Exploring the Relationship Between Economic Indicators and the Successful Arrest Rate in the City of Chicago

Pio Joe Sleiman, University of Toronto

Abstract

This paper investigates the relationship between the Successful Arrest Rate (SAR) in Chicago City, Illinois and socioeconomic variables between 2002 and 2022. The SAR is defined as the proportion of reported crimes for which an arrest has been made, and it is aggregated by year on a community level for all 77 communities in the city. Using data from the Chicago Police Department, this study examines the potential influence of socioeconomic variables such as poverty, unemployment, education, and healthcare level on the successful arrest rate at a community-level. The findings of this study indicate statistically significant associations between the successful arrest rate and some socioeconomic variables, however, the effect sizes were small and not considered to have significant implications. Interestingly, relatively large associations were found relating the total number of reports and arrests to the SAR at a significant level. It is important to note that this study was unable to analyze any causal relationships between the observed variables. Therefore, further research is needed to better understand the complex relationship between socioeconomic variables and successful arrest rates in Chicago city.

Introduction

The Successful Arrest Rate measures the proportion of crimes for which an arrest has been made out of the total number of reports; for this study, it is aggregated on a community-level from 2002-2022. This should not be confused with "arrest rate" which explains the number of arrests compared to the population, or "clearance rate", which refers to the proportion of reported crimes that have been cleared.

Researchers have come to find that throughout the last few years, Chicago's arrest rate and clearance rate have been decreasing steadily, while crime rates increase (Bloomberg, 2023). Additionally, we see a great amount of variation in economic welfare between the 77 communities throughout this time (Semuels, 2021). In light of this data, this study focuses on studying the relationship between economic welfare and the SAR in order to better understand the socioeconomic factors associated with it, if any. The study uses a range of different variables that provide measures of economic welfare, such as the unemployment rate, a hardship index, the percent of households below poverty, among others. In addition, variables reporting the total number of reports and arrests are also used as predictors of the SAR.

This is a relationship that has yet to be explored in this context, likely because of the difficulty of obtaining data that adequately represents the economic welfare of all communities of a city over a lengthy time period, such as twenty years, and over many variables, such as unemployment, birth rate, etc… Additionally, the successful arrest rate is not generally used as a measure of police effectiveness, but due to limitations on the available data, it had to be used in place of the clearance rate.

With this being said, a plethora of studies conducted have analyzed the causational relationship between a specific socioeconomic measure and arrest rates. Notably, "Youth, Underemployment, and

Property Crime: Differential Effects of Job Availability and Job Quality on Juvenile and Young Adult Arrest Rates", conducted by Emilie Allan and Darrell Steffensmeier, which studies the relationship between underemployment and the arrest rate among young adults (1989). Likewise, "The Effect of Changes in Intraracial Income Inequality and Educational Attainment on Changes in Arrest Rates for African Americans and Whites, 1957 to 1990", conducted by Gary LaFree and Kriss Drass, which studies the relationship between income inequality, educational attainment, and the arrest rate (1996).

In contrast to the previous studies, this paper focuses on the communities of Chicago City, Illinois over the 2002-2022 period. By analyzing the observed data, we can gain insights into the potential influence of socioeconomic factors on the SAR and further our understanding of the complex relationship between crime and social conditions in Chicago.

Data

The initial crime data used is collected from the Chicago Police Department, and has details regarding every report filed from January 2001 to January 2023. These details include the time, nature, location, community area number, and whether or not an arrest has been made, including other variables that were irrelevant to the study (Department, 2023)

The data was then merged with another dataset containing the names of each community area to ease interpretations and gain a better grasp of the results. Afterwards, the data was aggregated by community area and year, to include the year, total number of reports, total number of arrests, and the SAR of each community. This leaved the data with 1,617 observations, that being 77 communities over 21 years.

Furthermore, the aggregated-crime data was then joined by socioeconomic data for each community from three different sources. Firstly, socioeconomic indicators (such as per capita income, hardship index, and unemployment) on a community-level collected between 2008-2012 by the US Census Bureau was added to each community at every year (Bureau, 2014). Secondly, public health data, collected by the US Census and the Illinois Department of Public Health (IDPH), was added to the original data (Illinois Department of Public Health (IDPH) and U.S. Census Bureau, 2013). This includes variables relating to the Birth Rate, and crowded housing that were measured throughout 2005-2011. Lastly, using data collected by the IDPH, the 2010 life expectancy of each community was added to the data (IDPH, 2014).

With this, the total of number of observations remained at 1,617 observations, but the number of variables being measured and taken into account for increased. It is important to note that the socioeconomic data was gathered in different time frames, but set for all years. This means that the data used does not properly represent the changes that could have happened over time. This is because of limitations with accessing data that is continuous over the period of the study and at a community-level.

Summary Statistics and Visualizations

The following are visualizations and tables that better present the data being worked with and showcase the importance of the question being addressed by the paper.

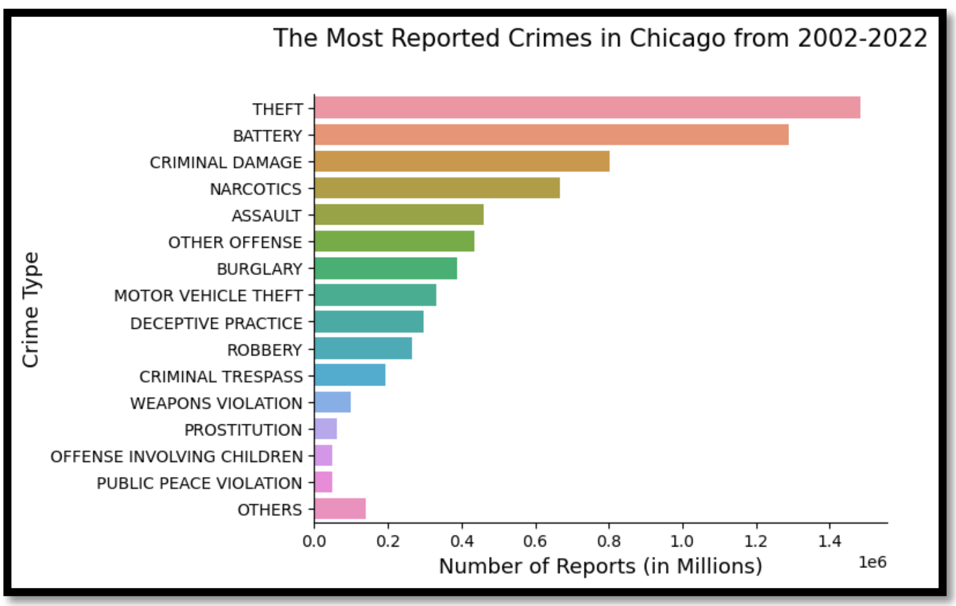
Table 1

Correlation Matrix	**Year**	**Total Reports**	**Total Arrests**	**Successful Arrest Rate**
Year	*1.00*	*-0.27*	*-0.32*	*-0.51*
Total Reports	*-0.27*	*1.00*	*0.94*	*0.53*
Total Arrests	*-0.32*	*0.94*	*1.00*	*0.68*
Successful Arrest Rate	*-0.51*	*0.53*	*0.68*	*1.00*

Note: The table shows the correlation between the four key variables of the aggregated crime data: Year, Total Reports, Total Arrests, and the SAR.

The table above shows a positive correlation between total arrests and the SAR, at 0.68, which is relatively expected considering that the SAR includes total arrests in its measurement. In addition, the table shows a moderately positive correlation between total reports and the SAR, at approximately, 0.53, which is not intuitive since it is expected that the more reports filed, the less resources departments are able to allocate per report, leading to a lower SAR; however, this does not appear to be the case. Therefore, it is highly likely that there are other variables, specifically socioeconomic, that contribute to this surprising result.

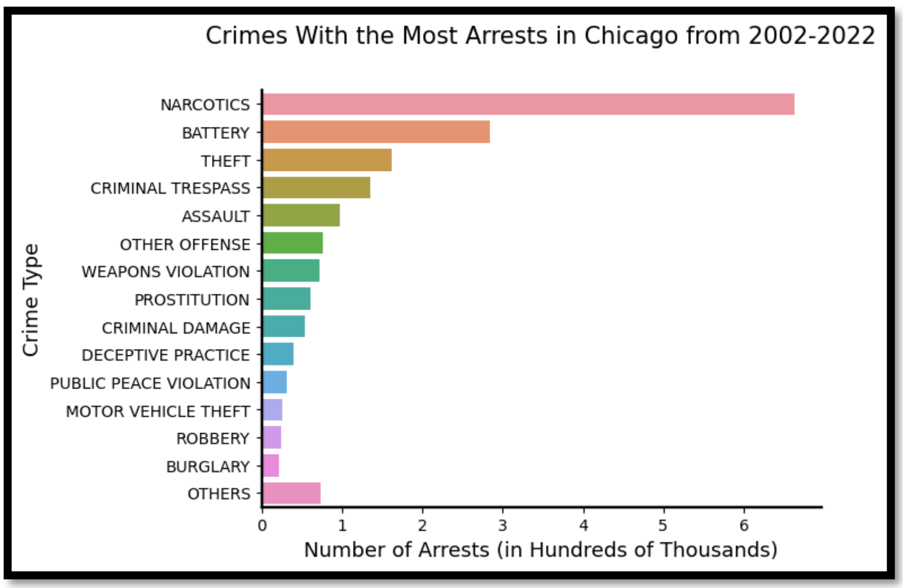
Figure 1



The above graph shows the most reported crimes in Chicago over the last 20 years. "Theft" is the most reported crime, closely followed by "Battery," which is defined as the use of force against another that results in harmful contact. This helps us see that the variety of crimes being reported is relatively small, with the top four most reported crimes representing nearly half of all the reported crimes in that time.

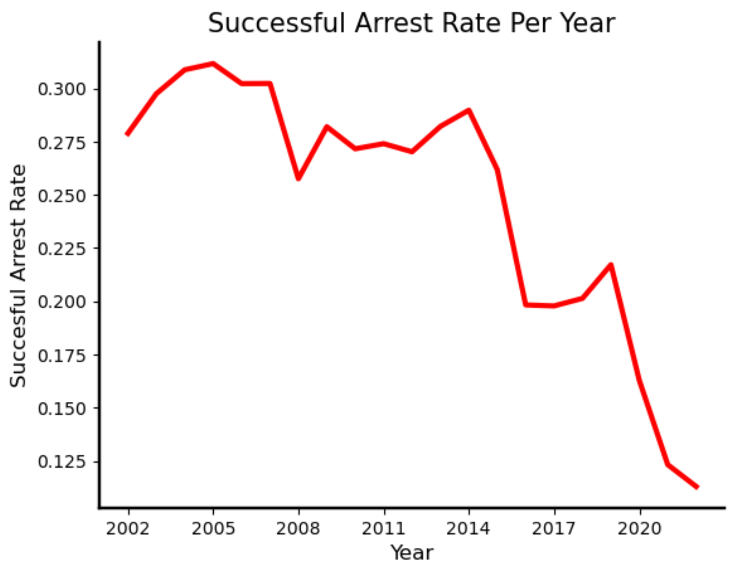
This, paired with the previously identified low SAR, could mean that we are witnessing the same criminals commit similar offenses multiple times since they are not being caught, rather than a large pool of criminal offenders committing the same crime. Alternatively, the police department could have a difficulty with limiting and making arrests for these specific crimes, making them more common compared to others. In order to further analyze this, the following graph will illustrate the types of crime for which the most arrests have been made.

Figure 2



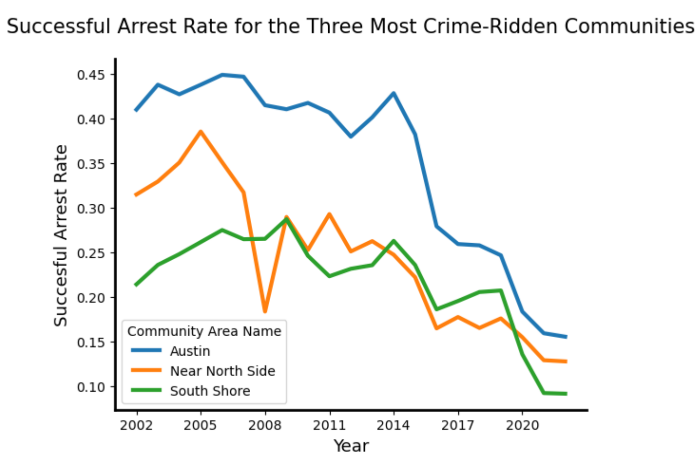
"Narcotics" is the crime with the most arrests, having over double the amount of arrests as second place, "Battery". Importantly, "Theft", the highest reported crime, is only the third highest in terms of arrests, suggesting that a very large majority of theft crimes are reported without an arrest. These results follow along with the previously stated hypothesis, and show that the SAR varies significantly among the different crimes committed.

Figure 3



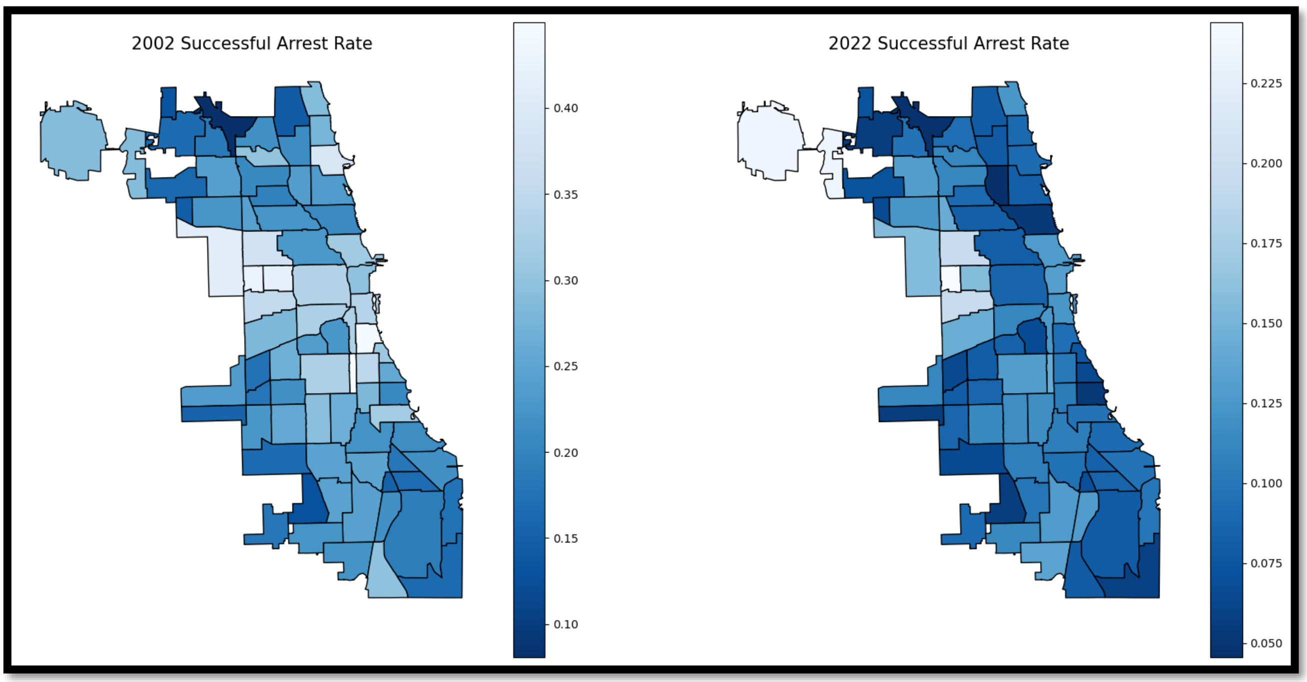
Here, we see a visualization of the Successful Arrest Rate of each year over the last 20 years in Chicago. Most importantly, there is a clear downward trend, showing that the SAR is likely to continue decreasing for the foreseeable future. There is a global maximum of the observed data in 2004, where the rate reached as high as 30%; apart from this peak, we see a local maximum approximately 10 years later in 2014 at about 28.5%, but no point in the recent past comes close to this number. The global minimum of the data lies in 2022, the most recent set of observations, at nearly 10%.

Figure 4



The figure above depicts the decrease in SAR observed in the three most crime-ridden communities, Austin, Near North Side, and South Shore. We see that in 2002, there was a relatively moderate difference between all three communities, with Austin, the most crime-ridden community, having an SAR as high as approximately 0.4. However, as we move further to the right, we see the difference between these communities' SAR decrease significantly as the SAR itself decreases. Seeing as there is quite a lot of variation between the three communities, the decrease suggests that there are other variables playing a role in the outcome that we see.

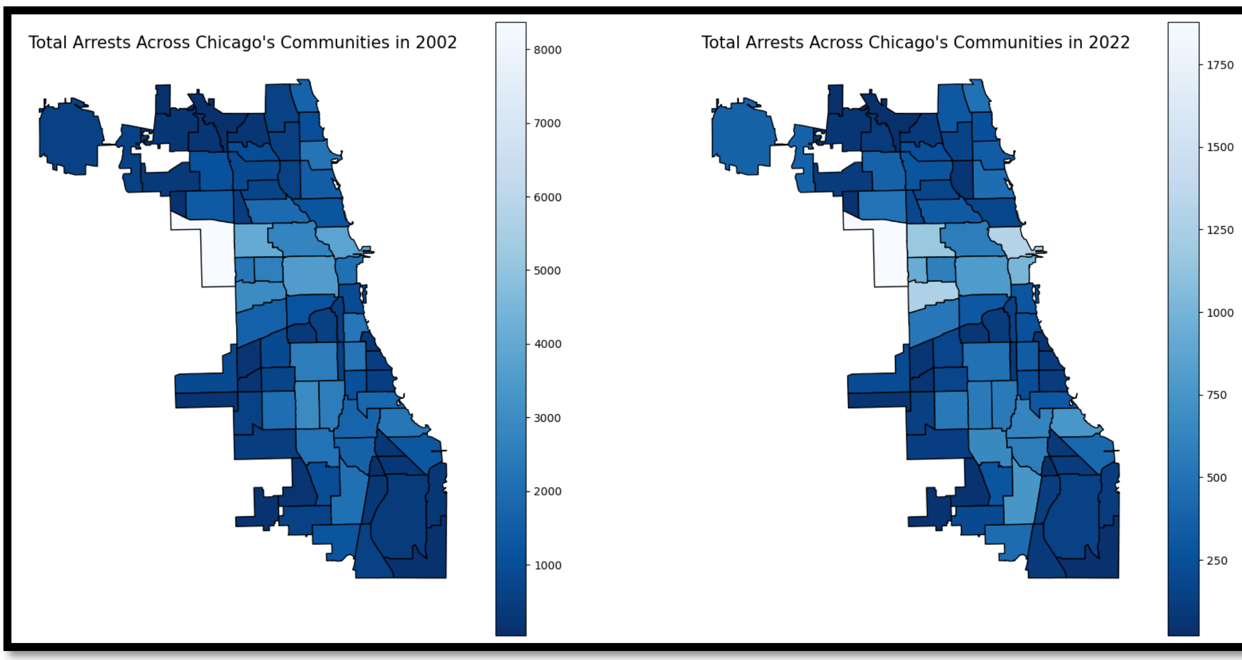
Figure 5

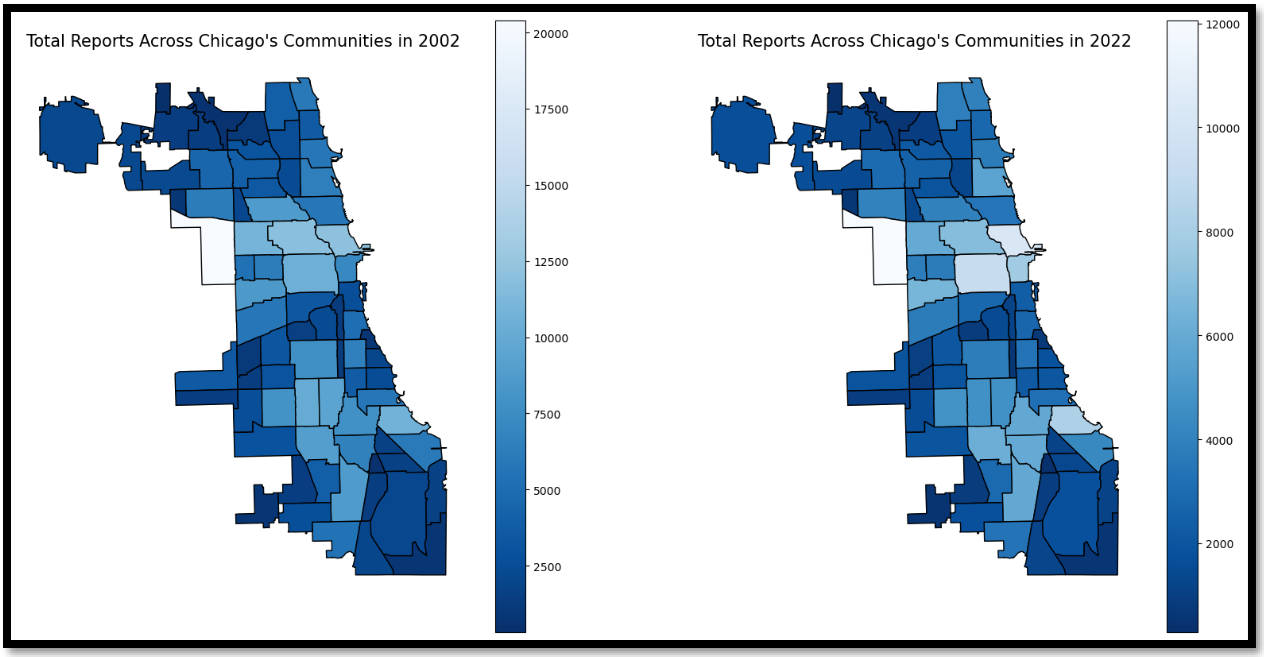


The main idea we learn from this visualization is that nearly all communities suffer from a decrease in their successful arrest rate. This helps us infer that the average SAR for all communities is not being brought down by a few bad communities, but rather a relatively similar decrease in all of the communities.

With this in mind, we can now shift our focus to studying variables that alter the entire city of Chicago rather than a few communities. It is also important to note that one of the only areas to experience an increase in their successful arrest rate is the community in which the Chicago International Airport is located (top left of the map); this increase is likely because of a shift of police department focus to these areas to ensure safety of incoming tourists.

Figure 6





The most important observation that can be made from the above maps is the fact that the number of total arrests per community has decreased significantly, but the total number of reports has remained relatively similar. The map really puts into context how rapid of a decline the successful arrest rate has taken in Chicago. We also notice that, in general, communities have held similar ranks relative to other communities, meaning that we are witnessing a general, and equal, decrease among all communities and not just a selected few.

Regression Results

One of the best methods of studying the association between variables is to use a linear regression model. In this case, seeing as our variables are likely complexly related, a multiple linear regression model must be used. This is because a multiple linear regression displays the association between the dependent and explanatory variable while holding all other variables fixed.

Additionally, for this section of the study, the SAR is measured on a 0-100 scale, which can effectively be thought of as the *percentage* of reports for which an arrest has been made. This is because it makes each table easier to interpret as the previous SAR scale (0-1) deals with small numbers.

Regression Table 1

	Dependent variable:		
	Model 1	Model 2	Model 3
	(1)	(2)	(3)
Total Arrests		5.594***	18.023***
		(0.117)	(0.192)
Total Reports	5.336***		-16.880***
	(0.201)		(0.293)
Population			-0.941***
			(0.127)
Constant	-20.862***	-13.908***	51.348***
	(1.620)	(0.761)	(0.883)
Observations	1,617	1,617	1,617
R^2	0.303	0.585	0.925
Adjusted R^2	0.303	0.585	0.924
Residual Std. Error	7.357(df = 1615)	5.674(df = 1615)	2.423(df = 1613)
F Statistic	702.532*** (df = 1.0; 1615.0)	2280.842*** (df = 1.0; 1615.0)	6584.126*** (df = 3.0; 1613.0)
Note:			*p<0.1; **p<0.05; ***p<0.01

The above table shows three regression models. The three models are related to crime variables, and not with socioeconomic factors. This is to be able to compare the effectiveness of a model using solely crime data, and a model using both crime data and socioeconomic data.

The first is a simple linear regression between the Total Reports and SAR; this model is relatively ineffective as only 30% of the variation in the SAR is explained by variations in the total number of arrests. Moreover, we see a positive association between the total number of reports and the SAR, which is unexpected. One would believe that the greater the number of reports, the lower the proportion of reports with an arrest, but the model suggests otherwise.

The second model, shown in column (2), is a simple linear regression between Total Arrests and the SAR. This model, similar to the first, is rather ineffective as only 59% of the variation in the SAR is explained by variation in the total number of arrests. Furthermore, the relationship between the SAR and the total number of arrests demonstrated by the model is intuitive: the greater the total number of arrests made, the greater the SAR.

The third model, show in column (3) is by far the best of the three. This is a multiple linear regression model with three dependent variables: total arrests, total reports, and population. Approximately 92.5% of the variation in SAR is explained by the variation between the three variables, meaning that the model is a good predictor of the successful arrest rate when given values of the total number of reports, arrests and the population. The following is the formal representation of the model:

$$\mathbf{\widehat{SAR}_i = 51.358 + 18.023 * (Total\ Arrests)_i - 16.88 * (Total\ Reports)_i - 0.941(Population)_i}$$

The intercept, in this case, has no interpretation as it outside the scope of the model. The regression suggests that, when holding fixed the total number of reports and population, a 10% increase in total arrests is associated with, on average, a 1.8 percentage point increase in the SAR. Furthermore, when holding fixed the total number of arrests and the population, a 10% increase in total reports is associated with, on average, a 1.7 percentage point decrease in the SAR. Lastly, the model suggests that, when holding fixed the total number of reports and arrests, a 10% increase in population is associated with, on average, 0.09 percentage point decrease in the SAR, which is a relatively small change. Overall, the model is relatively significant considering the high statistical significance of each slope coefficient and the moderate change in each variable when holding the others fixed. It is important to note that a 1 percentage point increase in this context could equate to thousands of more reports having an arrest.

Moreover, we can see that Model 3 follows the previously outlined intuitive relationship between total reports and the SAR, meaning that controlling for total arrests and population allowed for the model to self-correct. This is important as it implies that the model makes sense and is better suited for our predictions.

Regression Table 2

	Dependent variable:			
	Model 4	Model 5	Model 6	Model 7
	(1)	(2)	(3)	(4)
Total Arrests			17.964***	18.078***
			(0.200)	(0.203)
Total Reports			-16.991***	-17.505***
			(0.326)	(0.350)
Population			-0.827***	-0.471**
			(0.190)	(0.216)
Per Capita Income			2.013***	0.933***
			(0.421)	(0.304)
Unemployment	0.016	-0.174***	-0.011	-0.008
	(0.042)	(0.061)	(0.017)	(0.017)
Birth Rate	0.442***	0.211***		0.063***
	(0.067)	(0.073)		(0.023)
No High School Diploma	-0.016	0.208***		0.004
	(0.020)	(0.048)		(0.010)
Percent Households Below Poverty	0.341***	0.368***		0.070***
	(0.027)	(0.034)		(0.010)
Constant	7.554***	73.639***	29.338***	39.367***
	(0.939)	(8.741)	(4.558)	(3.302)
Observations	1,617	1,617	1,617	1,617
R^2	0.269	0.295	0.926	0.927
Adjusted R^2	0.267	0.292	0.926	0.927
Residual Std. Error	7.542(df = 1612)	7.413(df = 1610)	2.403(df = 1610)	2.379(df = 1608)
F Statistic	148.199*** (df = 4.0; 1612.0)	112.064*** (df = 6.0; 1610.0)	3353.109*** (df = 6.0; 1610.0)	2569.710*** (df = 8.0; 1608.0)
Note:				*p<0.1; **p<0.05; ***p<0.01

The above table shows 4 separate multiple linear regression models, with each of the first two controlling for a different set of socioeconomic variables, and the third including crime data that was used in table 1.

It is clear to see that the two models that use only socioeconomic factors, shown in columns (1) and (2), are largely ineffective, with less than 30% of the variation in the SAR being explained by variation in the dependent variables. However, the third and fourth models which include crime data do a much better job, with approximately 93% of the variation in the SAR being explained by variation in the dependent variables.

With this in mind, the model 3 of Regression table 1 is preferred over models 6 and 7 in Regression Table 2. This is because of the large amount of variables used in the last two models. The number of variables in the last two models is an issue because it poses the large risk of overfitting. Overfitting occurs when the model retains a high success rate in predictions simply because it covers every possible variation of observations. In addition to this, the last two models are only marginally better predictors of the SAR than Model 3, meaning that we would be running the risk of overfitting with no benefit by using models 6 and 7 over model 3.

Overall, it is important to note that models 4-7 show us that the socioeconomic indicators are quite bad predictors of the SAR and do not come close to the effectiveness of models use crime data. It also shows that the associations between the variables and the SAR are, for the most part, statistically significant, but not economically significant. This implies that the association between socioeconomic factors and the successful arrest rate is really weak, and this is applied to effectively all measures of economic welfare used in this study. Models using crime data are significantly more effective and better predictors of the SAR of a community in a given year than models using socioeconomic variables as a measure of the SAR.

Conclusion

In conclusion, this study aimed to investigate the relationship between the successful arrest rate (SAR) in Chicago City and socioeconomic variables relating to all 77 communities over the 2002-2022 period. Using data from the Chicago Police Department, the US Census Bureau, and the Illinois Department of Public Health (IDPH), this study found statistically significant associations between the successful arrest rate and some socioeconomic variables, such as the birth rate, per capita income, the percent of households below poverty, among others. Interestingly, the study also found relatively large associations between the total number of reports, arrests, and the population to the SAR at a highly significant level, indicating that communities with higher numbers of arrests, lower numbers of reports and population are associated with having a higher SAR, which follows general intuition on the matter.

It is important to note that the effect sizes of the significant associations were small and may not have significant implications. Therefore, it is crucial to interpret these findings with caution and acknowledge the limitations of this study. For instance, this study was unable to examine any causal relationships between the observed variables, and other factors that were not included in this study could also influence the SAR.

Moreover, the socioeconomic measures used in the study were constant throughout the study-period, meaning that the data did not account for any change in the values of these socioeconomic variables for the entirety of the study. Future research should explore the complex relationship between socioeconomic variables and successful arrest rates in areas larger than Chicago, while also taking into account potential confounding variables that could affect the results. This is because larger cities, or even countries, would imply access to more socioeconomic variables over a longer period of time, making the study highly more reliable. Overall, this study adds to the little existing literature on the relationship between socioeconomic status and crime, and highlights the need for continued research in this area on a larger scale.

References

1. Bloomberg, T. E. |. (2023, March 2). *Analysis | in Chicago, high crime leads to a mayor's downfall*. The Washington Post. Retrieved April 2023, from https://www.washingtonpost.com/business/in-chicago-high-crime-leads-to-a-mayors-downfall/2023/03/02/410cad44-b8fa-11ed-b0df-8ca14de679ad_story.html
2. Semuels, A. (2021, July 30). *Chicago's awful divide*. The Atlantic. Retrieved April, 2023, from https://www.theatlantic.com/business/archive/2018/03/chicago-segregation-poverty/556649/
3. Allan, E. A., & Steffensmeier, D. J. (1989). Youth, underemployment, and property crime: Differential effects of job availability and job quality on juvenile and young adult arrest rates. *American Sociological Review*, *54*(1), 107. https://doi.org/10.2307/2095665
4. LaFree, G., & Drass, K. A. (1996). The effect of changes in intraracial income inequality and educational attainment on changes in arrest rates for African Americans and whites, 1957 to 1990. *American Sociological Review*, *61*(4), 614. https://doi.org/10.2307/2096396
5. Klingner, J. (2023, April 14). *Chicago 2022 arrest rates collapse to just 5% – wirepoints: Wirepoints*. WirePoints Illinois Financial News. Retrieved April 2023, from https://wirepoints.org/chicago-2022-arrest-rates-collapse-to-just-5-wirepoints/
6. Department, C. P. (2023, April 16). *Crimes - 2001 to present: City of Chicago: Data Portal*. Chicago Data Portal. Retrieved April 16, 2023, from https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2
7. Bureau, U. S. C. (2014, September 12). *Census data - selected socioeconomic indicators in Chicago, 2008 – 2012: City of chicago: Data Portal*. Chicago Data Portal. Retrieved April 2023, from

https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

8. Illinois Department of Public Health (IDPH) and U.S. Census Bureau. (2013, May 30). *Public health statistics - selected public health indicators by Chicago Community Area - Historical: City of Chicago: Data Portal*. Chicago Data Portal. Retrieved April 2023, from https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Selected-public-health-in/iqnk-2tcu
9. *Public health statistics - life expectancy by Community Area - historical: City of chicago: Data Portal*. Chicago Data Portal. (2014, June 16). Retrieved April 2023, from https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Life-Expectancy-By-Commun/qjr3-bm53