

Does the Crown Court Discriminate Against Muslim-Named Offenders?

A Novel Investigation Based on Text Mining Techniques



Jose Pina-Sánchez, Julian Roberts, and Dimitrios Sferopoulos

Background: The study of sentencing discrimination in England and Wales has relied mainly on aggregate analyses comparing disparities by ethnic group. Examples of this approach include the Lammy Review of 2017 and the Ministry of Justice biannual ‘Statistics on Race’ reports. These publications have shown that for certain offence categories, some minority groups receive more severe punishments than White offenders. However, these studies fail to take into account differences in the individual characteristics of the cases processed.

Exploring in any further detail the presence of sentencing discrimination has been seriously limited by the dearth of official sentencing data capturing individual level information on the ethnicity of the defendant. Normally, the options available to researchers include the use of aggregate data or the collection of primary data through methods such as court observation or file review, both of them expensive and time-consuming undertakings. However, there is a third way: defendants can be classified into certain minority groups using text mining techniques and the defendants’ names obtained from court records. Names perceived to be associated with particular ethnic identities may trigger a discriminatory response from actors in the criminal justice system. This study reports original analyses conducted on a sentencing database to determine whether a defendant with traditional Muslim names would attract a different sentence.

Methods: To circumvent the limitations in the England and Wales official sentence data we explored an alternative research approach: we accessed, scanned and analysed sentence transcripts uploaded online at www.thelawpages.com. We designed a ‘data-scraping’ algorithm in *Perl* and *Selenium* to download the entire archive of records from ‘The Law Pages’ available in August 2017. These records do not capture the ethnic background of the offenders, yet, we have been able to match the offenders’s names against a list of traditional Muslim names (at www.alquranic.com) and classify 7.6% of offenders in our sample as having a Muslim name. Further text mining of ‘The Law Pages’ records allowed us to code a total of 47 case-relevant variables. These include three key sentence outcomes (whether custody was imposed, the number of months of the imprisonment, and whether the custodial sentence was indeterminate), three defendant characteristics (gender, whether Muslim-named or not, and the presence of previous convictions), and 42 other case characteristics (the presence of co-defendants, multiple counts of the same offence, plea, whether mitigating factors were recorded, whether a public protection sentence was imposed, the existence of a victim personal statement, whether the victim sustained injuries, and whether the defendant had been remanded in detention pre-trial, plus 36 specific offence types). The possibility of differentiating between specific offence types as opposed to the broad offence categories employed in the Ministry reports and most of the literature from England and Wales is crucial in order to control for the seriousness of the offence.

A subsample 8,437 of violent and sexual offenders resulting in a sentence of imprisonment imposed in the Crown Court from 2007 to 2017 was analysed. The analytical strategy is based on the comparison of custodial sentence lengths for Muslim-named and all other offenders, which we undertake in two stages. The first stage involves the comparison of the average sentence length for each of the two groups and the application of a t-test to ascertain whether any observed differences are statistically significant. This is the type of analysis that we could perform to investigate discrimination against BAME based on the official data currently available to researchers in the UK. This approach, however, is only capable of establishing gross disparities. That is, disparities that could be due to discriminatory practices, conflated with legitimate disparities reflecting differences in the legally-relevant characteristics of the cases processed. To test for unwarranted disparities, in the second stage of our analysis we specify a statistical model controlling for all the case characteristics derived from the The Law Pages records. To account for the typical right-skewness observed in the distribution of custodial sentence length and for the presence of indeterminate and life imprisonment sentences we used a proportional hazards Cox model.

Results: The average custodial sentence length for the group of offenders carrying traditional Muslim names was 133.6 months, while that of the non-Muslim-named group was 122.2 months. Thus, Muslim-named offenders received on average sentences 9.3% longer than the other offenders, a difference that is statistically significant (Welch two-sample t-test, $p\text{-value} < 0.007$). However, this difference disappeared once we accounted for the type of offence and other key case characteristics in our Cox model (Table 1). Given the proportional hazards specification of our model, the regression coefficients obtained should be understood as the effect of each of the variables included on the hazard rate (i.e. representing the risk of prison terms to be terminated at a specific point in time). Hence, negative coefficients are associated with the imposition of longer sentences, while positive coefficients indicate a shorter sentence. The model presented in Table 1 was repeated including interaction effects involving Muslim-name and the terrorism offences available at ‘The Law Pages’ archive. None of those interaction terms were statistically significant. Thus we found no evidence of discrimination even when the offence involved terrorism.

Conclusion: Without access to official sentencing data including various defendant identities at the individual level, we drew upon the application of ‘data-scraping’ and text mining techniques to collect a large sample of cases stored in an online legal database. We were then able to capture specific type of offence and a series of other legally-relevant case characteristics. Our analyses found no evidence of discrimination against defendants with traditional Muslim names, whether the analysis included all offences or only terrorism-related crimes. These findings should not be taken as conclusive proof of a lack of discrimination in sentencing against Muslim offenders; there is clearly a need for further research; most notably, our sample over-represents serious crimes. The principal contribution of our study lies in highlighting how the use of simple descriptive statistics in this area of research can be misleading.

Variable	Coefficient	Std. error	P-value
<i>Response variable: custodial sentence length (in months)</i>			
<i>Offender characteristics</i>			
Defendant male	-0.336	0.052	< 0.001
Defendant Muslim-name	-0.019	0.052	0.720
<i>Case characteristics</i>			
Co-defendants	-0.191	0.031	< 0.001
Public protection sentence	-0.756	0.045	< 0.001
Guilty plea entered	0.611	0.029	< 0.001
Mitigating factors	0.349	0.043	< 0.001
On remand	-0.537	0.033	< 0.001
Sentenced for a different second offence	-0.656	0.035	< 0.001
Sentenced for a different third offence	-0.422	0.049	< 0.001
Sentenced for a different fourth offence	-0.556	0.078	< 0.001
Sentenced for more than four different offences	0.136	0.103	0.190
Multiple counts of the same offence	-0.094	0.009	< 0.001
Victim impact statement	-0.253	0.032	< 0.001
Victim sustained injuries	0.113	0.038	0.003
<i>Principal offence (reference category: murder)</i>			
Assault occasioning actual bodily harm	4.581	0.088	< 0.001
Inflicting grievous bodily harm	4.99	0.106	< 0.001
Conspiracy to commit grievous bodily harm	3.311	0.198	< 0.001
Causing grievous bodily harm with intent	3.207	0.083	< 0.001
Causing an affray	5.656	0.121	< 0.001
Violent disorder	5.24	0.113	< 0.001
Unlawful (malicious) wounding	5.031	0.124	< 0.001
Common Assault / Assault by beating	5.124	0.192	< 0.001
Attempted murder	1.821	0.107	< 0.001
Manslaughter	3.328	0.085	< 0.001
Arson	4.461	0.198	< 0.001
Arson reckless / with intent to endanger life	3.754	0.135	< 0.001
Conspiracy to commit arson	3.526	0.347	< 0.001
Robbery	3.505	0.091	< 0.001
Attempted robbery	3.909	0.140	< 0.001
Conspiracy to commit robbery	3.065	0.115	< 0.001
Rape	3.017	0.088	< 0.001
Indecent assault	4.058	0.115	< 0.001
Indecent assault of a child	3.752	0.188	< 0.001
Sexual activity with a child	3.987	0.103	< 0.001
Sexual assault	4.432	0.115	< 0.001
Rape of a child	2.802	0.127	< 0.001
Attempted rape	3.661	0.238	< 0.001
Attempted rape of a child	3.383	0.292	< 0.001
Assault by penetration	3.817	0.238	< 0.001
Dangerous driving	5.343	0.118	< 0.001
Causing death by careless driving	5.35	0.150	< 0.001
Causing death by dangerous driving	3.563	0.098	< 0.001
Causing death by careless driving while over the alcohol limit	3.512	0.182	< 0.001
Causing serious injury by dangerous driving	4.238	0.205	< 0.001
Kidnap	3.569	0.137	< 0.001
Conspiracy to kidnap	3.573	0.275	< 0.001
Restraint of persons freedom of movement	3.684	0.118	< 0.001
Collecting a record of information to commit acts of terrorism	4.701	0.175	< 0.001
Engaging in conduct in preparation for acts of terrorism	3.338	0.162	< 0.001
<i>Random effects</i>			
Standard deviation random intercept	0.183		
<i>Sample size</i>			
Level 1 – Case: $N = 8437$			
Level 2 – Court: $N = 86$			

Table 1: Results from the Cox Model on Sentence Length