

First Names and Crime: Does Unpopularity Spell Trouble?

David E. Kalist and Daniel Y. Lee

Department of Economics
Shippensburg University
Shippensburg, PA 17257

(717) 477-1687; (717) 477-1556
dekali@ship.edu; dylee@ship.edu

To be presented at
the Canadian Economics Association's 42nd Annual Meetings
Saturday June 7, 2008
University of British Columbia, Vancouver

Abstract

We investigate the relationship between first name popularity and juvenile delinquency by comparing the names of juveniles in a state's population with the names of juveniles who received substantiated charges in that state's juvenile justice system. The distribution of first names in the state population is different from the names of juvenile delinquents. Our results show that unpopular names are positively correlated with juvenile delinquency for both blacks and whites. Furthermore, unpopular names are correlated with factors that increase the tendency towards juvenile delinquency, such as a disadvantaged home environment and residence in a county with low socioeconomic status.

Introduction

We examine whether a relationship exists between juvenile delinquency and first names by answering a basic question: Are juveniles with unpopular names more or less likely to become juvenile delinquents? (We measure the popularity of a name by its frequency in the population via an index.) Studies have documented that certain first name characteristics are associated with lower socioeconomic status; we extend this literature. For example, Aura and Hess (2004) report that first name characteristics are predictors of an individual's income and educational attainment, noting that popular names were positively correlated with these attributes but black-sounding names, which by definition are names infrequently given to whites and have a low popularity index, were negatively correlated. Mehrabian and Piercy (1993) find that persons with unconventionally spelled names are perceived by others to have undesirable characteristics in terms of popularity, morality, warmth, and success. Names convey information in which people use to judge an individual's masculinity, success, and trustworthiness (Mehrabian, 2001). Twenge and Manis (1998) discuss other studies that show a negative association between unpopular names and life outcomes (e.g., psychological problems, dropping out of school, and low achievement test scores).

Names may also provide a mechanism for discrimination. Bertrand and Mullainathan (2004) conducted a field experiment that sent out 5,000 resumes and found that job applicants with black-sounding names were less likely to receive interview callbacks, with white-sounding names receiving 50 percent more callbacks. However, none of the studies investigated the possibility of a correlation between names and crime. To the best of our knowledge, Figlio (2003) is the only research that studies the relationship between names and disruptive behavior.

He finds that, especially for blacks, boys with names commonly given to girls are more likely to be suspended from school.

A possible reason for a name-crime link is that people with unpopular names may be more likely to have grown up in a disadvantaged environment, such as having parents with low socioeconomic status. Lieberman and Bell (1992) find that the mother's education level affects naming patterns—a case in point is Allison, a name rarely given by mothers without a high school diploma, but frequently given by mothers with 17 or more years of schooling. In general, Lieberman and Bell (1992) find that the least educated women are less likely to give a top 20 name to their children but more likely to give a unique name. Aura and Hess (2004) find that unpopular names are associated with having a mother and father who obtained fewer formal years of schooling. Fryer and Levitt (2004), who studied the blackness of first names, conclude that people with black-sounding names are more likely to have parents with lower levels of education. Discrimination against people with unpopular names in labor markets may reduce their employment and income opportunities, making illegal activities an attractive option.

Discussing the looking-glass theory of self model of names and adjustment, Twenge and Manis (1998) write, "Having an unusual name...leads to unfavorable reactions in others, which then lead to unfavorable evaluations of the self." Figlio (2005) contends that teachers' perceptions of students are dependent upon the names of their students which, in turn, may affect student test scores. This feedback effect provides another possibility of how names and crime may be related.

The results of this study may shed light on audit studies that show applicants with certain name characteristics receive fewer callbacks (Cross et. al, 1990; Kenney and Wissoker, 1994; Bertrand and Mullainathan, 2004; Carpusor and Loges, 2006; King et. al, 2006). For example, if

people with unpopular names in the population are more likely to have criminal histories, employers, renters, and others may avoid transacting with these applicants—a statistical discrimination explanation. This study also has potential implications for identifying youths who may engage in disruptive behavior or relapse into criminal behavior.

In summary, first name characteristics may relate to crime through several mechanisms. First, as documented in the literature, parents of low socioeconomic status are more likely to give their children unpopular names, and a large theoretical literature links children living in low socioeconomic households with juvenile delinquency. Heimer (1997) reports that youths from low socioeconomic families “are more likely to engage in violent delinquency in the future.” Second, because some employers may discriminate against job applicants with unpopular names, juveniles with these names may have fewer employment opportunities and thus they may turn to crime. Finally, juveniles with unpopular names may be more prone to crime because they are treated differently by their peers, making it more difficult for them to form relationships. Juveniles with unpopular names may also act out because they have a conscious or unconscious dislike for their name.

Data

To test for a relationship between first names and juvenile delinquency, we use two datasets from a large state.¹ The first dataset provides the first names for all males born during the period 1987-1991. From this dataset, we construct a popularity-name index (PNI) for each name. The PNI is then matched and assigned to each respective name in the juvenile delinquent dataset. The PNI for the i^{th} name is calculated as

$$\frac{\text{Frequency of Name } i}{\text{Frequency of Most Popular Name}} \times 100.$$

The PNI for Michael is 100, the most frequently given name during the period. The PNI for David is 50, a name given half as frequently as Michael. The PNI is approximately 1 for names such as Alec, Ernest, Ivan, Kareem, Malcolm, Preston, and Tyrell, indicating that Michael is 100 times more popular than these names.² Overall, there are 15,012 different names in the data. We treat names that are spelled differently but sound the same as distinct names. Furthermore, we do not change names that could be perceived as misspelled since it is possible that the unique spellings are intentional (e.g., Adriaan, Kristofer, and Patric). It should be noted that this dataset does not include any other information, such as the city or county of birth, surname, family structure, or parental sociodemographics.

The second dataset consists of all persons in the state who were referred to a county's juvenile justice system for alleged delinquent offenses that were later substantiated during the period 1997-2005.³ For each delinquent, we have information on the juvenile's first name, last name, date of birth, sex, race, and family living arrangements. The data are for approximately half the state's counties for the period 1997-2005, with the population in these counties accounting for approximately 50 percent of the state's population in 2006. These data provide a representative mix of the state's urban and rural counties, and includes the state's largest metropolitan area. For the counties under analysis, the mix between urban and rural population is almost identical to that of the state population. Our selected counties are also geographically spread throughout the state. The reported regressions are based on relatively large samples (exceeding 10,000 observations) increasing the precision of the results. It should be noted that those born during the period 1987-1991 reached their teenage years during the period 1997-2005. For example, those born in 1987 are aged 10 at the end of 1997 and aged 18 by the end of 2005.

In other words, for most of the 1987-1991 cohort, the dataset covers the most formidable years of juvenile delinquency.

Summary statistics for the PNI are presented in Table 1. The mean PNI for the state population (i.e., the cohort born 1987-1991) is 26.31 but falls to 22.0 for the juvenile delinquents. A t test reveals the difference in means is statistically significant ($p < 0.01$). There is a considerable difference in the reported medians. About half of the names in the state population have a PNI greater than 20, while about half the names of juvenile delinquents have a PNI greater than 11. Therefore, compared to the state population of names, a larger proportion of juvenile delinquents have unpopular names.

Empirical Results

To more closely examine the link between names and crime, we estimate several regression models of the basic form:

$$\ln(N_i) = a + b \ln(PNI_i) + e_i$$

where N_i is the ratio of the number of juvenile delinquents with name i to the number of people in the state population with the same name. The PNI_i is the popularity-name index for name i and e is a random error term. Both variables are measured in natural logs and thus the coefficient on PNI gives the percentage change in the number of juveniles who are delinquents with name i from a 1 percent increase in the PNI (i.e., the elasticity of N_i with respect to PNI_i). The regression equation is also estimated separately for white and black juvenile delinquents. All standard errors are corrected for arbitrary heteroscedasticity of unknown form.

The results of the regression are presented in Table 2. The coefficient on the PNI is negative and precisely estimated in all cases, indicating that less popular names are associated

with a greater likelihood of juvenile delinquency. For example, a 10 percent increase in the popularity-name index decreases the number of juvenile delinquents by 3.67 percent. The R-squared statistic, ranging from 0.28 to 0.50, reveals that the PNI is an important factor in explaining variation in the dependent variable. In separate regressions for black and white juveniles, we find that PNI coefficient for blacks is -0.192 compared to -0.368 for whites. A Chow test confirms that these coefficients are statistically different ($p < 0.001$). If blacks are more likely to commit delinquent acts at higher rates than whites (perhaps attributable to living in socially disadvantaged areas) for any given level of the PNI, it can be mathematically shown that the elasticity for blacks will be smaller in absolute value.

Names and Family Structure

In Table 3, we estimate a series of logit regressions using the juvenile delinquent dataset in which the binary dependent variable is one of two measures of family structure: 1) whether the juvenile lives with only his mother, and 2) whether the juvenile lives with both parents. The key independent variable is the PNI. Each observation in the data represents a unique individual, accomplished by matching first name, last name and date of birth across the different years of the data.⁴ In light of Lieberman and Bell's (1992) findings that less popular names are associated with lower socioeconomic status, we suspect that the PNI may be correlated with certain family structures.

The results in Table 3 are presented separately for black and white juveniles. The coefficients represent marginal effects. We find that, regardless of race, the PNI is positively associated with the juvenile living with both parents but negatively related to living only with his mother. For blacks, the probability of living with both parents increases from approximately 10

to 16 percent as the PNI rises from 0 to 100, but the probability of living with only their mother decreases from 55 to 48 percent as the PNI rises from 0 to 100. Likewise for whites, the probability of living with both parents increases from approximately 30 to 36 percent as the PNI rises from 0 to 100, but the probability of living with only their mother decreases from 34 to 29 percent as the PNI rises from 0 to 100. To check the sensitivity of the results presented in Table 3, we re-estimated the logit regressions by including dummy variables for the juvenile's county of residence. The inclusion of the county dummy variables had virtually no effect on the coefficient estimates of the PNI and their statistical significance.

These results suggest that the popularity of a juvenile's name may not have a causal effect on crime. In other words, the correlation between names and crime may be the result of a third variable, the juvenile's family structure.⁵ As observed in Table 3, juvenile delinquents with unpopular names are more likely to live in households with less parental supervision. There are undoubtedly many other factors that could cause a correlation between first names and crime. These factors include a lack of parental resources (e.g., income, wealth, child care, and health care), an abusive home environment, the quality of the juvenile's school, and peer effects. Neighborhood effects of the juvenile's residence may also be associated with delinquency such as the presence of gangs, amount of drug trafficking, and availability of alcohol, drugs, and guns. The resources of the juvenile justice system and law enforcement agencies to counter crime are also important.

Names and Sociodemographics of Juvenile's County of Residence

We further examine the relationship between juvenile delinquency and first names by determining whether juvenile delinquents with unpopular names are more likely to live in

disadvantage counties, as measured by the county's unemployment rate, per capita Temporary Assistance to Needy Families (TANF), and per capita income. The regression models take the following form:

$$C_j = a + bPNI_j + e_j$$

where C_j is the county characteristics (unemployment rate, per capita TANF, and per capita income) of the j^{th} juvenile delinquent's residence. The dependent variables are measured using their respective mean values during the period 1997-2005. Both per capita TANF and per capita income are measured in constant 2005 dollars. The county-level unemployment data are from the Bureau of Labor Statistics (Local Area Unemployment Statistics). The data for TANF and income are from the Bureau of Economic Analysis (Regional Economic Information System). We estimate the regressions separately for white and black juveniles.

In all six regressions reported in Table 4, juvenile delinquents with more popular names tend to reside in counties that are less socioeconomically disadvantaged. We suspect that the same pattern holds true for all juveniles, not just juvenile delinquents, though we cannot test this hypothesis with our data. The possibility arises that the name-crime link is confounded by these neighborhood effects, implying the correlation between a name's popularity and crime results, at least in part, from the sociodemographic characteristics of the juvenile's residence.

A 50-point increase in the PNI is associated with almost a 0.10 percentage point reduction (50×-0.0018) in the county unemployment rate in which black juveniles reside; the percentage reduction is less than half for white juveniles. The county's per capita TANF decreases by about \$10.00 for black juveniles and \$2.00 for white juveniles from a 50-point increase in their PNI. Unlike the other dependent variables in Table 4, per capita income is

directly related to PNI. For example, per capita income is almost \$400 higher in counties in which the PNI for a black juvenile is 50 points higher.

Conclusion

We add to the literature on first names by finding, regardless of race, a positive correlation between unpopular first names and juvenile delinquency. The first names of juvenile delinquents do not represent a random sample of first names in the general population. A 10 percent increase in the popularity of a name is associated with a 3.7% decrease in the number of juvenile delinquents who have that name. Because unpopular names may signal an increased propensity to commit crime, this study provides additional insight (beyond that of a discrimination motive on part of employers) as to why job applicants with unpopular names may be disadvantaged.

We show that unpopular names are associated with juveniles who live in nontraditional households, such as female-headed households or households without two parents. In addition, juvenile delinquents with unpopular names are more likely to reside in counties with lower socioeconomic status. These two findings suggest that unpopular names may merely be correlated with omitted factors (disadvantage home environment) that affect the propensity towards juvenile delinquency rather than are the cause of juvenile delinquency. Nevertheless, if having an unpopular name constrains employment opportunities or negatively affects how others perceive one, it is possible that names could have a causal effect on crime. This hypothesis is consistent with the findings of Twenge and Manis (1998). They control for family background characteristics by using a paired-siblings design and report that “First names and identity appear to go hand in hand, with first names explaining a small but significant part of the variance in the

psychological adjustment of the individual.” With appropriate data, further research should explore more fully the issues of causality and the name-crime link. Any such future research should acknowledge, as in this study, a methodological limitation in using data on juveniles who are formally adjudicated. Previous research has shown that adjudicated delinquents are not the full population of all juvenile offenders, since official crime data do not contain self-reported crime. It is possible that juveniles whose crime is self-reported may have first name characteristics that differ from those who commit officially reported crime.

In research settings in which there are few control variables, especially on individual family background characteristics, first names could serve as a useful proxy and help address omitted variable bias. Gyimah-Brempong and Price (2006), for example, use the Scrabble score of a person’s first name as a tangential explanatory variable (their key independent variables measure skin hue) in regressions trying to explain age at incarceration and length of sentence.⁶ In the majority of their specifications, a higher Scrabble score is associated with either an increased hazard of criminal activity or a longer sentence. The authors conclude that, “...a dark skin hue, all things equal, induces a transition into criminal activity...as a result of dark skin hue constraining the set of legitimate non-criminal activities.” Based on the results presented herein, it is possible that the same could be said for people with unpopular names.

First name characteristics may have implications for other types of crime and law research. Are first names useful in predicting criminal recidivism? Do jurors use information on the defendant’s name to help decide guilt or punishment? In a related context, Figlio (2005) contends that teachers’ perceptions of students are dependent upon the names of their students which, in turn, may affect student test scores. It is possible that police officers may profile based on a person’s first name, causing officers to further interrogate and physically search people with

unpopular names. For example, police traffic stops may more frequently result in vehicle searches of drivers who have unique names. Depending on the information available to researchers, first name characteristics may be an important factor to help identify individuals at high risk of committing or recommitting crime, leading to more effective and targeted intervention programs.

References

- Aura, Saku and Gregory D. Hess, "What's in a Name?" CESifo Working Paper Series, No. 1190, 2004.
- Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, 94(4), 991-1013.
- Carpusor, Adrian G. and William E. Loges, "Rental Discrimination and Ethnicity in Names," *Journal of Applied Social Psychology*, 2006, 36(4), 934-952.
- Cross, Harry, Genevieve Kenney, Jane Mell and Wendy Zimmerman, *Employer Hiring Practices: Differential Treatment of Hispanic and Anglo Job Seekers*, Urban Institute Report 90-4, Washington DC: The Urban Institute, 1990.
- Figlio, David, "Names, Expectations, and the Black-White Test Score Gap," NBER Working Paper 11195, 2005.
- Figlio, David, "Boys Named Sue: Disruptive Children and Their Peers," 2003, Working Paper.
- Fryer, Roland and Steven Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, 2004, 767-805.
- Gyimah-Brempong, Kwabena and Gregory N. Price, "Crime and Punishment: And Skin Hue Too?" *American Economic Review*, 2006, 96(2), 246-250.
- Heimer, Karen, "Socioeconomic Status, Subcultural Definitions, and Violent Delinquency," *Social Forces*, 1997, 75(3), 799-833.
- Kenney, Genevieve M. and Douglas A. Wissoker, "An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers," *American Economic Review*, 1994, 84(3), 674-683.

King, Eden B., Juan M. Madera, Mikki R. Hebl and Jennifer L. Knight, "What's in a Name? A Multiracial Investigation of the Role of Occupational Stereotypes in Selection Decisions,"

Journal of Applied Social Psychology, 2006, 36(5), 1145-1159.

Lieberson, Stanley and Eleanor O Bell, "Children's First Names: An Empirical Study of Social Taste," *American Journal of Sociology*, 1992, 98(3), 511-554.

Mehrabian, Albert, "Characteristics Attributed to Individuals on the Basis of Their First Names,"

Genetic, Social, and General Psychology Monographs, 2001, 127(1), 59-88.

Mehrabian, Albert and Marlana Piercy, "Positive or Negative Connotations of Unconventionally and Conventionally Spelled Names," *Journal of Social Psychology*, 1993, 133(4), 445-451.

Twenge, Jean M. and Melvin Manis, "First-Name Desirability and Adjustment: Self-Satisfaction, Others' Ratings, and Family Background," *Journal of Applied Social Psychology*, 1998,

28(1), 41-51.

Table 1
 Summary Statistics of Popularity-Name Index

	State Population	Juvenile Delinquents
Mean (Std. Dev.)	26.31 (25.6)	22.00 (25.4)
Percentiles		
1%	0.01	0.01
5%	0.05	0.03
10%	0.33	0.13
25%	3.40	1.16
50%	20.41	11.30
75%	46.25	35.71
90%	54.64	54.14
95%	76.35	76.35
100%	100.00	100.00

Table 2
Effect of the Popularity-Name Index on Juvenile Delinquency:
Regression Results

	All Juveniles	Black Juveniles	White Juveniles
Constant	-2.076 (0.023)	-1.148 (0.024)	-2.313 (0.033)
ln(PNI)	-0.367 (0.006)	-0.192 (0.006)	-0.368 (0.01)
R-squared	0.4998	0.283	0.4927
obs.	3,777	2,613	1,585

Notes: Dependent variable is the natural log of the ratio of the number of juvenile delinquents with name i to the total number of individuals in the state population with name i . ln(PNI) is the natural log of the popularity-name index and its coefficient represents the elasticity of the number of juvenile delinquents with respect to the PNI. Robust standard errors are in parentheses.

Table 3
The Relationship between Family Structure and Popularity-Name Index:
Logit Regression Results

	<u>Black Juveniles</u>		<u>White Juveniles</u>	
	Live with Mother	Live with Both Parents	Live with Mother	Live with Both Parents
PNI	-0.0007 (0.0002)	0.0005 (0.0001)	-0.0005 (0.0001)	0.0006 (0.0001)

Notes: Marginal effects are shown and robust standard errors are in parentheses. PNI is the popularity-name index. For each regression, the number of observations exceeds 10,000.

Table 4
The Relationship between County Sociodemographic Characteristics and PNI:
Regression Results

	<u>Black Juveniles</u>			<u>White Juveniles</u>		
	Unemployment Rate	Per Capita TANF	Per Capita Income	Unemployment Rate	Per Capita TANF	Per Capita Income
Constant	6.111 (0.008)	166.47 (0.712)	31,022.2 (56.171)	5.401 (0.008)	52.807 (0.515)	31,463.5 (62.13)
PNI	-0.0018 (0.0003)	-0.2037 (0.0295)	7.609 (2.310)	-0.0007 (0.0002)	-0.0421 (0.0135)	10.019 (1.704)

Notes: Robust standard errors are in parentheses. The dependent variable measures the sociodemographic characteristics of the juvenile delinquent's county of residence. The unemployment rate is the county's average unemployment rate between 1997 and 2005. Per capita TANF (Temporary Assistance for Needy Families) is the county's average per capita payment (in 2005 dollars) received by recipients between 1997 and 2005. Per capita income is the county's average per capita income (in 2005 dollars) between 1997 and 2005. For each regression, the number of observations exceeds 10,000.

Endnotes

¹ Because of confidentiality concerns, we signed an agreement not to disclose the identity of the state.

² The following is a list of select names and their respective PNI: Matthew, 76; Christopher, 64; Ryan, 49; Brian, 30; Richard, 20; Charles, 16; Luke, 5; Walter, 2; and Garland, 0.06.

³ The data that we use are not limited to incarcerated juveniles. In most cases, substantiated charges lead only to warnings and counseling, probation, informal adjustment, fines, or consent decree.

⁴ Because of confidentiality, we cannot report the exact number of observations (since each observation represents a unique individual) but in each regression shown in Table 3 and 4, the number of observations exceeds 10,000.

⁵ Unfortunately, we cannot directly test this hypothesis since the dataset containing the names of all juveniles in the state does not include information on family structure.

⁶ Figlio (2005) uses a first name with a high Scrabble score (20 or more points) as an attribute indicating low socioeconomic status.