestic Gag Rule” 2019). Overall, many factors have been contributing to the decrease in women’s access to abortions and abortion clinics.

Furthermore, we noted early on how greatly access to abortion clinics varies from state to state. Compared to other countries, the United States has a unique federal government structure, dividing the responsibility between the federal government and state sub-governments. In fact, fewer than 30 modern countries have federal systems (“Federalism” 2019). This means that states have a unique amount of power over U.S. policymaking, which can result in wild variation between the different states. Given this unique structure, we wished to examine the effects of various variables on the abortion rate in each of the individual states, with a particular focus on politics and partisan composition. In addition, we were curious as to whether groups of states in similar geographic regions tend to have similar proportions of abortion clinics per individual after accounting for the trends that we found. Research into state-level policymaking is critical, especially in a time where many state governments are working to decrease or even remove access to abortions. There are now partial bans on abortions in some states, and many individual clinics are feeling the pressure of reduced funding (“Timeline of Attacks on Abortion” 2019).

Throughout this paper, we introduce the data we used and variables of interest, the modeling methods, and the evaluation criteria used to compare between models. We then introduce our final model, acknowledge its limitations, and finally utilize this model to synthesize important conclusions regarding the accessibility of abortion clinics by state.

2 Methods

2.1 The Data Set

In order to study state-by-state differences in abortion clinic accessibility, we primarily used data retrieved from the Guttmacher Institute’s U.S. States Abortion Data Center (“U.S. States Abortion Data Center” 2019). This database contains data on multiple abortion-related measures for each of the fifty states. These measures included the number of abortion clinics in 2017, the change in the number of clinics between 2014 and 2017, abortion rates and numbers, demographics data focusing on women of reproductive age, data on contraceptive services and supplies, federal and state abortion funding, as well as various other data. Specifically, it defines abortion clinics as “physicians’ offices reporting 400 or more abortions a year.”

It is worth noting that the measurements included in this database come from various years, ranging from 2013 to 2018. This is likely due to the large effort involved in obtaining all of this detailed data. It is also worth noting that certain columns of the dataset are particularly prone to missing values, which unfortunately made them less viable for our research. Data on funding and contraceptive supplies and care, for example, contained many missing values.

In addition to this data, we also merged in additional data on partisanship, area in square miles, and population by state. For our partisanship data, we retrieved data from the National Conference of State Legislatures on the partisan composition of each of the fifty states in 2015 (“State Partisan Composition” 2019). We chose this year because it is sufficiently recent to impact our data on the number of clinics in each state, which was measured in 2017, but also not too recent, as setting up new clinics and closing existing ones likely does not take place immediately when the partisan composition of a state changes. 2015 also falls in between 2014 and 2017, the period over which our dataset measures the change in the number of clinics by state. For our area data, we retrieved 2018 data on the total square miles of area by state from the U.S. Census Bureau (“State Area Measurements and Internal Point Coordinates” 2018). For population, we
retrieved 2015 data on state populations from the U.S. Census Bureau, the Population Division (“Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2015” 2016). This year was chosen because although our numbers of abortion clinics by state are from 2017, our primary dataset, from Guttmacher, included data from as far back as 2013, so 2015 was a reasonable middle ground. It is also likely that state populations had not changed too greatly between 2015 and 2017.

2.2 Variables of Interest

![Number of Abortion Clinics Per 100,000 People](image.png)

**Figure 1: U.S. plot of clinics per 100,000 people, by state.**

For our research, we chose to primarily focus on modeling the number of abortion clinics per 100,000 people by state, as this gave a relatively reasonable rating of the “intensity” of clinic-accessibility for each state. This key measure is plotted above. Note that this value is used rather than simply abortion clinics per capita, which is a very small value, because it gives us more interpretable results.

We considered many different measures of clinic-accessibility before settling on clinics per 100,000 people, and we felt this measure had the fewest drawbacks. For example, focusing on the total number of clinics would skew the numbers towards larger states. Studying measures such as clinics per square mile would also be less preferred, as these would be unfair to states that are more sparsely populated.

Nevertheless, we acknowledge that focusing on clinics per 100,000 still has its drawbacks, as this does not take the size of the abortion clinics into account. For example, imagine a smaller city only just large enough to have its own abortion clinic. Then consider a very densely populated city. Although the densely populated city is likely to have many more abortion clinics than the small city, it is also likely to have larger, more staffed clinics. This may actually result in the very dense city having a lower number of clinics per capita, and therefore also a lower number of clinics per 100,000 people, than the small, more sparsely populated city. There is no perfect measure of clinic-accessibility, and we felt that the measure we settled upon was appropriate for the scope of our project.

As for our explanatory variables, two main categories of variables came to stand out during our initial exploratory research as primary candidates for our model: state partisanship and need. (Note that we
generally considered data describing either pregnancy rates or proportions of the state population that would benefit from abortion clinics as measures of need). We quickly noticed that an interaction of these two variables appeared to influence the clinics per 100,000 people observed in each state. For state partisanship, we had several measures we considered: percentage of democratic seats in state legislature, governor’s party, legislature’s controlling party, and overall state control. Of these variables, we had the most success when including overall state control in our model. A plot of the clinics per 100,000 people separated by state control is shown below. Note that Alaska and Hawaii are excluded from these plots to allow for the other states to be more visible.

![Number of Abortion Clinics Per 100,000 People, by State Control](image)

Figure 2: U.S. plot of clinics per 100,000 people by state, grouped by state control.

For need, we considered the following variables: the number of pregnancies per 1000 women aged 15 to 19 by state of residence in 2013, the percentage of women aged 15 to 44 by state of residence in 2017, and the percentage of women aged 13 to 44 with potential demand for contraceptive services and supplies by state of residence in 2016. Of these variables, the number of pregnancies per 1000 women aged 15 to 19 by state of residence in 2013 (from now on referred to as “teen pregnancy rate”) stood out, and we obtained the best results when we included an interaction term between state control and logged teen pregnancy rate in our model. A plot displaying this relationship can be seen below. We noticed that states with the lowest logged teen pregnancy rates often had more clinics per 100,000 people.
It is worth noting that a limitation of this variable selection is that it fails to capture the change of state partisanship over time, which without a doubt also effects the number of abortion clinics in a state. For example, Rhode Island is the lowest-scoring point in the plot above for the democrat-led states, but it has also been under divided control in the past, such as in 2010, when Rhode Island had a Republican governor. A similar phenomenon also occurs for various divided states, many of which switch between being divided and democrat or between being divided and republican. In extension, some of the republican-led states are also not consistently under republican control. Nevertheless, our data on state control proved to be the most useful of all of our measures of state partisanship for predicting clinics per 100,000 people.

Beyond these two key variables, we also found two other interesting and useful variables: whether a state lost clinics between 2014 and 2017 and the total area of a state in square miles divided by the population, or area per capita. We were curious whether states with fewer clinics were the same states that were losing clinics, and we discovered that this relationship did appear to occur in our data. We also considered that area per capita may be likely to increase the number of clinics per 100,000 people, because although more sparsely populated states may have less need for clinics in the sense that they may have fewer people, they may still also have higher need for clinics due to the longer travelling distances required to actually reach a clinic in such a state. This variable of interest also proved to be useful in our model.

2.3 Statistical Methods and Model Selection Process

2.3.1 Neighborhood structure

When attempting to account for the correlation in areal data, it is necessary to select a neighborhood structure. Several common neighborhood structures are available, such as Rook, Bishop, Queen, KNN, and centroid distance. We selected a Queen structure for our work. This structure considers all states with borders touching a given state to be that state’s neighbors. For the context of our research, we felt that this neighborhood structure made logical sense, as states directly touching each other could easily influence each
other’s politics, and residents from one state may sometimes travel to a clinic in a neighboring state, if that clinic is the one that is the most accessible to them.

There is a weakness with choosing this neighborhood structure, as it leaves Alaska and Hawaii without any neighbors, and there is not consensus on how to treat areas that do not have neighbors. However, we feel that this neighborhood structure is the most accurate, even for these two states, as they are too distanced from the rest of the states to truly be influenced by them in the same way a more typical state would be influenced by its neighbor. Utilizing a k-nearest neighbors structure or a centroid distance structure with a very large distance, such that Hawaii and Alaska both have at least one neighbor, would not result in a realistic representation of potential correlation between states.

2.3.2 Correlation Structure & Method Choice

Before modeling, we used Moran’s I to confirm that the data of interest contains spatial correlation, determining whether a Simultaneous Autoregressive (SAR) model or Conditional Autoregressive (CAR) model would be necessary, as well as which of the two would be more suitable. Note that if there is not sufficient evidence of spatial correlation, a simpler Ordinary Least Squares (OLS) model is preferred. According to the Encyclopedia of GIS, SAR models are appropriate when there exists large-scale, global spatial autocorrelation, whereas CAR models are more appropriate for relatively local spatial autocorrelation (Shekhar and Xiong 2007). Furthermore, CAR models are also typically used in a Bayesian context (Heggeseth 2019). Moran’s I is a test statistic for which $H_0$ is that the outcomes $Y_i$ are independent and identically distributed, and thus are not correlated. The clinics per 100,000 people variable yielded a Moran’s I p-value of 0.002702. As this is below the standard 0.05, we have sufficient evidence to reject the null hypothesis; therefore, the data is spatially correlated.

Given the presence of spatial correlation, an OLS model will not be sufficient. We therefore must use either a SAR model or a CAR model. We chose to fit a SAR model in part because we were not focusing on fitting the model in a Bayesian context. Professor Brianna Heggeseth’s advice was also valuable in making this decision. This SAR model was fit using R’s spdep package, more specifically, the spautolm function. SAR models are fit as $Y = \lambda W Y + X \beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$, where $W$ is the proximity, or “neighbor,” matrix calculated for the areal data using the desired neighborhood structure, Queens in our case. $Y$ represents the outcomes of interest. Note that the $\lambda W Y$ term can therefore be thought of as accounting for the spatial correlation in the data. Typically, this model assumes that $Y$ are Gaussian. Using maximum likelihood estimation, $Y$ is approximated as a sample from the following normal distribution:

$$ Y \sim N(X \beta, \sigma^2(I - \lambda W)^{-1}(I - \lambda W^T)^{-1}) $$

2.3.3 Model Selection and Evaluation Criteria

We began our exploration by picking and choosing different explanatory variables, gradually refining our selections based on the results. Throughout this process, we used Moran’s I to confirm that there was no evidence to reject the null hypothesis that the remaining residuals were independent and identically distributed – therefore indicating to us that a given model successfully accounted for the correlation structure of the data. Additionally, we considered the p-values of the coefficients themselves, confirming that we had sufficient evidence to indicate that they were statistically significant. We also generated both the Q-Q plot and residuals plot for each model tried in order to confirm and compare their quality. Finally, we used the Akaike information criterion to assist in identifying the best candidate models, balancing both accuracy and the number of explanatory variables to avoid overfitting.

3 Results

After carrying out our research, we have come upon the following as our final model:

$$ \text{Clinics per 100,000} \sim \text{State Control} \times \log(\text{Teen Pregnancy rate}) + \text{Area per capita} + \text{Lost clinics?} $$
Clinics per 100,000 is the number of abortion clinics per 100,000 people in a state in 2017. State control is considered to be the dominant party in a given state in 2015. However, if the party of the governor and the majority party of the state legislature are different, the state control is classified as “divided.” It is worth noting that Nebraska is unique and lacks a classification here because its legislature consists of a nonpartisan senate, while its 2015 governor was Republican. Teen Pregnancy rate is the approximate number of women out of 1000 between ages 15 and 19 who were pregnant in 2013. Area per capita is the number of square miles per individual in a given state, while Lost clinics? is a boolean variable for whether the number of abortion clinics in the state decreased between 2014 and 2017. The model can be seen below:

|                          | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | 2.922    | 1.097      | 2.66    | 0.01     |
| State.ControlDivided     | -1.269   | 1.265      | -1.00   | 0.32     |
| State.ControlN/A         | -0.427   | 0.171      | -2.50   | 0.01     |
| State.ControlRep         | -2.776   | 1.217      | -2.28   | 0.02     |
| log(preg.rate.15.to.19)  | -0.673   | 0.305      | -2.21   | 0.03     |
| area.per.capita          | 0.346    | 0.191      | 1.81    | 0.07     |
| lost.clinics?            | -0.010   | 0.052      | -1.91   | 0.06     |
| State.ControlDivided:log(preg.rate.15.to.19) | 0.315 | 0.351 | 0.90 | 0.37 |
| State.ControlNA:log(preg.rate.15.to.19) | NA | NA | NA | NA |
| State.ControlRep:log(preg.rate.15.to.19) | 0.677 | 0.336 | 2.02 | 0.04 |

Table 1: Model Coefficient Summary: Clinics Per 100,000 People

By interpreting this model and its coefficients, we can expose several interesting trends in the data. Keeping all other variables fixed, a state with democrat control has 2.922 abortion clinics per 100,000 people. In contrast, a republican-controlled state has 0.146 abortion clinics per 100,000 people, which is 2.776 fewer than democrat-controlled states, when all other variables are held fixed. In between these two extremes, a divided state may have 1.653 clinics per 100,000 people, which is 1.269 fewer than a democrat-controlled state, but this coefficient is not statistically significant. With other variables held fixed, Nebraska has 2.495 clinics per 100,000 people, which is 0.427 fewer than a democrat-controlled state. Holding other variables fixed, a democrat-controlled state loses 0.673 clinics per 100,000 people for every point higher that their logged teen pregnancy rate increases. (Note that this rate is measured as the number of pregnancies per 1000 women aged 15 to 19 by state of residence in 2013). For a republican-controlled state, when other variables are held fixed, every point increase in the logged teen pregnancy rate results in 0.004 additional clinics per 100,000 people (a very slight incline, instead of a decline). For a divided state, this slope may be -0.358 (in between the rates for democrat and republican-controlled states), but this coefficient is not statistically significant. For Nebraska, this slope is not altered, as you cannot estimate a slope for only a single case. Holding all other variables constant, a state that lost clinics in between 2014 and 2017 will have 0.010 fewer clinics per 100,000 people than a state that did not lose any clinics in that time period, but this coefficient is barely not statistically significant with a p-value of 0.06. Also, holding all other variables constant, a one-unit increase in area per capita (measured in square miles per capita) results in a 0.346 increase in clinics per 100,000 people, although this coefficient is also barely not statistically significant with a p-value of 0.07.

Next, we consider the quality of these results. Note that the row for the slope adjustment to the logged teen pregnancy rate where the state control is NA has all “NA” values as this row only applies for a single state, Nebraska, and a coefficient for a slope adjustment cannot be estimated for a single case. As can be seen above, all of the interaction coefficients but this one, the intercept adjustment for divided states, and the logged teen pregnancy slope adjustment for divided states have statistically significant p-values when using a significance threshold of $\alpha = 0.05$. Given that “Divided” is only a single category of the state control variable, it is acceptable for it and its respective logged teen pregnancy rate slope adjustment coefficient to not be statistically significant, since all the other interaction coefficients do meet this criteria. Unfortunately, for area per capita and for whether a state lost any clinics between 2014 and 2017, both of these coefficients have p-values that are barely too high to be significant, with values of 0.07 and 0.06 respectively. Despite this, these p-values are so close to the significance threshold of $\alpha = 0.05$, and including these two variables
increased the quality of both our Q-Q plot and AIC value. The Q-Q plot is shown below, and it almost perfectly follows the line except for the right-most point. This model’s AIC score of -15.2 was also the lowest AIC of all our models. We therefore felt this slight breach in the significance of our coefficients was acceptable, since $\alpha = 0.05$ is simply an arbitrary threshold.

![Normal Q-Q Plot](image)

Figure 4: Model Q-Q plot, showing a slight upturn indicating the presence of potentially more outliers than we might expect if it were Normal.

The next evaluation criteria we considered were the residuals of our model. These can be seen in the plot below. These residuals performed, overall, either better than or comparably to the other models that we tried. It is worth noting that Maine appears to have the worst residual by far, but Maine was an outlier state that was under divided control yet still had the most clinics per 100,000 people of all the states. Given the lack of a pattern in states with divided control, we considered this singular large residual to be reasonable. Furthermore, the residuals appear to be quite random and do not follow any obvious spatial patterns.
It appears this model performs well thus far, but how well does our model account for the correlation between neighboring states? After fitting this model, we find that the residuals do not show any evidence of being correlated, with a Moran’s I p-value of 0.51, which is much higher than \( \alpha = 0.05 \). Furthermore, we actually observe a lambda of -0.472 after fitting this model, suggesting that neighbor states may even be negatively correlated with each other after taking the relationships seen in our model into account. This could potentially suggest that a higher number of clinics in one state may compensate for lower numbers of clinics in neighboring states, as residents of neighboring states may rely on the clinics in this state. This may especially make sense for smaller states, such as the states along the east coast. However, the Numerical Hessian Standard error suggests that this value is not statistically significant, but it is still an interesting feature to note.

Overall, we were quite pleased with the performance of our final model in comparison to our other models, 

\[
\text{Clinics per 100,000} \sim \text{State Control} \ast \log(\text{Teen Pregnancy rate}) + \text{Area per capita} + \text{Lost clinics?},
\]

and from it we learn some very interesting relationships.

## 4 Conclusion

Through our research we have come upon a model describing the patterns influencing variation in abortion clinic accessibility in the fifty states of the U.S., but our model does have limitations.

Moran’s I, though useful, is not a perfect test statistic. The results depend heavily on the structure of \( W \), and therefore our given correlation structure. Additionally, both non-constant variance and spatial trends may result in observations not appearing independent and identically distributed. Finally – and particularly relevant in this case for isolated states like Hawaii and Alaska – there is not consensus on how to treat areas that do not have neighbors.

In addition to potential limitations with the statistical methods, states’ governments on their own are already complicated and also dynamic structures. We attempted during our analysis to use state partisanship as a
factor, but we only included a single snapshot of state partisan composition from 2015. This means that states that frequently shift between divided and either republican or democrat state control may not be described as accurately by our data. In potential future work, we would like to attempt to address this issue and find a better measure that takes this into a count. Another weakness of our approach is that it struggles to categorize Nebraska, which is unique in its officially nonpartisan state senate.

Lastly, the per 100,000 people standardization still does not entirely capture the “true” level of access to an abortion procedure in a given state. This is because any given clinic might be significantly larger or better staffed and equipped than another. This may be particularly relevant when comparing clinics per 100,000 people between a more densely populated area, which may prefer to have a few larger clinics, and a more sparsely populated area, which may have just one small clinic but also such a small population that it may have a higher number of clinics per 100,000 people.

Despite these limitations, given the data and methods available, we made choices we felt best balanced the tradeoffs involved. From our model we can learn a lot about the accessibility of abortion clinics in different states. A key takeaway is that with all other factors held constant, democrat-controlled states tend to have significantly more clinics per 100,000 people overall. Republican-controlled states, in comparison, have much fewer. States with divided control demonstrate no significant pattern, which is logically reinforced by our understanding that divided states likely switch back and forth between divided and either republican or democrat control.

Another key pattern we observed is that democrat-controlled states tend to lose clinics as logged teen pregnancy rates go up. We hypothesize that several factors may underlie this trend. It is possible that states with higher teen pregnancy rates may have more stigma surrounding abortion. It is also likely that states with lower quality sex education and less access to contraceptive care and supplies, which may correlate with states that have fewer clinics per 100,000 people, potentially have higher teen pregnancy rates. It is worth noting that this pattern is different for states with republican control; they actually see a very slight increase in clinics per 100,000 people as the logged teen pregnancy rate rises. We believe this can be attributed to the fact, however, that you cannot have a negative number of clinics, and republican-controlled states already have much fewer clinics per 100,000 people in general.

Finally, we also observed some relationships between the number of abortion clinics per 100,000 people and area per capita and whether a state lost any abortion clinics between 2014 and 2017. Our model found coefficients that barely failed to meet the standard significance threshold, but we still found these variables to be interesting and helpful. It appears that as area per capita increases, the number of clinics also increases. This likely is because although some states may have smaller populations, if the population is sufficiently spread out, there is still need for additional clinics in order to decrease the travelling distance to reach a clinic. We also found that states losing clinics appear to be more likely to have smaller numbers of clinics per 100,000 people, suggesting that states with more clinics may also be more likely to be maintaining their number of clinics.

As can be seen from our work, state-level government has a large impact on our lives, influencing in particular, our access to important reproductive care. However, statistics reported by FairVote.org indicate that voter turnout for state elections is highly variable across different states, ranging in the 2018 elections from as low as 39.3% in Hawaii to as high as 64.2% in Minnesota (“Voter Turnout” 2018). In 2014, these numbers are even worse, ranging from 28.3% in Texas to 58.5% in Maine (“Voter Turnout” 2018). In comparison, 61.4% of eligible voters participated in the 2016 presidential election (“Voting in America: A Look at the 2016 Presidential Election” 2017). Overall, we hope that our findings can both spur further research into these issues and also bring into the sharp realization the importance of voter participation.
References


