THE UNIVERSITY OF CHICAGO

Local Immigration Enforcement and Crime in the United States: A Nationwide Analysis of the 287(g) Program

By
Amanda Peters

August 2021

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts Program in the Social Sciences

Faculty Advisor: Guanglei Hong
Preceptor: Amit Anshumali

Abstract

The 1996 Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA) created the 287(g) program, which authorizes Immigration and Customs Enforcement (ICE) to delegate immigration enforcement to local law enforcement agencies. Under the program, city police departments, county sheriffs' offices, and state law enforcement bodies voluntarily apply to the 287(g) program. If selected, ICE provides training in immigration enforcement practices to employees of the law enforcement agency, who are authorized to investigate individuals' immigration statuses and detain undocumented individuals prior to immigration proceedings. The earliest 287(g) agreements between ICE and local law enforcement agencies were signed in 2002, and the program continues to this day. Proponents of 287(g) agreements argue that participating law enforcement agencies identify and detain dangerous "criminal aliens," improving public safety and quality of life. Opponents claim that the 287(g) program undermines trust between immigrant communities and law enforcement, is costly for local law enforcement agencies, and contributes to the phenomenon of "crimmigration," in which criminal and immigration law increasingly overlap.

In order to evaluate the effectiveness of 287(g) programs, I examine whether these programs reduce violent crime. In particular, I use propensity-score matching to pair counties which implemented 287(g) agreements with comparable counties that did not. I then use a differences-in-differences design to compare changes in county crime rates in the treatment and control groups. I find that after adjusting for potential confounding variables, 287(g) agreements have no effect on the trend in violent crime rates. I then repeat my analysis using only murder and nonnegligent manslaughter to assess whether 287(g) agreements cause underreporting of other violent crimes. I observe an increase in murders in 287(g) jurisdictions relative to other

counties, suggesting that underreporting of other violent crimes may have occurred. This analysis suggests that 287(g) agreements may undercut public safety by damaging trust between immigrant communities and law enforcement. This study suggests that the evidence does not support the adoption of 287(g) agreements as a public safety measure and that more lenient policing practices towards immigrants may have a neutral or positive effect on public safety.

Introduction

In 1994, California voters backed Proposition 187, a ballot initiative intended to prevent undocumented immigrants from accessing public assistance. While Proposition 187 was deemed unconstitutional, it revealed Californians' deep dissatisfaction with the presence of undocumented individuals in their state and with lax federal immigration enforcement. In response, Congress adopted the Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA) in 1996, which tightened immigration enforcement and border security in a number of ways. Notably, IIRIRA added Section 287(g) to the Immigration and Nationality Act, allowing Immigration and Customs Enforcement (ICE) to delegate immigration enforcement responsibilities to local law enforcement agencies (Provine et. al. 2016). The 287(g) program was intended to allow local law enforcement agencies to voluntarily act as "force multipliers" for ICE. Under the program, ICE signs a memorandum of understanding (MOU) with a local law enforcement agency and provides training in immigration enforcement to select employees of the local agency. While ICE covers the costs of training, local law enforcement agencies are responsible for travel expenses, salaries, and overtime hours associated with immigration enforcement training and practice (American Immigration Council 2021).

Initially, Section 287(g) allowed for two models of agreements between local law enforcement and ICE. Under the first, known as the Jail Enforcement Model (JEM), law enforcement officers interrogate suspected noncitizens who have been arrested for state and local offenses about their immigration status. If officers believe arrestees are subject to deportation, they may detain them for immigration purposes even if there is insufficient evidence to continue detention for criminal offenses (American Immigration Council 2021). The second and most controversial model of 287(g) agreement was the task force model. Under the task force model, deputized officers could identify, question, and arrest individuals they suspected of violating federal immigration law in the course of their daily activities (American Immigration Council 2021). Notably, this model allows officers to stop and arrest individuals who are not suspected of any offense other than unauthorized presence in the United States. ICE later introduced two additional models of 287(g) agreements. The hybrid model of 287(g) agreement combined elements of multiple models. In 2012, the task force and hybrid models were discontinued due to costs, oversight challenges, and concerns about racial profiling (Provine et. al. 2016). Finally, the warrant service officer (WSO) model was introduced in May 2019. The warrant service officer model allows local law enforcement officers to arrest individuals in local jails with active immigration warrants. While the warrant service officer model is similar to the jail enforcement model, the former does not allow law enforcement to interrogate alleged noncitizens about their immigration status (American Immigration Council 2021).

In 2008, the Obama administration introduced the Secure Communities program, which gradually rolled out a mandatory role for local law enforcement agencies alongside the voluntary 287(g) program. Under Secure Communities, local law enforcement agencies were required to compare arrestees' fingerprints to a national database of immigration warrants. If the fingerprints matched, agencies were required to detain the arrestee for up to 72 hours to facilitate immigration proceedings (Provine et. al. 2016). In 2014, the federal government replaced Secure

Communities with a more limited program, PEPS, which was ultimately abandoned in 2017 (ICE 2017). The Trump administration saw a revival of the 287(g) program as local law enforcement agencies signaled support for President Trump's aggressive immigration enforcement policies. Compared to the first wave of 287(g) agreements, agencies that adopted 287(g) agreements in the second wave were predominantly rural and Southern; in particular, 26 of the 70 currently active JEM agreements as of 2021 are located in Texas, and 46 of the 76 active WSO agreements are in Florida (ICE 2021).

ICE has predominantly promoted the 287(g) program as a public safety measure. By asserting that so-called "criminal aliens" are a serious threat to public safety, proponents of the 287(g) program reframe immigration enforcement as a local law enforcement issue. Within this logic, devoting local law enforcement efforts to immigration enforcement will ultimately result in lower crime rates. In order to illustrate this objective, ICE publishes a monthly report describing particularly dangerous individuals apprehended under 287(g) programs (ICE 2021). These reports are primarily illustrative—instead of a random sample or comprehensive summary of all arrestees under 287(g), each encounter report focuses on particularly serious crimes. Relative to the total population of undocumented immigrants, arrestees featured in ICE's 287(g) encounter reports are disproportionately Latin Americans who crossed the border illegally (as opposed to individuals who overstayed their visas).

Since undocumented immigrants have weaker constitutional protections than lawful permanent residents and citizens, the 287(g) program provides local law enforcement with more tools to address offenses by undocumented immigrants. A 1975 Supreme Court decision allows immigration enforcement officers to consider race in deciding whether to stop individuals on suspicion of unauthorized presence in the United States, although race cannot be the only factor

(United States v. Brignoni Ponce). On the other hand, explicit racial profiling is clearly prohibited in criminal law enforcement. Further, the 287(g) program allows law enforcement officers to continue to detain immigrants for longer than criminal procedure dictates. For example, if an undocumented immigrant were arrested on suspicion of committing a robbery, but prosecutors decided not to file charges, the 287(g) program could prevent their release. In some 287(g) jurisdictions, traffic violations that ordinarily would not lead to arrest provide the basis for detention and removal of undocumented immigrants (American Immigration Council 2021).

On a broader theoretical level, the 287(g) program is one element of the larger trend of decentralization in U.S. immigration policy. Proponents of this general trend argue that increasing local involvement in immigration policy allows cities and counties to more closely match their residents' preferences. In some cases, this devolution of immigration enforcement authority might lead to policies more inclusive of immigrants (Provine et. al. 2016). This argument is particularly relevant in the case of county sheriffs, who are directly elected and often promise to sign or end 287(g) agreements on the campaign trail. In these cases, the 287(g) program provides a means for citizens to directly affect immigration enforcement policy at a local level.

On the other hand, opponents to the 287(g) program often identify it with the larger trend of "crimmigration," a term coined by Juliet Stumpf in 2006 to describe the increasing overlap between criminal and immigration law in the United States. Stumpf contends that both immigration and criminal law are primarily forms of inclusion and exclusion. In criminal law, the state excludes offenders through incarceration and through the collateral consequences of conviction, which mark offenders as outsiders even when they return to mainstream society. Similarly, immigration law separates immigrants into lawfully admitted and unauthorized, legally and illegally present (Stumpf 2006). Stumpf observes that the number and nature of crimes that can lead to deportation of lawful permanent residents have expanded dramatically since the 1980s. She also notes growing similarities in legal procedure and enforcement of immigration and criminal law (Stumpf 2006). As Felicia Arriaga argues, the 287(g) program is fundamentally a form of crimmigration (Arriaga 2017). By joining the 287(g) program, local law enforcement agencies frame immigration as a criminal rather than civil enforcement issue. For undocumented immigrants accused of crimes, criminal and immigration proceedings overlap and involve the same enforcement agents.

Prior studies of local immigration enforcement at the state level have generally found little effect on crime rates and have suggested that 287(g) agreements undermine trust between immigrant communities and law enforcement. In their study of 287(g) agreements in North Carolina, Andrew Forrester and Alex Nowrasteh (2018) find no relationship between the implementation of 287(g) on overall crime rates or crime clearance rates, although they do observe an increase in assaults on police officers in counties that implemented 287(g) programs, which may indicate reduced trust in law enforcement. Miles and Cox (2014) use a differences-in-differences design to analyze the impact of a similar immigration enforcement program, Secure Communities, on crime rates nationwide and also find no effect. Using survey data from San Diego's immigrant community, Tom K. Wong et. al. (2019) assess the hypothetical effect of cooperation between police and ICE on immigrants' willingness to interact with police and access public services. They find that if a 287(g) agreement between San Diego law enforcement and ICE were introduced, undocumented immigrants would be substantially less likely to report crimes.

While researchers have examined the national impact of 287(g) agreements on ethnic segregation (Hall & Rugh 2019), school enrollment (Dee & Murphy 2020), and food security (Potochnik, Chen & Perreira 2017), there has been no analysis at the national level of whether 287(g) agreements reduce violent crime. Such a study is necessary to evaluate the intended purpose of the 287(g) program, which is costly for local law enforcement agencies, leads to racially discrepant policing (Coleman et. al. 2019), and erodes trust with immigrant communities. If the program does not meaningfully reduce violent crime, policymakers may decide that its costs outweigh its benefits. Further, evaluating the effectiveness of the 287(g) program will contribute to a broader understanding of "crimmigration" and its implications.

Based on the existing literature, I hypothesize that a national study of 287(g) agreements will reveal no causal effect on violent crime rates, except possibly through underreporting of crimes within immigrant communities. If 287(g) agreements indeed lead immigrant communities to underreport crimes, I expect the reduction in reported crime rates to be larger for crimes that are especially prone to underreporting (e.g., rape, domestic assault) and smaller for crimes that are very likely to be reported (e.g., murder). I exploit this distinction by analyzing the trends in murder rates alone as a supplement to my analysis of all violent crimes. If 287(g) jurisdictions show larger reductions in all violent crimes than murder, underreporting may have exaggerated the effect of 287(g) agreements in preventing violent crime.

Data and Methods

Ideally, in order to determine the relationship between 287(g) agreements and violent crime, I would obtain a list of all current and historical 287(g) agreements, along with their effective dates. Unfortunately, no official record of inactive 287(g) agreements exists. While ICE maintains a list of jurisdictions with currently active 287(g) agreements, it is difficult to establish when those agreements were initially adopted. Further, ICE provides only an incomplete archive of previous 287(g) agreements. Some organizations that advocate on behalf of immigrants have compiled lists of 287(g) agreements active at particular times; however, these lists rarely contain accurate effective dates of the agreements. For the purposes of this study, I use the Immigrant Legal Resource Center's interactive map of 287(g) agreements as a starting point to determine which county sheriffs' offices signed 287(g) agreements in 2017 or 2018. I then use local news sources and individual sheriff's office websites to determine which of those agreements were still active through the end of 2019. Since only the Jail Enforcement Model (JEM) of 287(g) agreement was available in 2017 and 2018, all agreements in my treatment group are governed by identical memoranda of understanding (MOUs). However, even within the same model, different sheriffs' offices vary in the extent to which they prioritize serious or violent crimes while implementing 287(g). In Cobb County, Florida, more than half of individuals detained under the 287(g) program were initially stopped for minor traffic offenses (Robinson 2021); other counties do not prioritize traffic violations to the same extent. Despite significant local variation in the implementation of 287(g), I use a binary variable to represent the presence or absence of a 287(g) program in a county in a given time period. I claim that this approach is reasonable for two theoretical reasons: first, there is evidence to support the hypothesis that local immigration policy has a strong symbolic effect. In the context of 287(g), this would suggest that much of the effect on local crime rates is attributable to the announcement of a 287(g) agreement rather than its actual implementation. Second, I claim that the standardized language of JEM agreements, the centralized 287(g) training process, and the similar political objectives of 287(g) jurisdictions under the Trump administration limit practical variation in 287(g) implementation during the period under study.

Establishing construct validity and reliability for the dependent variable of interest, violent crime, is also complicated. I will use the FBI's Uniform Crime Reporting Statistics (UCR), which reports four index categories of violent crime: murder and nonnegligent manslaughter, rape, robbery, and aggravated assault. With the exception of rape, all of these categories of crime have had stable, federally standardized definitions throughout the duration of the 287(g) program. The FBI revised its definition of rape in 2011 to include rape victims who are not women. A small number of agencies continued to report rape under the old definition after 2011; in those cases, the UCR statistics do not report the total number of violent crimes in the agency's jurisdiction that year. Since the period under study is after 2011, all statistics included in my analysis will have reported rape using the revised definition. While it is difficult to empirically test the reliability of these statistics, the fact that the FBI's reporting procedures and definitions are standardized and stable over time provides some reassurance of reliability. It is important to note that the FBI explicitly cautions against directly ranking and comparing crime rates between jurisdictions (FBI 2020). In the case of county-level statistics, this is particularly dangerous. While most U.S. counties are predominantly served by county sheriffs' offices, some counties have county police departments, either independently or alongside county sheriffs' offices. Additionally, county-level and city-level law enforcement jurisdictions often overlap. Depending on the jurisdiction, the county crime totals reported to the FBI might include only crimes reported to the county sheriff's office or all crimes reported to any law enforcement agency that took place in the county. For these reasons, this study compares counties to themselves over time rather than to each other in any given year. To justify this strategy, I make the assumption that variation in county reporting practices between 2016 and 2019 is limited and independent of participation in the 287(g) program. While difficult to test, this assumption is typical of the current literature on crime rates in the United States.

Perhaps the most serious limitation of UCR is that these statistics only include crimes that were reported to law enforcement. Some crimes, such as rape and domestic violence, are likely to be underreported, while others, such as murder, are almost always known to law enforcement (Miller & Segal 2019). Since 287(g) agreements theoretically affect immigrants' willingness to report crimes (Wong et. al. 2019), underreporting presents a serious challenge for this study. To assess the extent to which underreporting has affected the results of this analysis, I repeat the analysis using only murder and nonnegligent manslaughter. Since most jurisdictions observe few or no murders in any given year, these data are significantly noisier than violent crime totals. Nevertheless, a quick analysis of changes in murder rates in 287(g) and non-287(g) jurisdictions should suggest the extent to which the results are biased by underreporting.

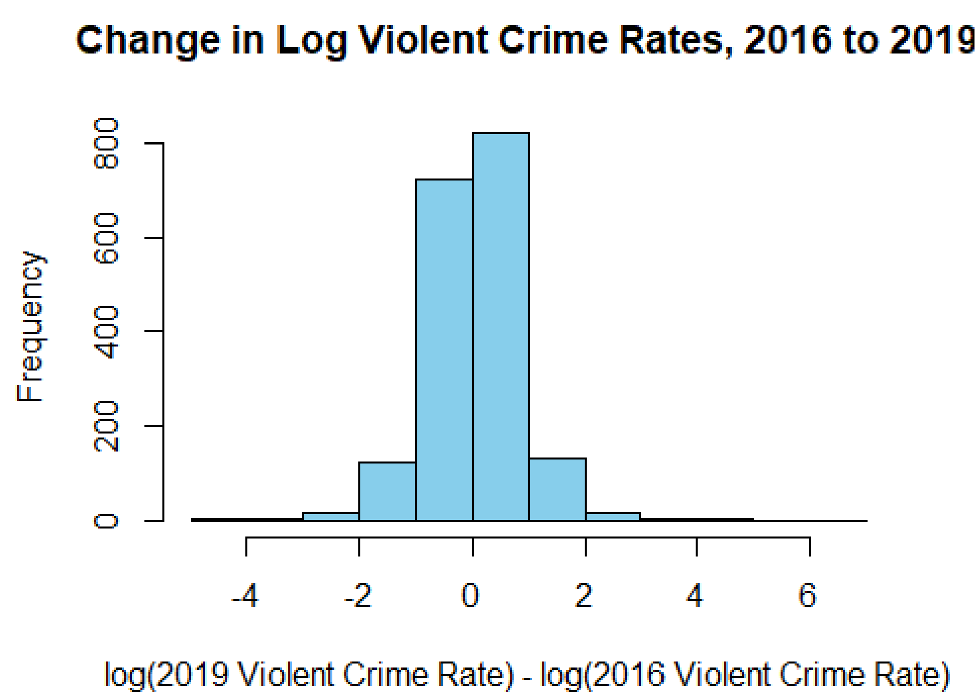
Since the vast majority of local law enforcement agencies that have adopted 287(g) agreements are county sheriffs' offices, I restrict my analysis to counties and county-equivalent regions. Since Connecticut and Alaska do not have counties or equivalent units, these states do not appear in my analysis. I also exclude independent cities that do not belong to any county, which are common in Virginia. Since Louisiana's parishes are functionally equivalent to counties, I make no distinction between parishes and counties. If a county has both a sheriff's office and a police department but only one participates in the UCR statistics, I use the reported value as the crime rate for that county. If both agencies report UCR statistics, I take the larger of the two violent crime totals. I do not add the totals in order to avoid double-counting individual crimes. In most cases, one agency reports far fewer violent crimes than the other, suggesting that one agency acts as the primary law enforcement agency in the county. A brief analysis shows that this approach yields crime rates similar to those I would have expected based on population and demographic composition.

In order to investigate the relationship between 287(g) agreements and violent crime rates, I combine propensity score matching[1] with a differences-in-differences design. First, I build a generalized linear model to predict the likelihood of adopting a 287(g) agreement for each county. For interpretability, I use a logit rather than probit link function.[2] I include several covariates as predictors in this generalized linear model, including racial and ethnic composition of the county, income distribution, Republican vote share in previous elections, prior crime rate, and unemployment. I also include an indicator variable for location in Texas, which is highly predictive of adopting a 287(g) agreement in 2017 or 2018. I then remove insignificant covariates from the model and add interaction terms as needed. By fitting this model, I invoke several assumptions. First, I assume that errors are independent. Since counties are clustered within states, this assumption may be violated. However, since the treatment group only includes 40 counties with complete crime data, a multilevel model would result in prohibitively small sample sizes. Further, a violation of the independence assumption would primarily affect standard errors. To address this challenge, I use robust standard errors. I also assume an approximately linear relationship between the logit of the propensity score and the covariates. Finally, I assume that the variance of the predicted values is a known function of the mean predicted value.

[1] Propensity-score matching is a research strategy used to provide estimates of causal effects in quasi-experimental studies. Researchers first develop a model that predicts whether individual units are likely to receive the treatment. They then use this model to match each treatment unit with a comparable control unit, minimizing confounding. For more information on propensity-score matching and other propensity-score-based strategies, see Rubin 1997.

[2] Researchers frequently use either a logit or a probit link function to transform probabilities into a suitable dependent variable for regression analysis. While the logit and probit functions have different origins and mathematical definitions, they yield very similar results. In this study, I opt for a logit link function to ease the interpretation of my results. For more information on generalized linear models and link functions, see Dobson & Barnett 2018.

After constructing the propensity score model, I check the goodness of fit and then identify the region of common support. That is, I restrict my analysis to observations with values of the propensity score represented in both the treatment and the control groups. I then use statistical software to match counties in the treatment group (those that implemented 287(g) agreements in the given time period) with counties in the control group (those that did not) using the propensity scores. I then check balance in the matched pairs to make sure that the matched treatment and control counties are similar in pre-treatment covariates. If the matched pairs are comparable, I then regress the change in violent crime from 2016 to 2019 on the treatment indicator variable. This outcome variable is a difference of Poisson variables. Since the natural logarithm of each Poisson variable is approximately normally distributed, the difference in the natural logarithms of each crime rate is also approximately normally distributed. I include a histogram of the relevant outcome variable to justify this claim.



As with the propensity score model, the second stage regression invokes a number of model-based assumptions. Again, I assume that errors are independent, although this assumption can be relaxed by using robust standard errors. I also assume an approximately linear relationship between the change in the natural logarithm of crime rates and the treatment. This assumption is satisfied because the treatment variable is binary. If these model-based assumptions are satisfied, I can report the hypothesis testing result and the effect size. I then repeat this analysis using murder rates, as described above.

In order to identify the true average treatment effect on the treated (ATT), I need a few more assumptions. First, I claim that, conditional on propensity score, if no counties had adopted 287(g) agreements, county crime rates would have evolved similarly between the treated and untreated counties. This is a weaker version of the parallel trends assumption typically used in a differences-in-differences analysis, since I am conditioning on the pretreatment characteristics summarized in the propensity score. If the propensity score model includes most potential confounders, this assumption should approximately hold. Second, I assume that the propensity score is positive for all counties—that is, each county in the sample has a nonzero likelihood of adopting a 287(g) agreement. County sheriffs enjoy considerable political independence, so opposition from other local governments or their state governments to the 287(g) program most likely does not prevent sheriffs from signing these agreements. Further, the literature shows that anti-immigrant sentiment is not restricted to parts of the country with significant immigrant populations (Butz & Kehrberg 2019). I conclude that this positivity assumption is reasonable for counties within my sample. Finally, I restrict the interpretation of my results to counties likely to participate in UCR. These counties are not necessarily representative of all U.S. counties or all local law enforcement jurisdictions; on average, counties that participate in UCR are more populous and have a higher proportion of non-Hispanic white residents than the average U.S. county. Notably, counties that adopt 287(g) agreements are more likely to participate in UCR than counties that do not adopt 287(g) agreements, suggesting that my results are largely valid for the population of counties considering 287(g) agreements. UCR participation might also be correlated with unmeasured covariates. It's plausible, for example, that sheriffs' offices which participate in the UCR program are better organized, have more staff, and cooperate more fully with the federal government. My analysis is therefore only valid for the subset of counties that participate in UCR and have a nonzero likelihood of adopting a 287(g) agreement.

Results

After cleaning the data and identifying the treatment and control groups, I conduct a naive t-test to check whether the treatment and control groups have different trends in violent crime rates between 2016 and 2019. To avoid undefined values, I take the natural logarithm of each crime rate plus one.

Test statistic	df	P value	Alternative hypothesis	mean of x	mean of y
-2.477	40.49	0.01752 *	two.sided	-0.4466	0.04992

We find that for treated counties,

$$E(log(v_{2019} + 1) - log(v_{2016} + 1)) = -0.4466121,$$

while for control counties, we have

$$E(log(v_{2019} + 1) - log(v_{2016} + 1)) = 0.04992375.$$

Equivalently, we have

$$\frac{v_{2019} + 1}{v_{2016} + 1} = e^{-0.447} = 0.640$$

for treated counties and

$$\frac{v_{2019} + 1}{v_{2016} + 1} = e^{0.050} = 1.051$$

for control counties. That is, violent crime significantly decreased in the treatment counties and slightly increased in the control counties. This difference is statistically significant, with a t-statistic of -2.477 and a p-value of 0.01752. We therefore reject the null hypothesis that the treatment and control counties have similar trends in crime rates between 2016 and 2019. We can compute the estimated effect size using the standard deviation of the natural logarithm of control counties' violent crime rate, plus 1. This is 1.217, so the effect size is about 0.4 standard deviations. Significantly, this effect is not necessarily causal. Since we have not controlled for potential confounding variables, we can only observe that, on average, counties that implemented 287(g) agreements in 2017 or 2018 had larger decreases in violent crime rates between 2016 and 2019 than counties which did not implement 287(g) agreements in this period.

I now shift to constructing a propensity score model. That is, I use covariates to predict how likely each county is to implement a 287(g) agreement in 2017 or 2018. At first, I include the share of each county's population that was white, Hispanic, and Black based on the 2016 American Communities Survey (ACS). I also include the share of households that speak only English at home, the share of households receiving public assistance, the share of households with an income less than $35,000 per year, and the percentages of the population that are younger than 30 years old and male. Finally, I include the 2016 violent crime rate, the vote share Trump received in the 2016 presidential election, and an indicator variable for location in the state of Texas. I remove the covariates that are insignificant, add significant interaction terms, and report the final logistic model below.

Term	estimate	std.error	statistic	p.value
(Intercept)	17.850342	7.7711490	2.297002	0.0216187
Tx	1.930999	0.4835051	3.993752	0.0000650
Pcthisp	5.781591	3.8742710	1.492304	0.1356194
Pctmale	-39.098443	13.6674961	-2.860688	0.0042272
Pctblack	-4.238142	3.9174540	-1.081861	0.2793141
Pctyoung	-8.111321	3.9970908	-2.029306	0.0424271
unemploy	-35.599527	28.1589501	-1.264235	0.2061457
Pctwhite	-10.314533	4.2942212	-2.401957	0.0163076
pctengonly	10.481953	4.4133336	2.375065	0.0175459
lowincome	-9.877994	2.4505411	-4.030944	0.0000556
pcthisp:pctblack	39.986949	16.3036906	2.452632	0.0141815
unemploy:pctwhite	74.369840	37.4356951	1.986602	0.0469665

Below, I list each term in the propensity score model from lowest to highest p-value, along with a brief explanation of the term.

1. lowincome: percent of households with income < 35k
2. tx: indicator variable for location in Texas
3. pctmale: percentage of the population that is male
4. pcthisp:pctblack: interaction term between the percentage of the population that is Hispanic and the percentage of the population that is Black
5. pctwhite: the percentage of the population that is Caucasian and not Hispanic.
6. pctengonly: the percentage of households that speak only English at home
7. pctyoung: the percentage of the population that is younger than 30

8. unemploy:pctwhite: interaction term between percent white and percent unemployed
9. pcthisp: percentage of the population that is Hispanic
10. unemploy: the percentage of the civilian labor force that is unemployed
11. pctblack: the percentage of the population that is Black

Notably, GOP vote share, education, and percentage of households receiving public assistance are not significant, either when controlling for relevant covariates or alone. Percent Hispanic has a quadratic relationship with log odds of treatment, but this relationship disappears after controlling for other covariates. To understand why GOP vote share is not strongly related to participation in the 287(g) program, it is important to remember that most United States counties are significantly more Republican than the population as a whole. This occurs because Democratic votes tend to cluster in more populous counties. As such, the mean U.S. county has a Republican vote share of about 64%. Counties that enact 287(g) programs are more conservative than the U.S. population as a whole. However, after controlling for relevant covariates, they are about as conservative as the mean county when counties are weighted equally.

Interestingly, neither the 2016 violent crime rate nor its natural logarithm is predictive of whether counties enact 287(g) agreements. This could indicate a distinction between nominal crime rates and subjective public safety. It's possible, for instance, that a large undocumented population makes citizens feel less safe, even if the undocumented population is not responsible for additional crimes. Since undocumented immigrants commit non-immigration-related crimes at a lower rate than citizens but are stereotyped as dangerous criminals (Light et. al. 2020), this explanation is plausible. Alternatively, the adoption of 287(g) agreements could have little to do with even subjective public safety. For example, it's possible that 287(g) agreements primarily address economic anxieties about immigration, even though ICE frames the 287(g) program as a public safety measure.

While constructing the propensity score model, I considered two education variables–the share of the population age 25 and older with a college degree and the share of the same population without a high school diploma or GED. Neither variable was significantly correlated with adoption of 287(g) agreements after controlling for other social and demographic covariates.

The percentage of households receiving public assistance was also uncorrelated with adoption of 287(g) agreements, although higher household income was associated with a greater probability of joining the 287(g) program. Interestingly, the correlation between the proportion of the population receiving public assistance and the proportion earning less than $35,000 was relatively weak, at 0.23. This occurs for a few reasons; first, not all households who are eligible for public assistance choose to accept it. Further, among those who do receive public assistance, only some choose to self-report their benefits to the Census Bureau. Additionally, the cutoff of $35,000 is somewhat arbitrary and does not necessarily correspond to eligibility for public benefits. Depending on the specific program and household characteristics, some households below the threshold may not qualify and some households above the threshold will qualify. Finally, public assistance itself bolsters income and may push some qualifying households above the $35,000 threshold. With the difference between the income and public assistance variables in mind, it is striking that the income variable is more predictive of whether a county adopts a 287(g) agreement. Counties with many low-income households are less likely to adopt 287(g) agreements even after adjusting for race and ethnicity. It's possible that income reflects political preferences not accounted for by 2016 Republican vote share. Alternatively, sheriffs' offices in low-income communities might have fewer resources, making them less likely to apply for and be approved to join the 287(g) program. Low-income communities that prefer harsher policies against immigrants may also be less effective in organizing to reach their policy objectives.

After income, the most significant predictor of whether a county will join the 287(g) program is whether the county is located in Texas. In part, this reflects the time period under study. Most 287(g) agreements in Texas were signed in 2017 or 2018. In Florida, on the other hand, most 287(g) agreements were signed under the Warrant Service Officer (WSO) model, which was introduced in May 2019. The proliferation of 287(g) agreements in Texas around this time period is the result of several factors. As a border state, Texas is home to a relatively large population of undocumented immigrants of Latin American origin. Texas counties also tend to be rural and politically conservative. Additionally, counties within the same state do not operate independently. Texas counties are subject to similar political pressures at the state level. At the same time, county sheriffs' offices influence each others' policies. When one Texas county adopts a 287(g) agreement, neighboring counties have greater awareness of the 287(g) program and may experience political pressure to join. Since the treatment group is very small, only the indicator variable for Texas was significant in the propensity score model; however, with more observations, we would likely see similar clustering effects within different states.

Notably, the percentage of the population that is male also significantly and negatively affects the probability a county will adopt a 287(g) agreement. Undocumented immigrants are more likely to be young men than any other demographic, which suggests that more heavily male counties are likely to have more undocumented immigrants and that male citizens are more likely to compete with undocumented immigrants for jobs. This result is therefore surprising, though it is highly statistically significant.

The next strongest predictor of whether a county will adopt a 287(g) agreement is an interaction term between the percentage of the population that is Hispanic and the percentage that is Black. Interestingly, this term is more significant than either percentage individually. Since the coefficient for this term is positive, we can conclude that among counties with a large Hispanic population, increasing the share of the population that is Black leads to an increase in the probability that the county has adopted a 287(g) agreement. On the other hand, in counties that have a very small Hispanic population, increasing the Black population share actually leads to a decrease in the probability that the county has adopted a 287(g) agreement. One plausible interpretation of this relationship is that Black residents have more favorable views of (predominantly Hispanic) undocumented immigrants when they are less likely to compete with them for jobs and resources. However, the intersection of race, politics, and immigration is complex and difficult to parse. For the purpose of this analysis, this term is primarily included for its predictive power rather than its interpretive clarity. Complicating the relationship between race and 287(g) agreements further, we see that the percentage of white residents is negatively correlated with the likelihood of adopting a 287(g) agreement. One plausible interpretation of this coefficient is that white residents feel most threatened by undocumented immigrants when white people comprise a smaller share of the population. However, this explanation is only plausible when the total share of white residents is relatively large; as people of color account for a larger share of the population, we would expect that the attitudes of white residents would have less influence on local immigration policy.

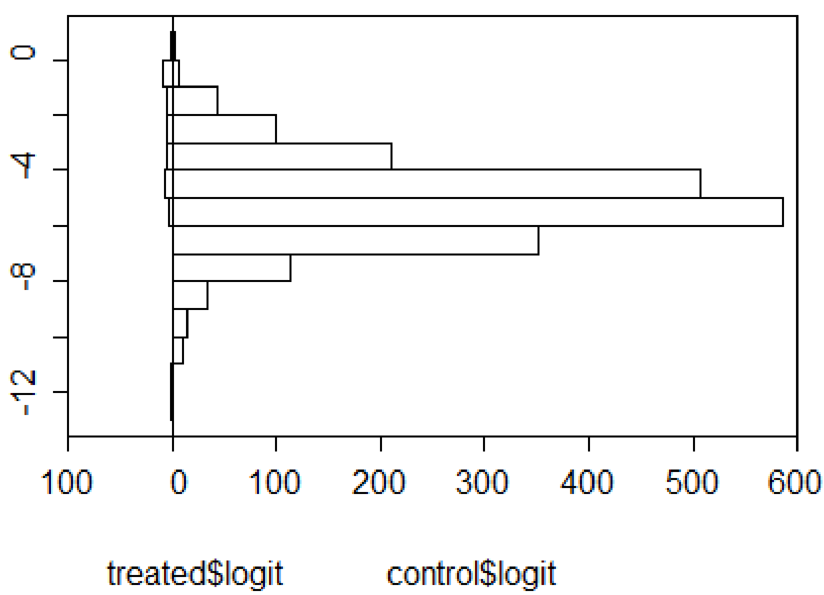
Next, the percent of households in a county that speak only English at home is positively associated with 287(g) agreements. The interpretation of this coefficient is relatively straightforward. Individuals in households that speak only English are much less likely to be recent immigrants and may be less sympathetic to undocumented immigrants. More broadly, it's plausible that multilingualism is linked to more accepting attitudes towards immigrants in general. However, the relationship between monoligualism and race is difficult to interpret. Since most multilingual households in the United States are either Hispanic or Asian American, the monolingual English population may be a proxy for racial characteristics.

The share of a county's population that is younger than 30 years old is negatively correlated with the likelihood of adopting a 287(g) agreement. The simplest explanation of this phenomenon is that younger populations are more welcoming towards immigrants, all else held constant. It's also plausible that criminal justice and public safety issues are more salient to older segments of the population. The sign of this term provides more evidence that the actual crime rate is unrelated to the adoption of 287(g) agreements. Since most crimes in the United States are committed by individuals under 30 years of age (Kearney et. al. 2014), we might have expected that a larger proportion of young people in a county would lead to a more negative subjective assessment of public safety, which in turn could lead to more punitive policies towards immigrants.

The interaction term between the proportion of white residents and the civilian unemployment rate is also positive and significant. Since the unemployment rate alone is not significant, the best interpretation of this coefficient is that a high unemployment rate increases the likelihood of adopting a 287(g) agreement when a large proportion of the population is white but has little effect if the population is predominantly nonwhite. It's plausible that unemployed and economically insecure white residents are more likely to favor harsh policies towards undocumented immigrants.

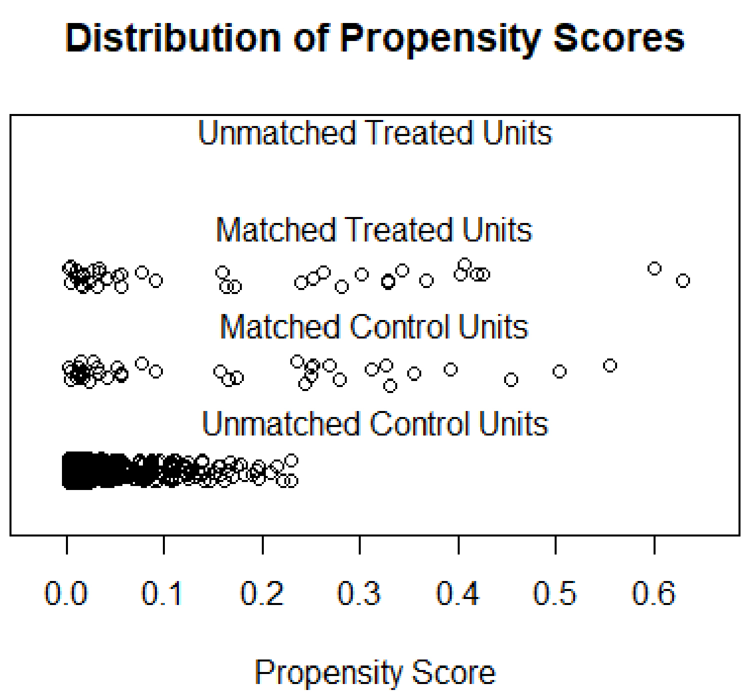
Finally, I include unemployment, percent Hispanic, and percent Black as covariates even though they are not significant because they are included in significant interaction terms. After controlling for other covariates and adding relevant interaction terms, these covariates provide little additional information about whether a county will adopt a 287(g) agreement. That's not to say that economic conditions and race are unrelated to 287(g) agreements; on the contrary, economic insecurity and race are connected to immigration policy in intimate and complex ways. However, in this model, interaction terms capture these relationships more effectively than linear regression terms.

After developing a propensity score model, I identify the region of common support between the treatment and control groups. As expected, the treatment group generally has larger values of the propensity score and its logit. Below, I include a histogram comparing the distribution of propensity scores between the two groups.



Treatment	n_counties	min_logit	med_logit	max_logit
0	1982	-12.277805	-5.197339	0.8857348
1	40	-5.846602	-2.682851	0.4622474

I restrict my sample to the region of common support–that is, the counties with logit propensity scores represented in both the treatment and control groups. After restricting the sample, there are 1,533 control units and 40 treated units. I then use the MatchIt package to match each treated county to one control county. Below, I include a plot of the propensity scores of the matched and unmatched treated and control units.



As shown, the matched control units have a distribution of propensity scores more similar to the treated units than to the unmatched control units. This is ideal, since the matched control units should be similar to the matched treated units on all relevant covariates. To formally check balance within the matched pairs, I display the standardized difference and variance ratio in logit propensity scores for the matched pairs.

Balance Measures

	Type	Diff.Adj	V.Ratio.Adj
Distance	Distance	0.0800	1.2701
Logit	Contin.	0.0365	1.0912

Sample Sizes

	Control	Treated
All	1533	40
Matched	40	40
Unmatched	1493	0

We see that the balance within the matched pairs is quite good, with a standardized difference of only 0.08. I include a table showing each treatment county and its matched control county.

Next, I regress the difference in log 2019 and 2016 crime rates on the treatment variable for the matched data. I display the exponent of the regression coefficient, which should be interpreted as the ratio of average violent crime rates plus one in treatment and control counties, respectively. I use robust standard errors and display a 95% confidence interval for the exponentiated coefficient.

	Exp(Est.)	2.5%	97.5%	t val.	p
(Intercept)	0.71086	0.52019	0.97141	-2.14201	0.03219

Treatment	0.90003	0.54644	1.48241	-0.41372	0.67908

The treatment effect is not significant, with a t-statistic of -0.414 and a p-value of 0.680. Strikingly, the R^2 value for the model is only 0.002, indicating that 287(g) agreements account for only 0.2% of the variation in the change in log crime rates among the matched units. This result holds using different matching methods, including full matching and many-to-one matching, although the standardized difference is minimized by one-to-one nearest neighbor matching.

I now repeat the analysis using murder rates rather than violent crime rates with the same matched pairs. In this case, the outcome variable is

$$log(m_{2019} + 1) - log(m_{2016} + 1),$$

where m_{2016} and m_{2019} are annual per capita murder rates. As before, I use robust standard errors and display the 95% confidence interval for the exponentiated coefficient.

	Exp(Est.)	2.5%	97.5%	t val.	p
(Intercept)	1.00000	0.99999	1.00000	-1.09598	0.27309
Treatment	1.00002	1.00000	1.00003	2.01896	0.04349

Strikingly, per capita murder rates actually increased slightly in 287(g) counties relative to control counties. This result is statistically significant, with a t-statistic of 2.019 and a p-value of 0.043. However, it is possible that since per capita murder rates are relatively low, adding one inside the natural logarithm may have introduced bias. To ensure that the significant result is not an artifact of this distortion, I also run the model using the actual number of murders as the dependent variable. The results, which are quite similar, are shown below.

	Est.	S.E.	t val.	p
(Intercept)	-0.00000	0.00000	-1.09597	0.27755
Treatment	0.00002	0.00001	2.01895	0.04804

Since murder rates are noisy in small counties and the treatment group is very small, we should be cautious in interpreting this effect causally. However, this result suggests that underreporting may have biased the results for violent crime rates. In particular, it is plausible that violent crimes other than murder were less likely to be reported after counties implemented the 287(g) program. This underreporting effect may have offset a slight increase in the actual number of violent crimes that occurred. Such an increase could have arisen from local law enforcement devoting more resources to immigration enforcement and less to local violent crimes. It is also important to note that underreporting itself has a causal effect on the number of crimes committed in a particular jurisdiction. When prospective offenders believe that crimes are less likely to be reported to police, they are more likely to commit them.

Naturally, far more research is necessary to determine whether 287(g) agreements lead to an increase in murders or a decrease in crime reporting rates. Such analysis is beyond the scope of this paper and requires more detailed data on 287(g) jurisdictions and crime reporting. Nevertheless, the results of this section are broadly consistent with the literature suggesting that 287(g) agreements undermine trust between law enforcement and immigrant communities, making law enforcement less effective.

Discussion

In this analysis, I examine whether 287(g) agreements were associated with reductions in county violent crime rates between 2016 and 2019 after controlling for potential confounding variables. While a naive analysis reveals that counties which implemented 287(g) agreements generally experienced larger decreases in violent crime between 2016 and 2019, this effect disappears after controlling for relevant confounders through propensity score matching. In particular, counties that adopted 287(g) agreements differ from counties that did not in their demographic, socioeconomic, and geographic characteristics. 287(g) jurisdictions were generally wealthier, less white, less male, older, more monolingual, and more likely to be located in Texas than the average U.S. county. Race, ethnicity, and unemployment also interact in complex ways to influence whether counties adopt 287(g) agreements. Interestingly, 287(g) jurisdictions are not more politically conservative than non-287(g) jurisdictions after controlling for other demographic characteristics. Counties that adopted 287(g) agreements also did not have higher prior crime rates than counties that did not. The propensity score model generally suggests that racial and socioeconomic anxieties, rather than concerns about public safety, were more likely to lead to 287(g) agreements.

After accounting for these differences, counties that adopted 287(g) agreements between 2017 and 2018 did not have larger reductions in violent crime rates than counties that did not adopt these agreements during this time period. This finding is consistent with the results in Forrester and Nowratesh (2018) and Miles and Cox (2014), and it suggests that the 287(g) program is not effective in promoting public safety. Further, after adjusting for confounding variables, 287(g) jurisdictions had larger increases in murder rates relative to counties without 287(g) agreements. The latter result suggests that 287(g) agreements might lower the rate at which immigrant communities report crimes other than murder, masking an increase in the overall number of crimes committed. While this result is based on a small sample size and a relatively noisy outcome variable, it provides some evidence in support of the claim that 287(g) agreements weaken relationships between law enforcement and immigrant communities.

The results of this analysis, along with the current literature, cast doubt on the effectiveness of the 287(g) program as a public safety measure. However, the reasons for its failure to reduce violent crime are unclear. It’s plausible that, in practice, counties that implement the 287(g) program engage in relatively little immigration enforcement activity, especially compared to non-287(g) jurisdictions with similar demographic characteristics. Without detailed data on detentions and deportations associated with the 287(g) agreement, this explanation is difficult to evaluate. This hypothesis does not exclude the possibility that the existence of a 287(g) agreement in a county leads immigrant communities to underreport crimes as witnesses or victims. Indeed, Wong et. al.’s research provides evidence that awareness of 287(g) agreements, regardless of their actual implementation, can contribute to underreporting. It’s plausible that underreporting itself, rather than any particular policing practices, leads to an increase in actual but not reported violent crimes.

Another possibility is that 287(g) jurisdictions shift law enforcement resources toward immigrant communities, which makes them somewhat less effective in preventing crime generally. This hypothesis is consistent with Michael Coon’s analysis of the 287(g) program in Frederick County, MD, which found that police became much more likely to arrest Hispanic individuals after the implementation of the agreement. In this case, the program directly affects violent crime rates through policing practices. Any underreporting of crime rates masks increases in the actual numbers of violent crimes. To evaluate this explanation, researchers will need more detailed information about the ethnicities or immigration statuses of arrestees. Qualitative evidence on policing practices and immigrant communities' perceptions of the 287(g) agreement would also provide valuable perspectives on both interpretations. Self-reported crime victimization data could also shed light on the extent and nature of crime underreporting arising from 287(g) agreements.

Regardless of whether 287(g) agreements directly affect crime through policing practices, this study undermines a common argument in favor of the 287(g) program. ICE advertises the program to local law enforcement agencies as an effective public safety measure. If that argument contradicts the empirical evidence, local law enforcement agencies may be less likely to sign 287(g) agreements in the future. Of course, violent crime reduction is not the only possible justification for local involvement in immigration enforcement. If law enforcement agencies perceive reducing the number of undocumented immigrants in their jurisdictions as an end in itself, the 287(g) program may remain attractive. Implementing a 287(g) agreement may also serve a symbolic political function; sheriffs who sign 287(g) agreements signal to their constituents and to the federal government that they favor punitive immigration policies.

On the other hand, the results of this study are encouraging for jurisdictions that are considering more lenient local immigration policies but are concerned about the possibility of higher crime rates. 287(g) agreements represent a high degree of cooperation between local law enforcement agencies and ICE, at least symbolically and potentially practically. If punitive law enforcement policies towards immigrants are ineffective in reducing violent crime, it's unlikely that more lenient policies will result in a dramatic increase in violent crime. More lenient policies towards immigrants might also improve trust between undocumented immigrants and law enforcement, resulting in higher crime reporting rates. If this is the case, so-called sanctuary policies at the state and local level might have a neutral or beneficial effect on public safety. Naturally, more research is necessary to evaluate this argument. For instance, it's possible that eliminating punitive immigration policies where they already exist has a different effect on crime than never implementing punitive policies in the first place. Nevertheless, this study supports a cautious optimism about the effect of lenient local immigration policies on public safety.

References

American Immigration Council. (2021, June 8). *The 287(g) Program: An Overview.* https://www.americanimmigrationcouncil.org/research/287g-program-immigration

Arriaga, F. (2017). Relationships between the Public and Crimmigration Entities in North Carolina: A 287(g) Program Focus. Sociology of Race and Ethnicity, 3(3), 417–431.

Butz, A. M., & Kehrberg, J. E. (2019). Anti-Immigrant Sentiment and the Adoption of State Immigration Policy. *Policy Studies Journal, 47*(3), 605-623.

Coleman, M., Kocher, A., Garip, F., Gleeson, S., & Hall, M. (2019). Rethinking the "Gold Standard" of Racial Profiling: §287(g), Secure Communities and Racially Discrepant Police Power. *American Behavioral Scientist, 63(9),* 1185-1220.

Coon, M. (2017). Local Immigration Enforcement and Arrests of the Hispanic Population. Journal on Migration and Human Security, 5(3), 645–666. https://doi.org/10.1177/233150241700500305

Dee, T. S. & Murphy, M. (2020). Vanished Classmates: The Effects of Local Immigration Enforcement on School Enrollment. *American Educational Research Journal, 57(2),* 694.

Dobson, A., & Barnett, A. (2018). An Introduction to Generalized Linear Models. CRC Press.

Federal Bureau of Investigation. (2017). *Crime in the United States, 2016.* https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016

Federal Bureau of Investigation. (2020). *Crime in the United States, 2019.* https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/home

Forrester, A., & Nowrasteh, A. (2018). (Rep.). Cato Institute. Retrieved February 18, 2021, from http://www.jstor.org/stable/resrep16983

Hall, M. & Rugh, J. (2019). Local Immigration Enforcement and Latino Segregation. *Conference Papers –American Sociological Association,* 1-30.

Immigrant Legal Resource Center. (2020, October 21). *National Map of 287(g) Agreements.* https://www.ilrc.org/national-map-287g-agreements

Immigration and Customs Enforcement (2017). *Priority Enforcement Program.* https://www.ice.gov/pep

Immigration and Customs Enforcement (2020). *287(g) – Memorandums of Agreement/Understanding (Old).* https://www.ice.gov/identify-and-arrest/287g/287g-archive#

Immigration and Customs Enforcement. (2021). *Delegation of Immigration Authority Section 287(g) Immigration and Nationality Act.* https://www.ice.gov/identify-and-arrest/287g

Kearney, M. S., Harris, B. H., Jacome, E., & Parker, L. (2014). Ten Economic Facts about Crime and Incarceration in the United States.

Light, M. T., He, J., & Robey, J. P. (2020). Comparing Crime Rates Between Undocumented Immigrants, Legal Immigrants, and Native-Born U.S. Citizens in Texas. *Proceedings of the National Academy of Sciences of the United States of America, 117*(51), 32340-32347.

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., & Ruggles, S. (2021). IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V16.0

Miles, T., & Cox, A. (2014). Does Immigration Enforcement Reduce Crime? Evidence from Secure Communities. *The Journal of Law & Economics, 57*(4), 937-973. doi:10.1086/680935

Miller, A. R. and Segal, C. (2019). Do Female Police Officers Improve Law Enforcement Quality? Effects on crime reporting and domestic violence. *The Review of Economic Studies 86*(5), 2220-2247.

Potochnick, S., Chen, J.-H., & Perreira, K. (2017). Local-Level Immigration Enforcement and Food Insecurity Risk among Hispanic Immigrant Families with Children: National-Level Evidence. *Journal of Immigrant & Minority Health*, *19(5)*, 1042.

Provine, D. M., Varsanyi, M. W., Lewis, P. G., & Decker, S. H. (2016). Policing Immigrants: Local law enforcement on the front lines. University of Chicago Press.

Robinson, A. (2021, January 20). *Owens Ends Participation in 287(g) Program.* Cobb County Courier. https://cobbcountycourier.com/2021/01/owens-ends-cobb-participation-in-287g-program/

Rubin, D. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. Annals of Internal Medicine, 127(8), 757-763.

Stumpf, J. (2006). The crimmigration crisis: Immigrants, crime, and sovereign power. American University Law Review, 56(2), 367-420.

United States v. Brignoni-Ponce, 422 U.S. 873 (1975).

Wong, T. K., Shklyan, K., Isorena, A., & Peng, S. (2019). The Impact of Interior Immigration Enforcement on the Day-to-Day Behaviors of Undocumented Immigrants. *U.S. Immigration Policy Center.* Working Paper 1.