



UNIVERSITY OF LEEDS

This is a repository copy of *Have The England and Wales Guidelines Affected Sentencing Severity? An Empirical Analysis Using a Scale of Severity and Time-Series Analyses*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/140337/>

Version: Accepted Version

Article:

Pina-Sánchez, J, Gosling, JP orcid.org/0000-0002-4072-3022, Chung, H-I et al. (3 more authors) (2019) *Have The England and Wales Guidelines Affected Sentencing Severity? An Empirical Analysis Using a Scale of Severity and Time-Series Analyses*. *British Journal of Criminology*, 59 (4). pp. 979-1001. ISSN 0007-0955

<https://doi.org/10.1093/bjc/azz005>

© 2019, The Author(s). Published by Oxford University Press on behalf of the Centre for Crime and Justice Studies (ISTD). All rights reserved. This is an author produced version of a paper published in *The British Journal of Criminology*. Uploaded in accordance with the publisher's self-archiving policy. For permissions, please e-mail: journals.permissions@oup.com This article is published and distributed under the terms of the Oxford University Press, Standard Journals Publication Model (https://academic.oup.com/journals/pages/open_access/funder_policies/chorus/standard_publication_model).

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Have the England and Wales Guidelines Influenced Sentencing Severity? An Empirical Analysis Using a Scale of Severity and Time-Series Analyses

Jose Pina-Sánchez*

John Paul Gosling

Hye-In Chung

Elizabeth Bourgeois

Sara Geneletti

Ian Marder

University of Leeds, School of Law, The Liberty Building, Leeds, LS2 9JT, UK

*Corresponding author details. Telephone number: 07909414923, email:

j.pinasanchez@leeds.ac.uk

Abstract

Sentence severity has increased in England and Wales in recent years. The causes of the increase remain unclear. One possible explanation relates to the introduction of sentencing guidelines, which seem to coincide in time with the increase in sentence severity. To date, investigations of this hypothesis have been limited to simple exploratory analyses or to specific offences. We use a new scale of sentence severity - developed using Thurstone scaling and the participation of 21 magistrates - and time-series modelling to explore whether a causal effect can be attributed to seven different guidelines. We corroborate the existence of an increase in sentence severity, however, we do not find conclusive evidence pointing at the guidelines having caused it.

Keywords: Sentencing, severity, guidelines, time-series, Thurstone

Funding Information:

This work was supported by the National Centre for Research Methods [RIS14241/06].

Acknowledgements:

We are grateful to Jill Gramann, Robin Linacre, Amber Isaac, Emma Marshall, Jenna Downs, Marco Doretti, Anthea Hucklesby, Julian Roberts, Sebastian Walker, David Hayes, Ruth Pope, Eleanor Nicholl, Julian Berg, and to the 21 magistrates and seventeen members of the sentencing and punishment working group of the European Society of Criminology that participated in our survey.

1. Introduction

In recent decades, England and Wales has experienced a substantial inflation in sentence severity. The range of different disposals available to courts complicates attempts to quantify the exact change. Nevertheless, separate analyses for different disposal types provide evidence of an unequivocal increase in severity. For example, according to the Ministry of Justice (MoJ) quarterly statistics¹, from 1999 to 2017 the use of suspended sentences rose from 0.2% to 4.4%, the use of community orders dropped from 10.8% to 7.9% over that same period, and the average custodial sentence length increased by 47%.

The prison population is clearly affected by increases in the custody rate or the average sentence length and therefore constitutes another measure of the increase in severity. Analysis of the prison population trends corroborate the severity uplift seen in the disposal trends. Between 1993 and 2016 the total prison population rose by 92.4% (MoJ, 2016). This may of course reflect an increase in the volume of offending. However, statistics reveal that crime rates and sentencing caseloads declined over the same period (see Crown Prosecution Service; Office for National Statistics)². Taken together, these trends suggest sentencing has become harsher independent of the volume of crime recorded and cases processed. The key question is then, what is behind this increase in severity?

A number of explanations have been advanced. Penal populism is one obvious candidate. Scholars have suggested that courts and Parliament have simply reflected growing public intolerance with offending, particularly serious and sexual offending (Pratt, 2007). The extension of mandatory minimum³ and life sentences for an ever-wider range of offences (Ashworth, 2015) exemplifies the harsher criminal justice legislation that could be associated to the concept of penal populism.

More recently, it has been suggested that sentencing guidelines have contributed to this trend towards greater severity (Allen, 2016; Padfield, 2016). Sentencing guidelines have been evolving for well over a decade now (Roberts and Ashworth, 2016; Ashworth and Roberts, 2013). At present, most major offences are covered by definitive guidelines (Wasik, 2014). To be precise, fifteen offence specific guidelines are in operation in England and Wales, structuring the sentencing process of more than 149 offence types⁴, and by 2020 the Sentencing Council intends to have guidelines covering all of the most common offences.

Allen (2016) suggests that the increase in sentence severity might be a consequence of the guidelines' emphasis on consistency and proportionality. If consistency is achieved by more upward adjustments to sentences rather than downwards, increasing sentence severity would be a natural outcome (Sentencing Council, 2011a). This hypothesis is supported by some of the empirical evidence from the US, where some studies have detected increases in sentence severity following the introduction of guidelines (D'Alessio and Stolzenberg, 1995; Griset, 1999). Similarly, the central part played by the principle of proportionality in the design of the guidelines (Cooper, 2013; Hutton, 2013) has raised concerns regarding the potential relegation of personal mitigating circumstances in determining the sentence outcome (Cooper, 2013).

More specifically, some of the guidelines have intentionally shifted tariffs for specific offence types that were deemed either too lenient or too harsh. An example of the latter is the reduction in the

¹ Available at: <https://www.gov.uk/government/collections/criminal-justice-statistics-quarterly>

² <https://www.cps.gov.uk/underlying-data/cps-magistrates-court-caseload-2014-2017>
<https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingdecember2017>

³ A wide range of provisions based on an overall tougher approach is well reflected in the Legal Aid Sentencing and Punishment of Offenders (LASPO) Act 2012. <https://www.gov.uk/government/news/new-sentences-and-criminal-offences-come-into-effect>

⁴ Excluding the Magistrates' Courts Sentencing Guidelines.

starting point for ‘drug mule’ offences following considerations about the vulnerable nature of offenders of this type of crime (Sentencing Council, 2011b). Examples of changes in the opposite direction include those of assault offences, such as causing grievous bodily harm with intent, rape and possession of a bladed article, for which tariffs have been increased (Sentencing Council, 2015, 2018a). Similarly, the new guidelines for magistrates’ courts (Sentencing Council 2017a) suggest increasing fine levels for the most serious cases of speeding, whilst the new guidelines for domestic abuse seek to “*ensure that the seriousness of these offences is properly taken into account*” (Sentencing Council, 2018b).

In short, there are well-founded reasons to believe that the connection hypothesised between the introduction of sentencing guidelines and the increase in severity is not spurious. The temporal coincidence of the increase in sentence severity with the publication of sentencing guidelines raises legitimate questions about whether the latter is responsible for the former.

Besides the evaluations undertaken by the Sentencing Council, the empirical literature exploring the England and Wales guidelines experience is scarce, and it has provided mixed results regarding the impact of the guidelines on sentence severity. XXXX (2016) studied the consequences of the 2011 summer riots which coincided with the publication of the guidelines for cases of commercial burglary. The authors found that severity increased together with disparities in sentencing. Fleetwood, Radcliffe and Stevens (2015) found that the new drugs guidelines achieved its original aim to reduce the severity with which offences of ‘importing a Class A drug’ are treated.

The scope of these two studies is limited since they both focus on a specific type of offence. In addition, the methodology used to assess changes in severity are simple before and after comparisons. This is particularly problematic given the possibility that the alleged changes in severity might be due to ongoing trends in sentence severity that were not caused by the guidelines. In this article, we aim to provide new and more robust evidence on the hypothetical effect that the sentencing guidelines might have had on sentence severity. To do so, we follow the approach taken by the Sentencing Council in their evaluation of some of their guidelines (Sentencing Council, 2015, 2017b, 2018c, 2018d), where it is shown that the guidelines might have been responsible for a statistically significant increase in severity in specific offences of causing grievous bodily harm with intent, actual bodily harm, and commercial burglary. Our approach is based on time-series analyses and a new scale of sentence severity that encompasses all disposal types.

Time-series analyses have been used previously in the literature to assess the effect of different sentencing policy reforms in the US (Chen, 2008; D’Alessio and Stolzenberg, 1995; Marvell, 1995; Marvell and Moody, 1996; Nicholson-Crotty, 2004; Stolzenberg and D’Alessio, 1994, 1997; Vujić, Commandeur and Koopman, 2016). The use of time-series analysis, as opposed to simpler before and after comparisons, is key to establishing whether any observed changes following the implementation of sentencing guidelines can be attributed to them, rather than an ongoing prior trend. In so doing, this research design offers more robust evidence regarding the potential causal effect of the guidelines.

The development of a scale of severity is another crucial contribution of this article. With the exception of the recent evaluations carried out by the Sentencing Council, the literature assessing the impact of sentencing policy reforms - in the UK and elsewhere - are based on the analysis of specific disposal types, mainly the relative use of custodial sentences and their average length. This is a convenient approach, but it has some important methodological implications. Research based on changes in custodial sentence length ignores non-custodial sentences, and as such, is severely affected by selection bias (Bushway, Johnson and Slocum, 2007; Sullivan, McGloin and Piquero, 2008; Zatz and Hagan, 1985); e.g. in England and Wales only 7.2% of cases are sentenced to custody (MoJ, 2018). On the other hand, analyses of the relative use of custodial sentences are particularly inefficient, since this approach reduces the heterogeneity of the different disposal types available (e.g. conditional discharge, fine, community order, etc.) to a simple zero/one variable.

Using a scale of sentence severity, we can aggregate the different sentence types based on their relative severity and in so doing study all of the cases processed. This allows us to eliminate the problem of selection bias, while accounting for changes in the use of disposals beyond the ‘custody or not’ dichotomy.

Lastly, this article complements the evidence produced by the Sentencing Council evaluations in two important ways. First, we create a more robust, statistically principled, and transparent scale of sentence severity. Second, we undertake analyses of the major guidelines designed by the Sentencing Council up to 2016. Based on these two important considerations, we provide new and more comprehensive evidence regarding the effect of the new England and Wales sentencing guidelines on sentence severity. We proceed by presenting the strategy we followed to estimate a scale of sentence severity. This is followed by a short description of the data, and by the analyses carried out. The implications of our findings are discussed in the last section.

2. Measuring Sentence Severity

Tremblay (1988) coined the term ‘penal metric theory’ to refer to studies investigating the measurement of the relative severity of different disposal types available to sentencers. Besides being a relevant academic topic in its own right, the measurement of sentence severity can be used as a tool to explore other sentencing related research questions. For example, the empirical investigation of the concept of proportionality in sentencing requires that sentences using different units of measurement (e.g. pounds for fines, hours for community orders, or days for custodial sentences) are combined into a single scale of severity (Hebenton and Pease, 1995; Leclerc and Tremblay, 2016; Lovegrove, 2001). Perhaps more importantly, such a scale can be used to eliminate the problem of selection bias affecting current studies on deterrence, discrimination and other disparities in sentencing (Buchner, 1979; Erickson and Gibbs, 1979; Sebba, 1980; Sebba and Nathan, 1984). However, in spite of the value of such academic endeavour, we are only aware of fifteen studies where a scale of sentence severity has been estimated, most of them carried out in North America and dated (Apospori and Alpert, 1993; Buchner, 1979; Croyle, 1983; Erickson and Gibbs, 1979; Harlow, Darley and Robinson, 1995; Leclerc and Tremblay, 2016; McClelland and Alpert 1985; McDavid and Stupack, 1981; Schiff, 1997; Sebba, 1980; Sebba and Nathan, 1984; Spelman, 1995; Tremblay, 1988; Van Kesteren, 2009; Warr, Meier and Erickson, 1983).

As a result of the different disposal types used across jurisdictions, and how these keep evolving to explore alternative forms of non-custodial punishments (Bottoms, Rex and Robinson, 2004), the use of a scale of severity as a research tool needs to fit specific place and time requirements. Given the inadequacy of the existing literature, the England and Wales Sentencing Council decided to design its own scale of severity, which the Council’s researchers have used to evaluate their sentencing guidelines (Sentencing Council, 2015, 2017b, 2018c, and 2018d). The Council’s scale of severity is based on the starting points coded in the sentencing guidelines for specific types of offences. Each starting point reflects a category of seriousness to which a specific sentence is attached. For example, offences of ‘grievous bodily harm with intent’ are divided in three categories of seriousness, with starting points of four, six, and twelve years’ custody attached to them (Sentencing Council, 2011c).

The estimation process of the Council’s scale of severity can be summarised in four steps: a) using only offences with custodial sentences as their starting points, a non-linear function is estimated representing the relative change in sentence length associated with the increase in one category of seriousness; b) a value of severity is chosen for a one-month custodial sentence to anchor the relationship between sentence length and severity; c) values of severity for custodial sentences longer than a month are estimated using the function obtained in step-a; and d) values of severity

for non-custodial outcomes (fines and community orders) are extrapolated beyond the custodial range of the same function obtained in step-a.

Although original in its design, this scale has three important flaws. First, it assumes that the change in seriousness across different offences is equivalent. For example, a change in seriousness from category-1 to category-2 for a case of common assault is equivalent - in relative terms - to the same change of seriousness in a case of grievous bodily harm. This assumption is difficult to justify since the starting points attached to each category of seriousness were designed on an offence-by-offence basis, rather than applying a common rule across them (e.g. custodial sentence length to be increased by 75% with every additional level of seriousness). Second, the severity scores for non-custodial sentences are extrapolated from the relationship estimated in step-a, which is estimated using custodial sentences exclusively. This is equivalent to establishing a unique function of severity across custodial and non-custodial outcomes. This assumption is refuted by much of the penal metric theory (Erickson and Gibbs 1979; Leclerc and Tremblay, 2016; Tremblay, 1988), which points at varying marginal rates of severity for different disposal types. Third, the choice of the value for a one-month custodial sentence has important implications since it will determine the severity scores of all the other sentence outcomes. However, the rationale behind that choice is not well justified.

Our Approach: General Considerations

Designing a scale of severity entirely based on the sentencing guidelines can be affected by serious limitations. However, it would be unwise to disregard the guidelines as an important source of information on sentence severity. Our approach seeks to combine information encoded in the guidelines and enhance it with qualitative insights from sentencers. Specifically, we will borrow the sentencing ladder applied across all of the sentencing guidelines, which specifies the ordinal ranking of the different disposal types available to sentencers: *absolute discharge* < *conditional discharge* < *fine* < *community order* < *suspended sentence* < *immediate custody*. Sentencers' insights are used to identify and quantify those transitions across the sentencing ladder where there might be some overlap in severity.

By exploiting overlaps between disposal types to inform their relative severity, we accept the existence of *penalty exchangeability* (Erickson and Gibbs 1979; Harlow, Darley and Robinson, 1995; McClelland and Alpert 1985; Sebba and Nathan 1984; Leclerc and Tremblay, 2016; Lovegrove, 2001; Petersilia and Deschesnes 1994a, 1994b; Spelman 1995; Tremblay, 1988). That is, different disposal types can reach similar levels of severity and therefore become exchangeable when the intensity of the - in principle - less severe disposal type is high (e.g. a community order with multiple and onerous conditions attached), while the intensity of the more severe disposal type is low (e.g. a short suspended sentence with no conditions attached).

A sample of 21 magistrates was selected using snowball sampling⁵, and asked to respond to a short online questionnaire. Other studies in the literature have also used samples of offenders, probation officers, police officers, or members of the general public, showing substantial variations in the perceived severity of sentences (Erickson and Gibbs, 1979; Sebba and Nathan, 1984; Spelman, 1995). We chose to collect views from sentencers since we are exploring changes in sentencing practice. As such, it is of paramount importance that the scale of severity reflects the views on the topic held by sentencers.

Our Approach: Specific Considerations

⁵ The snowball sampling strategy was started using two magistrates from two different courts, both located in England (in the north and south of the country). These magistrates then circulated the questionnaire amongst their contacts from their own and nearby magistrates' courts. The specific courts where the 21 responses were gathered cannot be disclosed to protect respondents' anonymity.

The first method considered to elucidate the extent of the overlap between disposal types was *magnitude escalation*. This is the most commonly employed method in the penal metric theory (Apospori and Alpert, 1993; Erickson and Gibbs, 1979; Harlow, Darley and Robinson, 1995; Leclerc and Tremblay, 2016; McClelland and Alpert, 1985; Spelman, 1995; Tremblay, 1988; Warr, Meier and Erickson, 1983), possibly because of its simplicity. Subjects are given a list of sentence outcomes and asked to compare the relative severity of those outcomes against a common benchmark or ‘modulus’, often a one-year custodial sentence. Following piloting trials we decided to discard this method. Specifically, we noted an unexpectedly high variability in the responses obtained. This is a feature that can be observed in other studies where the method has been implemented (Erickson and Gibbs, 1979; Tremblay, 1988), and points at the possibility that interviewees follow different thought process while considering their responses.

Given the observed unreliability of responses under magnitude escalation, it was decided to opt for *Thurstone’s scaling method*. This approach has been previously used in the penal metric literature (Buchner, 1979), and is considered to be less cognitively demanding on the interviewees (Spelman, 1995). As developed by Thurstone (1927), subjects are presented with a series of pairs of choices (in our case, sentence outcomes). For each choice, they are asked to identify what option they perceive to be more intense - severe in our case. The number of times a sentence outcome is judged more severe than another, can then be used to rank their relative severity.

We utilise the *Thurstone model (Case V)* that estimates a severity scale based upon pairwise comparison data. Specifically, the data used in the model takes the form of the proportion of times a sentence type is more severe than another looking across all pairwise comparisons. The statistical model that underpins the method associates a normal distribution with same variance and changing mean for each of the sentence types (Morsteller, 1951). Each of the means in these normal distributions can be thought of as severity scores. The amount of overlap between the distributions dictates their closeness on the severity scale. For instance, if two sentence types had the same normal distribution, then the chances that one severity for one of the types exceeded the severity for the other would be 50%, which would correspond to the situation where we cannot judge which sentence would be more severe on average. In contrast, if the two normal distributions were far apart, then we may be certain that the severity for one type exceeds that for another, which corresponds to situations where we are certain one sentence type is more severe than another.

To estimate the means of the underlying normal distributions and, hence, the severity scores, a least squares approach is utilised where the reported proportions are compared with the probabilities of one sentence being greater than another conditional on different sets of means. The set of normal distributions that can best replicate all of the judged performances simultaneously will be chosen as the basis of our severity scale. This model has been implemented using the ‘thurstone’ command within the R statistical package ‘psych’ (Revelle, 2018).

The sentence outcomes included in our questionnaire reflect the information available in the datasets published by the MoJ⁶. Specifically, we did not differentiate between fines based on their band, or between other non-custodial sentences based on the conditions attached, because that information is not available in the MoJ data. Under such criteria, eleven sentence outcomes were included in the questionnaire⁷: absolute discharge, conditional discharge, fine, community order, four types of suspended sentences (one-month custody suspended for six months, one-month custody suspended for twelve months, six-month custody suspended for six months, and twelve-

⁶ <https://www.gov.uk/government/organisations/ministry-of-justice/about/statistics>

⁷ The full questionnaire is available here: <https://admin.onlinesurveys.ac.uk/account/leeds/preview/scale-of-sentence-severity-open?referer=dashboard>

month custody suspended for 24 months), and three immediate custodial sentences of one, two, and three-month's duration.

Yet, during the piloting of this questionnaire, another problem was identified. When asking which of the two options is more severe, respondents systematically chose the disposal type that sits higher in the sentencing ladder. As a result, we were not obtaining the necessary information regarding the extent of the overlap in severity between disposal types. To circumvent this problem, we decided to ask how often could *sentence A* have a more punitive effect than *sentence B*. To ensure that answers are as encompassing as possible, we also asked magistrates to consider all possible offender backgrounds (e.g. from first-time offenders to career criminals) and all possible variations of the sentence outcome to be compared (e.g. from low to high community orders).

This change in the wording of the questionnaire makes it more cognitively demanding than requesting participants to simply pick the more severe option, since now they also have to assess the intensity of their choice (i.e. identify how often their choice can be seen as more severe). To facilitate the task, multiple-choice answers were provided using plain English with a percentage attached to them (e.g. "*a*) - Always has a more punitive effect (100% of times)", "*b*) Almost always (90% of times)", etc.). Furthermore, out of the 55 possible combinations (based on the permutation of eleven items in pairs with no repeats), we decided to restrict comparisons to just seven pairs to reduce the duration of the questionnaire and simplify the study. This was achieved by removing comparisons where it seems clear which of the two sentences will be more punitive.

Our Approach: Results

Responses to the questionnaire are summarised in Table 1. Each cell represents the relative frequency with which the sentence outcome at the top of the column in which the cell is located is considered to be more severe than the sentence outcome to the left of its row. Cells showing 1s, reflect comparisons where one of the sentences is always considered more severe, the remaining figures represent the average proportions for each pair-comparison obtained from our questionnaire.

Severity scores for the sentence outcomes included in Table 1 are estimated by feeding this matrix to the Thurstone model (Case V) described above. Following the approach taken in Buchner (1979), severity scores for other suspended and immediate custodial sentences that are not included in the questionnaire, are estimated at a second stage through the specification of additional regression models. Custodial sentences longer than three months are estimated through extrapolations from the following simple linear regression model,

$$Y_i = \alpha + \beta X_i + \epsilon_i \quad (1)$$

Where, the number of months is $X_i = (1,2,3)$, Y_i is composed of the severity scores estimated by the Thurstone model for one-, two- and three-month custodial sentences, α represents the constant term, and ϵ_i the model residuals. The model for suspended sentences is a simple extension of Equation 1; two slopes are used to account for differences in the duration of the custodial sentence and the length of the suspension, respectively. The severity scores obtained through this two-stage process for a variety of sentence outcomes are shown in Table 2. For comparability's sake, the severity scores for the same sentence outcomes obtained using the Sentencing Council scale are also included.

(Table 1 here)

(Table 2 here)

A simple comparison between the two scales shows how our approach has superior face validity than the scale currently in operation by the Sentencing Council. We are able to distinguish between absolute and conditional discharge, and to discriminate between suspended sentences based on

their duration. Perhaps more importantly, we can see how the relative severity allocated by the Council's scale to non-custodial outcomes (like fines or community orders) is excessive. For example, it seems highly questionable to consider a five-year custodial sentence little more than twice as severe as a community order.

Given the relatively small sample of magistrates to which we have had access, an additional validity check in the form of a split sample sensitivity analysis was carried out. The scale of severity was re-estimated based on two different samples, each composed of ten magistrates from our sample. The severity scores obtained for the sentence outcomes shown in Table 2 were remarkably similar for each of the two samples, suggesting that our approach is not affected by sampling error⁸. This high level of stability is a consequence of the strong restrictions that we imposed in deciding the comparison of sentence outcomes to be included in our questionnaire. Knowing the more severe item for most of those comparisons (e.g. a two-month immediate custody will always be more severe than a one-month immediate custody), and assuming that knowledge for many others (e.g. immediate custodial sentences deemed to be always more severe than a fine), we were able to limit the input required from the sentencers.

3. Data

For the analyses presented in the next section, we have compiled a dataset from different MoJ quarterly statistics⁹. This dataset represents a census of all principal¹⁰ offences processed in England and Wales in the Crown Court and magistrates' courts. Employing data aggregated by quarter is key to both assessing the immediate impact that followed the introduction of the guidelines, and examining fluctuations in sentence severity within a year of the guidelines' introduction. The data can also be disaggregated for indictable offences by offence group, which can be used to assess the specific effect of the most important sentencing guidelines.

In most instances (such as for sex, robbery, fraud and drug offences), there is a high level of overlap between the specific offences covered in each guideline and the offence groups as defined in the MoJ data. There are two important instances where that overlap is far from perfect. To study the impact of the assault guidelines, we looked at offences of violence against the person, which include 'non-assault' offences such murder or manslaughter. Similarly, we used the offence group of theft offences to assess the impact of the burglary and theft guidelines since the MoJ data available does not distinguish between these two types of offences.

⁸ To assess the stability of our scale of severity if a different population of sentencing experts was to be used, we replicated our questionnaire with a sample of seventeen UK based academics, all members of the sentencing and punishment working group from the European Society of Criminology. The sampling frame was composed of a total of 43 UK-based academic members of this working group, which gives us a response rate of 39.5%. As for comparisons based on a split sample analysis, we found that the scale of severity derived from our sample of academics was remarkably similar to the scale derived from magistrates. This offers further reassurance about the stability of our method and suggests that additional sensitivity analyses should focus on other assumptions invoked in our methodology. To assess the robustness of the assumption that scores of severity are distributed as latent normal random variables invoked in the Thurstone model, we estimate a new severity scale using the Bradley-Terry model (Bradley and Terry, 1952) using the R package 'BradleyTerryScalable' (Kaye and Firth, 2017). The Bradley-Terry model is built upon an assumption of latent exponential random variables. This change from a symmetric distribution to a heavily skewed distribution has resulted in no noticeable difference. As before, we find that the two scales are quite similar, which gives us confidence that the method is robust to these distributional changes.

⁹ Available at: <https://www.gov.uk/government/collections/criminal-justice-statistics-quarterly>

¹⁰ In cases where multiple offences are considered, the principal offence represents the most serious of those offences.

There are other datasets released by the MoJ¹¹ that provide information by the specific type of offence, as opposed to broad groups of offences, which would facilitate the generation of samples of offences more accurately matched with the offences covered in each of the guidelines. However, this data is grouped by year, which makes it difficult to examine the trends of offence severity across time and to detect the specific effect of the guidelines. Using the quarterly data, we ensure that we have at least fourteen time-points before the introduction of the assault guidelines – the first of the new guidelines to come into force. Table 3 shows the offence group used to study each of the guidelines, the date when each guideline was published and the date when it came into force¹².

Other offence specific guidelines published over the last couple of years, such as the terrorism guidelines, were not included in the analysis given the limitations of our data, which only reaches the third quarter of 2017. In addition, there are other guidelines published for other minor (or less frequent) offences that cannot be analysed reliably using the broad group of offences provided in the MoJ data. The MoJ data that we used is also limited since it does not offer the specific duration of suspended sentences. We have taken a mean imputation approach and have given all suspended sentences a severity score of 3.88, the mean of the four suspended outcomes included in our scale. Other datasets from the Sentencing Council do capture these two variables but only cover the 2011-2015 period, an unduly narrow window of observation for the time-series analyses that we present next. Despite these limitations, our study covers all of the new sentencing guidelines published for the most common offences and all the main disposal types available to sentencers.

(Table 3 here)

4. Analysis

A simple exploration of the MoJ data for all offences provides evidence of an increase in sentence severity in England and Wales over the last couple of decades. It is difficult to observe any changes in the use of different disposal types (Figure 1.A) given the prevalence of fines. However, if we focus on changes in the average sentence length across time (Figure 1.C) we can clearly observe a steady upward trend from 1999. Roberts and Irwin-Rogers (2015) identified this same trend using data up to 2013. Here, we demonstrate that this phenomenon has not slowed down in the last four years. This increase in the average sentence length is coupled with a decrease in the use of community sentences. Specifically, for indictable offences, the use of community sentences went from a peak of 36.6% in 2005 (Figure 1.B) to 20.1% in 2017, while the use of suspended sentences went from 0.7% in 2004 to 16.9% in 2017.

Yet, these plots alone cannot provide an integrated view on the overall trend in sentence severity. Average sentence length is perhaps the more intuitive of the plotted outcomes, but custodial sentences only represent a small proportion of the sentences imposed. Similarly, looking at trends across other disposal types we can see which ones are becoming more or less prevalent. However, we cannot observe the specific trend in the overall level of severity. To do so, we need to estimate the relative severity of different disposal types, so we can transform the main sentence outcomes into a single scale of severity. Using the average scores of severity for the different disposal types estimated in Section 2 and the data presented in Figure 1, we can create a combined severity score

¹¹ See 'Data behind interactive data tools (CSV)' at <https://www.gov.uk/government/statistics/criminal-justice-system-statistics-quarterly-december-2017>

¹² The Sentencing Council deliberately spaces the date of publication of new guidelines from the date they come into force (i.e. the date sentencers are required to start using them) so sentencers can familiarise themselves with their content.

and monitor its behaviour across time. This is shown in Figure 2, which differentiates between all offences and the subgroup of indictable offences.

(Figure 1 here)

(Figure 2 here)

The trend of increased severity deduced before is much clearer now as presented in Figure 2. For the period analysed we can observe an 8.8% increase in sentence severity for all offences, and a shocking 34.7% for indictable offences. If we focus on when exactly those changes took place, we can identify most of the increase in severity for all offences taking place roughly from 2004 to 2011, which suggests a potential effect of legislation that preceded the foundation of the Sentencing Council. In particular, given the specific starting point of this trend, a connection with the 2003 Criminal Justice Act seems likely. On the other hand, if we shift the focus to indictable offences, we can see how the increase in severity accelerated in 2011, which coincides with the year when the first sentencing guideline designed by the new Sentencing Council comes into force. However, the evidence presented here does not conclusively attribute the change to the guidelines. There are many different types of indictable offences, not all of them are covered by guidelines; and those which are, had their guidelines introduced at different times.

To assess the individual effect of each of the guidelines, we look independently at each offence group that best matches the offences covered by the guidelines studied, as shown in Table 1. Figure 3 represents the trend in severity for each of the six groups of offences studied. The vertical solid lines represent the date when the guidelines came into force, and the vertical dashed lines preceding them represent the date when the guidelines were published. These publication dates have been introduced to help visualise potential anticipatory effects that the guidelines might have had before they came into force (XXXX, 2014), when sentencers were still familiarising themselves with the new guidelines. Notice that the dates for the publication and the coming into force of the burglary and theft guidelines were included in the same plot since the MoJ data does not distinguish between burglary and theft offences.

(Figure 3 here)

In each offence group, the level of severity was higher at the end of the window of observation than at its beginning. However, detecting specific shifts in severity following the immediate implementation of the guidelines is more difficult. The question remains as to whether the observed increase in severity is due to the introduction of the new guidelines, or whether it reflects an ongoing trend in severity preceding the introduction of the guidelines. To investigate the specific effect that might be attributed to each of the guidelines, we assess the relative change following the two-year period after the implementation of each guideline. We compare that to the average relative change observed for other two-year periods across the whole window of observation. This is shown in Table 4, where we can see how the increase in severity is higher than expected following the introduction of the guidelines of assault, fraud, burglary, and theft, while the increase in severity was lower than expected following the guidelines of drugs, sex, and robbery.

(Table 4 here)

Results for the assault and burglary guidelines need to be interpreted carefully. The observed increase in severity might also be associated with the punitive response that followed the 2011 summer riots (Lightowlers and Quirk, 2015; Pina-Sánchez, Lightowlers and Roberts, 2017), which involved types of crimes encompassed by these two guidelines, and took place roughly at the same time as the assault and burglary guidelines were introduced. In addition, some of the differences between the expected and the observed change in severity reported in Table 4 are quite small. In order to test whether these represent statistically significant differences, we proceed to estimate the changes in severity for each of the group of offences using time-series models.

Time-Series Analysis

Time-series models have been previously used in the literature to test for structural changes produced by the introduction of sentencing guidelines (see D'Alessio and Soltzenberg, 1995; Soltzenberg and D'Alessio, 1994; Sentencing Council 2015, 2017b, 2018c, 2018d). These models allow us to predict the average severity following the implementation of the guidelines and to provide measures of uncertainty such as confidence intervals around those predictions by taking into account the serial correlation of the data. If the observed value of severity falls outside the forecasted region, we can conclude that the sentencing guidelines have had a statistically significant effect on sentence severity. Importantly, the models that we use allow us to account for any potential seasonal fluctuations in the levels of severity and for the trend that preceded the implementation of the guidelines.

For each guideline, we used the information from the beginning of the window of observation to the quarter before the guideline came into force. The specific ARIMA models used are shown in Table 5. The three figures in each of the parentheses in Figure 4 represent the autoregressive (AR), integrated (I), and moving average (MA) components. The model choice was based on AIC comparisons, carried out through the *auto.arima* command from the R package *forecast* (Hyndman and Khandakar, 2008). The time-series showing the forecasted regions following the two-year period after the implementation of the guidelines are represented in Figure 4.

(Table 5 here)

A statistically significant increase in severity is detected following the coming into force of the guidelines of assault and theft. The severity increase for assault is so vast and sustained in time that it can be concluded that the guidelines have contributed to this increase, beyond any added effect that could be attributed to the 2011 riots. The previously detected abnormal increase in severity following the fraud guideline was not found to be statistically significant. In total, five out of the seven guidelines did not show a significant effect on severity. The same findings are obtained if the forecast region is limited to one year after the guidelines. Hence, most guidelines 'pass the test', in that they seem to reflect current sentencing practice adequately. The observed increases in severity following the coming into force of each of those guidelines seem to be the product of an ongoing trend that preceded the arrival of the guidelines, with the exception of drug offences for which sentencing was decreasing in severity before the guidelines came into force. The question, then is, what else might have caused the remarkable increase in sentence severity observed in England and Wales over the last couple of decades.

(Figure 4 here)

5. Discussion

In this study, we have demonstrated the potential of a scale of severity to examine trends in sentencing. This is a useful strategy that allows the aggregation of sentence outcomes based on different units of measurement, and in so doing: i) eliminate the problem of selection bias affecting studies limited to custodial sentences; ii) make the most of the information available (i.e. accounting for differences across custodial sentences but also across disposal types); and iii) obtain more meaningful, robust and clearly interpretable evidence about the level of severity in sentencing in England and Wales. It is difficult to understand how, in spite of the advantages listed above this approach has not been pursued actively in the literature. It seems that after a solid start between the 1970s and 80s, the discipline of penal metric theory has practically vanished. This is particularly shocking given the remarkable dynamism that the measurement of crime seriousness has recently experienced (e.g. Ignatans and Pease, 2016; Sherman, Neyroud and Neyroud, 2016), an area of research from which much could be learnt.

The scale of severity presented in this article is more refined than the one currently in operation by the Sentencing Council (it accounts for different durations of suspended sentences, allows for overlaps in severity between certain disposal types, and each stage is clearly defined and reproducible). As such, we believe that the Council would benefit from adopting our scale in the future. The scale presented here could be substituted for their current scale in the Council's ex-post assessments of the impact of their guidelines, but it could also be used in the development of new guidelines. Using a scale of severity in the design of new guidelines would allow the Council to consider more systematically the relationship between non-custodial and custodial sentences in their design of starting points across guidelines. In so doing, the Council would be promoting the principle of consistency of approach that sentencers are expected to uphold, only in this case the principle would be applied to the Council's own design of sentencing guidelines.

We would also like to encourage other researchers to use our scale of severity, which could be applied to a wide range of areas, such as recent research on: the effect of aggravating and mitigating factors (Belton, 2018; Maslen, 2015; XXXX, 2017), consistency in sentencing (XXXX, 2017; XXXX, 2013; Reid and MacAlister, 2018), discrimination (Brandon and O'Connell, 2017; Lammy, 2017), proportionality (Fleetwood, Radcliffe and Stevens, 2015; Vibla, 2015), or deterrence, recidivism, and risk assessment (Bell, Jaitman and Machin, 2014; Raaijmakers et al., 2017).

The scale of severity presented here stems from the sentencing context of England and Wales, and as such, it is more amenable to being applied to studies focusing on this jurisdiction. Researchers working in other jurisdictions could nonetheless replicate our scale using views from their local sentencers on the disposal types available to them. We have demonstrated how the methodology used here is not excessively complex, and how using what can be known in advance about the relative severity of different pairs of sentences, the necessary questionnaire could be limited to a small number of questions and sample size.

Regarding more substantive findings, our analyses corroborate the previously observed increase in sentence severity in England and Wales (see Allen, 2016; Roberts and Ashworth, 2016; Roberts and Irwin-Rogers, 2015). Using the latest data available we have been able to observe that, although on aggregate this phenomenon has stabilised throughout this decade, the severity applied to indictable offences has increased at an even faster pace. Specifically, applying our new scale of severity we can determine that sentences are now 8.8% more severe than in 1999 on aggregate, and 34.7% more severe for indictable offences. The possibility of expressing overall changes in severity in simple percentages is an innovative feature from our study, which facilitates the monitoring of sentence severity in the future. Future studies might be able to offer more precise severity estimates. Given current data limitations we were not able to account for different types of fines, community orders and suspended sentences. However, the Sentencing Council is currently collecting new datasets differentiating fine bands, low, medium and high community orders, and durations of suspended sentences.

The detected increase in severity for indictable offences is so vast, and its implications so far-reaching (mainly as a result of the social and economic costs associated with undue incarceration, Prison Reform Trust, 2013), that it is of paramount importance to understand the causes behind this change. In this article we chose to test the potential effect that might be attributed to the new sentencing guidelines. The choice of this particular hypothesis seemed obvious to us given the intensity of the process of reform undertaken in England and Wales through the creation of sentencing guidelines.

Based on a time-series analysis we have been able to determine that, for the most part, the increase in severity cannot be attributed to the guidelines, but to an ongoing trend that preceded the arrival of the guidelines. There are three important exceptions to this. First, the increase in severity following the adoption of the assault and theft guidelines was higher than expected, which suggests that these guidelines have contributed to the overall increase in severity. This is problematic,

especially given the commitment expressed by the Sentencing Council to reflect the current sentencing practice in their design of new guidelines. It is reassuring, however, to know that the Council is fully aware of the unintended increase in severity brought about by their guidelines of assault, as identified in their own evaluation (Sentencing Council, 2015), and that they are currently working on a new version of these particular guidelines. Second, the increase in severity following the theft guidelines was also higher than expected. The Sentencing Council has not yet published an assessment of these guidelines but, in light of the evidence offered here, it might be worth prioritising the assessment of these particular guidelines over others that have been found to have no effect on severity. If our findings are corroborated the Council should prioritise the reformulation of these guidelines over older guidelines that have not been found problematic. Third, drug offences was the only group for which the average severity was declining before the publication of their guidelines, and although severity remained stable for the two years following its publication, it has sharply increased since then.

After falsifying - at least partially - the hypothesis linking the guidelines to the observed increase in severity, we need to ponder what other causes might be behind this phenomenon. In doing so, it might be worth taking a step back. Following the ongoing debate in the literature (Allen, 2016; XXXX, 2016; Roberts and Ashworth, 2016; Roberts and Irwin-Rogers, 2015) we have assumed that changes in sentence severity are the product of changes in sentencing practice, be that through the specific effect of the guidelines, or through more diffuse mechanisms associated with the phenomenon of penal populism. However, there is a simpler, yet possibly more meaningful, explanation for the observed increase in sentence severity. It is possible that the mix of cases processed through court has changed. If, i) the average seriousness of the cases processed, ii) the proportion of repeat offenders, or iii) those pleading guilty has changed, then we would expect to see a change in severity that could not be attributed to a change in sentencing.

Further analyses of MoJ statistics allowed us to rule out the last of these hypotheses. The proportion of defendants pleading guilty has remained relatively static over time. In 2006, 77.9% defendants entered a guilty plea for 79.3% in 2016. Similarly, the proportion of first-time entrants to court has remained relatively stable, accounting for 11.3% of all offenders in 2006 and 12.6% in 2016, although the proportion of the offending population with a long criminal career has increased from 28% in 2010 to 36% in 2017 (MoJ, 2018). Hence, some of the observed change in severity might be driven not by a change in sentencing practice, but by a different mix of offenders being processed. In addition, the first hypothesis remains to be explored. It is possible that even within the same group of offences, the cases processed today have become more serious. Indeed, using MoJ data covering the same period, and focusing on drug offences, we have found that, while the volume of drug offences has remained stable 39,478 cases processed in 2006 for 41,831 in 2016, the seriousness of the cases processed has changed remarkably. For example, the ratio of sentenced offences for the 'production, supply and possession with intent to supply class A drugs' (those deemed to be the most dangerous, e.g. heroin, cocaine, etc.) compared to class C drugs (those deemed to be the least dangerous, e.g. ketamine, anabolic steroids, etc.) was 64/36 in 2006, with that ratio reaching the lowest point in the series, 57/43, in 2008, and then rising dramatically to 96/4 in 2016 (following the reclassification of cannabis from class C to B in 2009). These changes in the seriousness of drug offences follow the time-plot presented in Figure 4 remarkably well, which highlights the need to control for the seriousness of the cases processed through court to be able to disentangle the different mechanisms affecting sentence severity. If the MoJ was to disclose their record level data currently stored at the MoJ was to be disclosed, future research could replicate the analyses undertaken here while controlling for key case characteristics such as the specific offence type, guilty plea, or the number of previous convictions, and in so doing improve our understanding of the causes behind the observed increase in sentence severity.

References

- Allen, R. (2016), 'The Sentencing Council for England and Wales: Brake or Accelerator on the Use of Prison?' *Transform Justice*. Available at: <http://www.transformjustice.org.uk/wp-content/uploads/2016/12/TJ-DEC-8.12.16-1.pdf>
- Apospori, E., and Alpert, G. (1993), 'Research Note: The Role of Differential Experience with the Criminal Justice System in Changes in Perceptions of Severity of Legal Sanctions Over Time', *Crime and Delinquency*, 39/2: 184--194.
- Ashworth, A. (2015), *Sentencing and Criminal Justice*. Cambridge University Press.
- Bell, B., Jaitman, L., and Machin, S. (2014), 'Crime Deterrence: Evidence from the London 2011 Riots', *The Economic Journal*, 124/576: 480--506.
- Belton, I. (2018), *The Role of Personal Mitigating Factors in Criminal Sentencing Judgments: An Empirical Investigation*. Doctoral dissertation, Middlesex University.
- Bottoms, A. E., Rex, S., and Robinson, G. (2004), *Alternative to Prison: Options for an Insecure Society*. Willan.
- Bradley, R. A. and Terry, M. E. (1952), 'Rank analysis of incomplete block designs: I. The method of paired comparisons', *Biometrika*, 39: 324--345.
- Brandon, A. M., and O'Connell, M. (2017), 'Same Crime: Different Punishment? Investigating Sentencing Disparities Between Irish and Non-Irish Nationals in the Irish Criminal Justice System', *The British Journal of Criminology*.
- Buchner, D. (1979), 'Scale of Sentence Severity', *The Journal of Criminal Law and Criminology*, 70/2: 182--187.
- Bushway, S., Johnson, B. D., and Slocum, L. A. (2007), 'Is the Magic Still There? The Use of the Heckman Two-step Correction for Selection Bias in Criminology', *Journal of Quantitative Criminology*, 23/2: 151--178.
- Chen, E. Y. (2008), 'Impacts of "Three Strikes and You're Out" on Crime Trends in California and Throughout the United States', *Journal of Contemporary Criminal Justice*, 24/4: 345--370.
- Cooper, J. (2013), 'Nothing Personal: The Impact of Personal Mitigation at Sentencing Since the Creation of the Council', in A. Ashworth, and J. V. Roberts, eds., *Sentencing Guidelines: Exploring the English Model*, 157--164. Oxford University Press.
- Croyle, J. L. (1983), 'Measuring and Explaining Disparities in Felony Sentences: Courtroom Work Group Factors and Race, Sex, and Socioeconomic Influences on Sentence Severity', *Political Behavior*, 5/1: 135--153.
- D'Alessio, S. J., and Stolzenberg, L. (1995), 'The Impact of Sentencing Guidelines on Jail Incarceration in Minnesota', *Criminology*, 33/2: 283--302.
- Erickson, M. L., and Gibbs, J. P. (1979), 'On the Perceived Severity of Legal Penalties', *The Journal of Criminal Law and Criminology*, 70/1: 102--116.
- Fleetwood, J., Radcliffe, P., and Stevens, A. (2015), 'Shorter Sentences for Drug Mules: The Early Impact of the Sentencing Guidelines in England and Wales', *Drugs: Education, Prevention and Policy*, 22/5: 428--436.
- Griset, P. L. (1999), 'Criminal Sentencing in Florida: Determinate Sentencing's Hollow Shell', *Crime and Delinquency*, 45/3: 316--333.

- Harlow, R. E., Darley, J. M., and Robinson, P. H. (1995), 'The Severity of Intermediate Penal Sanctions: A Psychological Scaling Approach for Obtaining Community Perceptions', *Journal of Quantitative Criminology*, 11/1: 71--95.
- Hebenton, B., and Pease, K. (1995), 'Weighing the Pound of Flesh: The Psychology of Punishment', in R. Bull, and D. Carson, eds., *The Handbook of Psychology in Legal Contexts.*, 375--391. John Wiley & Sons.
- Hutton, N. (2013), 'The Definitive Guideline on Assault Offences: The Performance of Justice', in A. Ashworth, and J. V. Roberts, eds., *Sentencing Guidelines: Exploring the English Model*, 86--103. Oxford University Press.
- Hyndman, R. J., and Khandakar, Y. (2008), 'Automatic Time Series Forecasting: The Forecast Package for R', *Journal of Statistical Software*, 27/3: 1--22.
- Ignatans, D., and Pease, K. (2016), 'Taking Crime Seriously: Playing the Weighting Game', *Policing: a Journal of Policy and Practice*, 10/3: 184--193.
- Kaye, E., and Firth, D. (2017), 'BradleyTerryScalable: Fits the Bradley-Terry Model to Potentially Large and Sparse Networks of Comparison Data', *R package version 0.1.0*. Available at: <https://github.com/EllaKaye/BradleyTerryScalable>
- Lammy, D. (2017), 'An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System', *The Lammy Review*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643001/lammy-review-final-report.pdf
- Leclerc, Ch., and Tremblay, P. (2016), 'Looking at Penalty Scales: How Judicial Actors and the General Public Judge Penal Severity', *Canadian Journal of Criminology and Criminal Justice*, 58/3, 354--384.
- Lightowers, C., and Quirk, H. (2015), 'The 2011 English 'Riots': Prosecutorial Zeal and Judicial Abandon', *The British Journal of Criminology*, 55/1: 65--85.
- Lightowers, C., and Pina-Sánchez, J. (2017), 'Intoxication and Assault: an Analysis of Crown Court Sentencing Practices in England and Wales', *The British Journal of Criminology*, 58/1: 132--154.
- Lovegrove, A. (2001), 'Sanctions and Severity: To the Demise of Von Hirsch and Wasik's Sanction Hierarchy', *The Howard Journal of Crime and Justice*, 40/2: 126--144.
- Marvell, T. B. (1995), 'Sentencing Guidelines and Prison Population Growth', *Journal of Criminal Law and Criminology*, 85/3: 696--709.
- Marvell, T. B., and Moody, C. E. (1996), 'Determinate Sentencing and Abolishing Parole: The Long-term Impacts on Prisons and Crime', *Criminology*, 34/1: 107--128.
- Maslen, H (2015), 'Penitence and Persistence: How Should Sentencing Factors Interact?' in J. V. Roberts, ed., *Sentencing Guidelines: Exploring Sentencing Practice in England and Wales*, 173--193. Palgrave.
- McClelland, K. M., and Alpert, G. P. (1985), 'Factor Analysis Applied to Magnitude Estimates of Punishment Seriousness: Patterns of Individual Differences', *Journal of Quantitative Criminology*, 1/3: 307--318.
- McDavid, J. C., and Stipack, B. (1981), 'Simultaneous Scaling of Offense Seriousness and Sentence Severity through Canonical Correlation Analysis', *Law and Society Review*, 16/1: 147--162.
- Ministry of Justice (2016), 'Story of the Prison Population: 1993-2016 England and Wales. Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/541667/prison-population-story-1993-2016.pdf

Ministry of Justice (2018), 'Criminal Justice System Statistics Quarterly: December 2017. Available at:

<https://www.gov.uk/government/statistics/criminal-justice-system-statistics-quarterly-december-2017>

Nicholson-Crotty, S. (2004), 'The Impact of Sentencing Guidelines on State-level Sanctions: An Analysis Over Time', *Crime and Delinquency*, 50/3: 395--411.

Padfield, N. (2016), 'Guidelines Galore', *Criminal Law Review*, 5: 301--302.

Petersilia, J., and Deschesnes, E. (1994a), 'What Punishes? Inmates Rank the Severity of Prison vs. Intermediate Sanctions', *Federal Probation*, 58/1: 3--8.

Petersilia, J., and Deschesnes, E. (1994b), 'Perceptions of Punishment: Inmates and Staff Rank the Severity of Prison Versus Intermediate Sanctions', *Prison Journal*, 74/3: 306--328.

Pina-Sánchez, J., and Grech, D. C. (2017), 'Location and Sentencing: To What Extent Do Contextual Factors Explain between Court Disparities?' *The British Journal of Criminology*, 58/3: 529--549.

Pina-Sánchez, J., Lightowers, C., and Roberts, J. (2016), 'Exploring the Punitive Surge: Crown Court Sentencing Practices Before and After the 2011 English Riots', *Criminology and Criminal Justice*, 17/3: 319--339.

Pina-Sánchez, J., and Linacre, R. (2013), 'Sentence Consistency in England and Wales: Evidence from the Crown Court Sentencing Survey', *British Journal of Criminology*, 53/6: 1118--1138.

Pina-Sánchez, J. and Linacre, R. (2014), 'Enhancing Consistency in Sentencing: Exploring the Effects of Guidelines in England and Wales', *Journal of Quantitative Criminology*, 30/4: 731--748.

Pratt, J. (2007), *Penal Populism*, London: Routledge.

Prison Reform Trust (2013), 'Prison: The Facts'. Available at:

<http://www.prisonreformtrust.org.uk/portals/0/documents/bromley%20briefings/summer%202017%20factfile.pdf>

Raaijmakers, E. A., Loughran, T. A., de Keijser, J. W., Nieuwbeerta, P., and Dirkzwager, A. J. (2017), 'Exploring the Relationship between Subjectively Experienced Severity of Imprisonment and Recidivism: A Neglected Element in Testing Deterrence Theory', *Journal of Research in Crime and Delinquency*, 54/1: 3--28.

Reid, A. A., and MacAlister, D. (2018), 'Extending a Geographical Perspective to the Study of Jurisdictional Consistency in Sentencing Outcomes', *The British Journal of Criminology*.

Revelle, W. (2018), 'psych: Procedures for Personality and Psychological Research, Northwestern University'. Available at: <https://CRAN.R-project.org/package=psych> Version = 1.8.4.

Roberts, J. V. (1992), 'Public Opinion, Crime, and Criminal Justice', *Crime and Justice*, 16: 99--180.

Roberts, J. V., and Ashworth, A. (2016). 'The Evolution of Sentencing Policy and Practice in England and Wales, 2003--2015', *Crime and Justice*, 45/1: 307--358.

Roberts, J. V., and Irwin-Rogers, K. (2015), 'Sentencing Practices and Trends: 1999-2013', in J. V. Roberts, ed., *Sentencing Guidelines: Exploring Sentencing Practice in England and Wales*, 76--92. Palgrave.

Schiff, M. F. (1997), 'Gauging the Intensity of Criminal Sanctions: Developing the Criminal Punishment Severity Scale (CPSS)', *Criminal Justice Review*, 22/2: 175--206.

- Sebba, L. (1980), 'Mitigation in Sentence in Order to Deter?' *Monash Law Review*, 6: 268--293.
- Sebba, L., and Nathan, G. (1984), 'Further Explorations in the Scaling of Penalties', *The British Journal of Criminology*, 24/3: 221--249.
- Sentencing Council (2011a), 'Analytical Note: The Resource Effects of Increased Consistency in Sentencing'. Available at: https://www.sentencingcouncil.org.uk/wp-content/uploads/Consistency_in_sentencing.pdf
- Sentencing Council (2011b), 'Drug 'Mules': Twelve Case Studies'. Available at: https://www.sentencingcouncil.org.uk/wp-content/uploads/Drug_mules_bulletin.pdf
- Sentencing Council (2011c), 'Assault Definitive Guideline'. Available at: https://www.sentencingcouncil.org.uk/wp-content/uploads/Assault_definitive_guideline_-_Crown_Court.pdf
- Sentencing Council (2015), 'Assault Offences Guideline Assessment: Synthesis Report'. Available at: <https://www.sentencingcouncil.org.uk/wp-content/uploads/Assault-assessment-synthesis-report.pdf>
- Sentencing Council (2017a), 'New Sentencing Guidelines for Magistrates' Courts'. Available at: <https://www.sentencingcouncil.org.uk/news/item/new-sentencing-guidelines-for-magistrates-courts/>
- Sentencing Council (2017b), 'Burglary Offences Initial Guideline Assessment'. Available at: <https://www.sentencingcouncil.org.uk/wp-content/uploads/Burglary-assessment.pdf>
- Sentencing Council (2018a), 'New Sentencing Guideline Introduced for the Possession of Weapons and Threats to Use Them'. Available at: <https://www.sentencingcouncil.org.uk/news/item/new-sentencing-guideline-introduced-for-the-possession-of-weapons-and-threats-to-use-them/>
- Sentencing Council (2018b), 'Sentencing Council Publishes New Guidelines on Domestic Abuse'. Available at: <https://www.sentencingcouncil.org.uk/news/item/sentencing-council-publishes-new-guideline-on-domestic-abuse/>
- Sentencing Council (2018c), 'Assessing the Impact and Implementation of the Sentencing Council's Drug Offences Definitive Guideline'. Available at: <https://www.sentencingcouncil.org.uk/wp-content/uploads/Drug-offences-guideline-assessment.pdf>
- Sentencing Council (2018d), 'Assessing the Impact of the Sentencing Council's Fraud, Bribery and Money Laundering Definitive Guideline'. Available at: <https://www.sentencingcouncil.org.uk/wp-content/uploads/Fraud-Bribery-and-Money-Laundering-guideline-assessment.pdf>
- Sherman, L., Neyroud, P. W., and Neyroud, E. (2016), 'The Cambridge Crime Harm Index: Measuring Total Harm from Crime Based on Sentencing Guidelines', *Policing: A Journal of Policy and Practice*, 10/3: 171--183.
- Spelman, W. (1995), 'The Severity of Intermediate Sanctions', *Journal of Research in Crime and Delinquency*, 32/2: 107--135.
- Stolzenberg, L., and D'Alessio, S. J. (1994), 'Sentencing and Unwarranted Disparity: An Empirical Assessment of the Long-Term Impact of Sentencing Guidelines in Minnesota', *Criminology*, 32/2: 301--310.
- Stolzenberg, L., and D'alessio, S. J. (1997), "'Three Strikes and You're Out": The Impact of California's New Mandatory Sentencing Law on Serious Crime Rates', *Crime and Delinquency*, 43/4: 457--469.

- Sullivan, C. J., McGloin, J. M., and Piquero, A. R. (2008), 'Modeling the Deviant Y in Criminology: an Examination of the Assumptions of Censored Normal Regression and Potential Alternatives', *Journal of Quantitative Criminology*, 24/4: 399--421.
- Thurstone, L. L. (1927), 'A Law of Comparative Judgement', *Psychological Review*, 34: 273--286.
- Tremblay, P. (1988), 'On Penal Metrics', *Journal of Quantitative Criminology*, 4/3: 225--245.
- Van Kesteren, J. (2009), 'Public Attitudes and Sentencing Policies across the World', *European Journal on Criminal Policy and Research*, 15/1-2: 25--46.
- Vibla, N. (2015), 'More Than One Crime: Sentencing the Multiple Conviction Offender', In J. V. Roberts, ed., *Sentencing Guidelines: Exploring Sentencing Practice in England and Wales*, 221--234. Basingstoke.
- Vujić, S., Commandeur, J. J., and Koopman, S. J. (2016), 'Intervention Time Series Analysis of Crime Rates: The Case of Sentence Reform in Virginia', *Economic Modelling*, 57: 311--323.
- Warr, M., Meier, R. F., and Erickson, M. L. (1983), 'Norms, Theories of Punishment, and Publicly Preferred Penalties of Crimes', *Sociological Quarterly*, 24: 75--91.
- Wasik, M. (2014), *A Practical Approach to Sentencing*. Oxford: Oxford University Press.
- Zatz, M. S., and Hagan, J. (1985), 'Crime, Time, and Punishment: An Exploration of Selection Bias in Sentencing Research', *Journal of Quantitative Criminology*, 1/1: 103-126.

Table 1. Matrix of severity representing the frequency with which sentence outcomes at the top of each column are considered more severe than sentence outcomes to the left of each row (1 = *always more severe*)

| | absolute discharge | conditional discharge | fine | community order | 1 month custody suspended for 6 months | 1 month custody suspended for 12 months | 6 month custody suspended for 6 months | 12 month custody suspended for 24 months | 1 month immediate custody | 2 months immediate custody | 3 months immediate custody |
|--|--------------------|-----------------------|------|-----------------|--|---|--|--|---------------------------|----------------------------|----------------------------|
| absolute discharge | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| conditional discharge | | | 0.69 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| fine | | | | 0.77 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| community order | | | | | 0.39 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 month custody suspended for 6 months | | | | | | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 month custody suspended for 12 months | | | | | | | 0.74 | 1 | 1 | 1 | 1 |
| 6 month custody suspended for 6 months | | | | | | | | 1 | 1 | 1 | 1 |
| 12 month custody suspended for 24 months | | | | | | | | | 0.41 | 0.51 | 0.61 |
| 1 month immediate custody | | | | | | | | | | 1 | 1 |
| 2 months immediate custody | | | | | | | | | | | 1 |
| 3 months immediate custody | | | | | | | | | | | |

Table 2. Severity scores

| Sentence outcome | Council's scale | Our scale |
|---|-----------------|-----------|
| absolute discharge | 0 | 0 |
| conditional discharge | 0 | 0.97 |
| fine | 10.25 | 1.32 |
| community order | 21.65 | 2.12 |
| 1 month custody suspended for 6 months | - | 2.34 |
| 1 month custody suspended for 12 months | - | 3.66 |
| 6 months custody suspended for 6 months | - | 3.78 |
| 12 months custody suspended for 24 months | - | 5.74 |
| 1 month immediate custody | 29.37 | 5.05 |
| 2 months immediate custody | 29.93 | 5.75 |
| 3 months immediate custody | 30.48 | 6.45 |
| 1 year immediate custody | 35.26 | 13.45 |
| 5 years immediate custody | 55.89 | 47.05 |
| 20 years immediate custody | 100 | 173.05 |

Table 3. Offence group and sentencing guidelines

| Offence group | Sentencing guideline | Publication | Came into force |
|-----------------------------|----------------------|---------------|-----------------|
| Violence against the person | Assault | March 2011 | June 2011 |
| Sex | Sex | December 2013 | April 2014 |
| Drugs | Drugs | January 2012 | February 2012 |
| Fraud | Fraud | May 2014 | October 2014 |
| Theft | Burglary | October 2011 | January 2012 |
| Theft | Theft | October 2015 | February 2016 |
| Robbery | Robbery | January 2016 | April 2016 |

Figure 1. Trends in the use of different disposal types and on average custodial sentence length

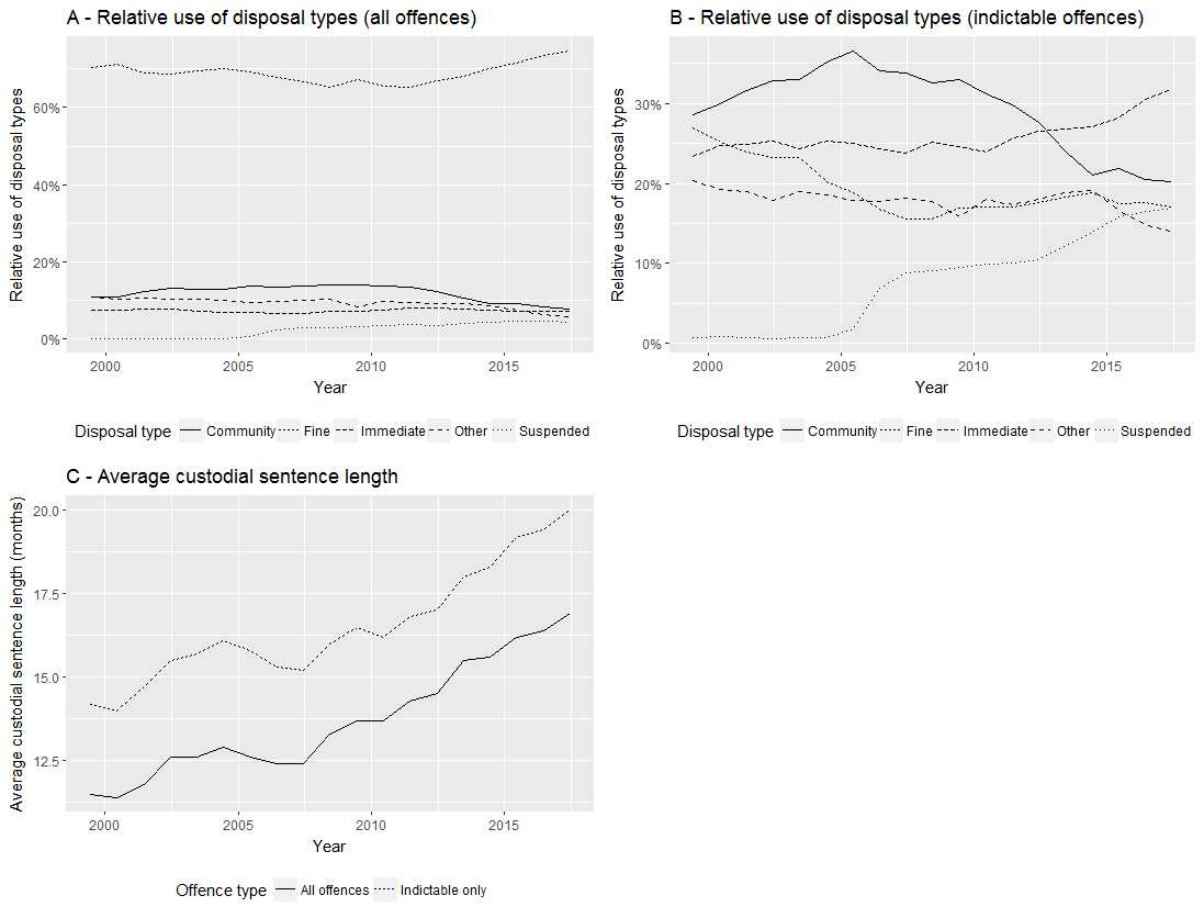


Figure 2. Trends in aggregated severity for all offences and for 'indictable only' offences

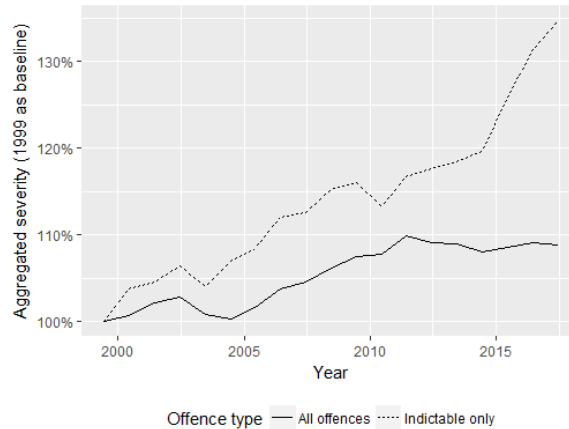


Figure 3: Sentence severity trends by offence group

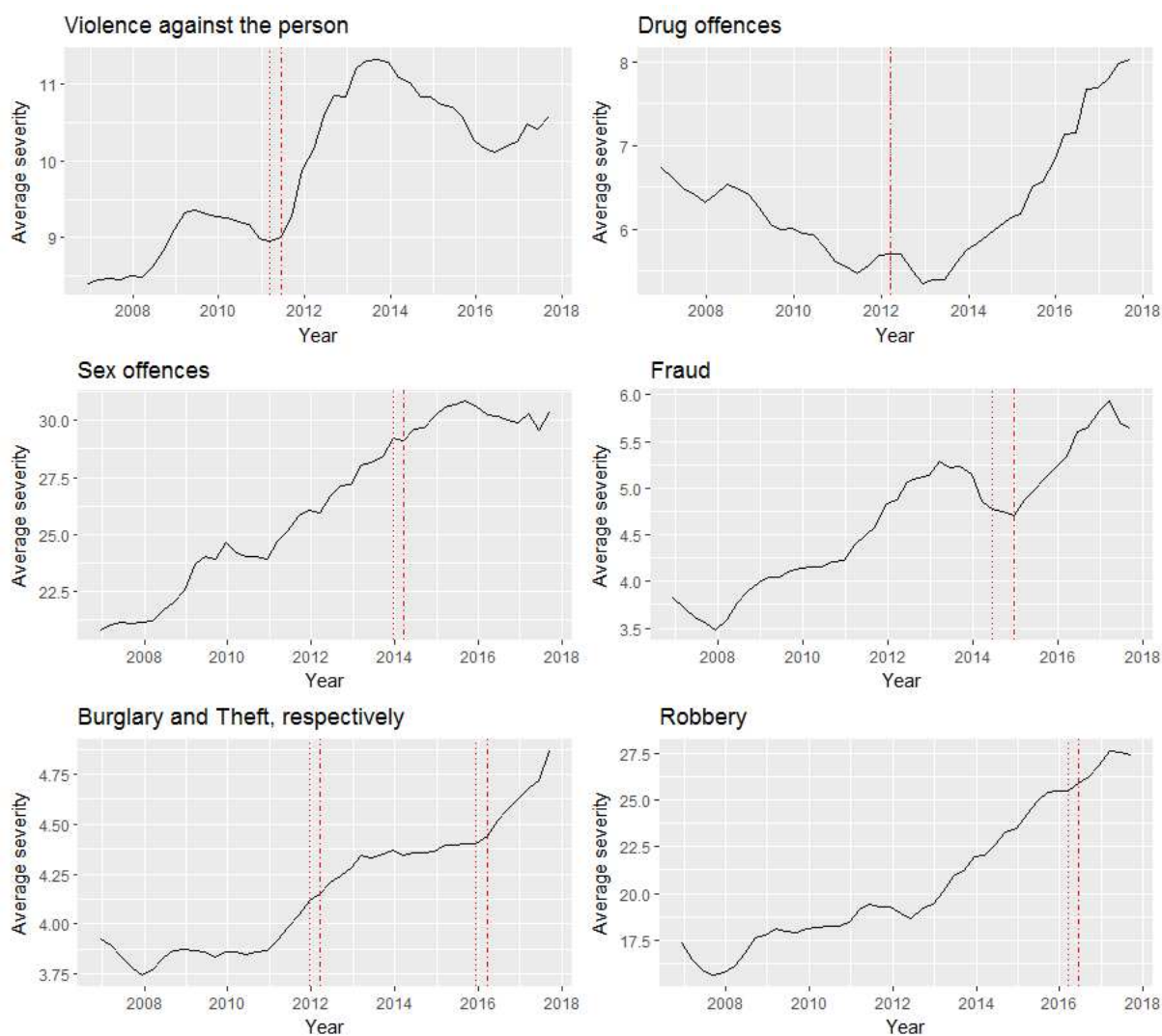


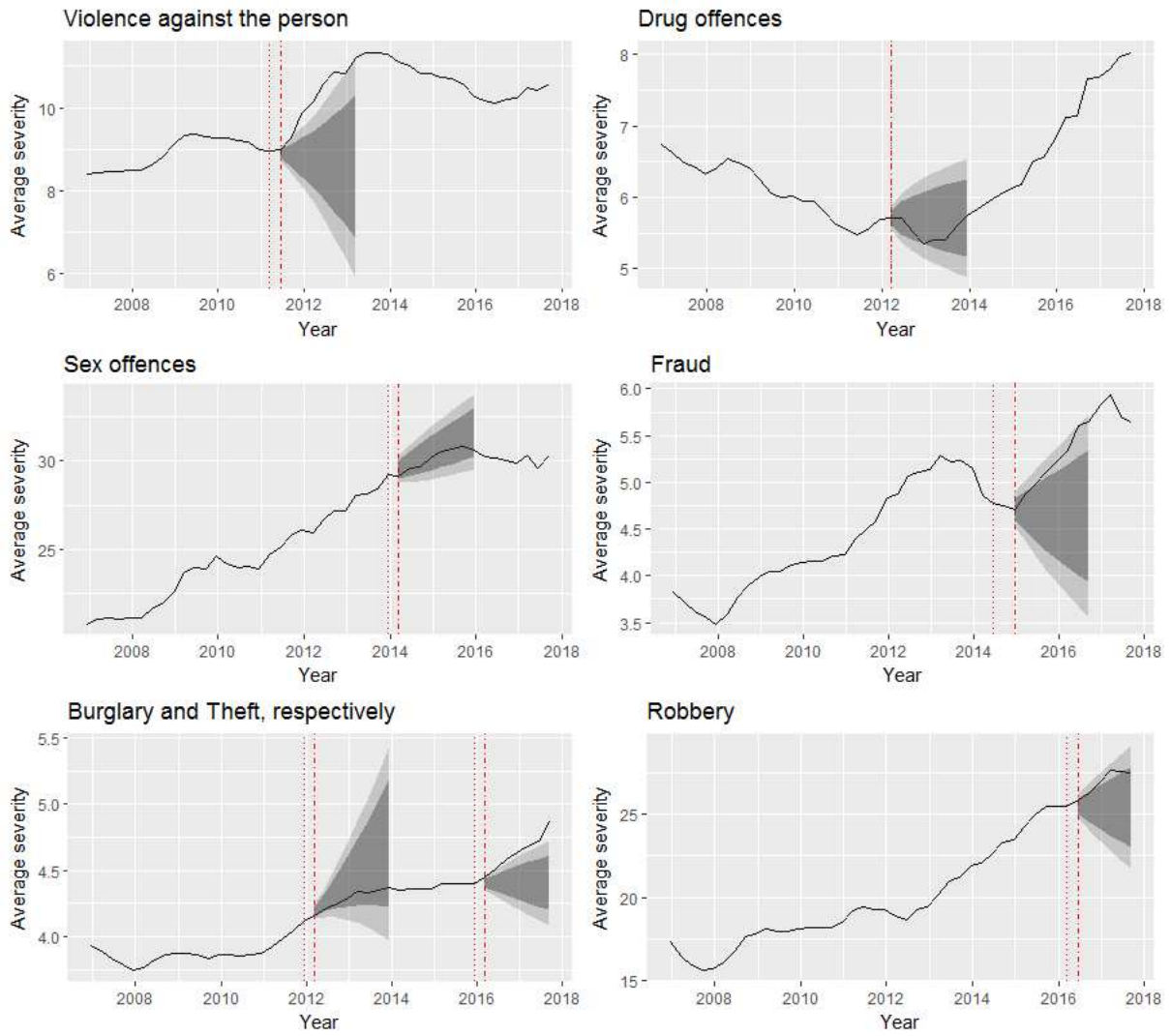
Table 4. Relative change in severity

| Guideline | Change from Dec 2007 | Average change over two-year periods | Change over the two years after the guideline |
|-----------|----------------------|--------------------------------------|---|
| Assault | 26.1% | 4.7% | 25.6% |
| Drugs | 19.1% | 3.5% | 2.1% |
| Sex | 45.8% | 8.3% | 3.7% |
| Fraud | 47.7% | 8.7% | 23.6% |
| Burglary | 23.9% | 4.3% | 4.7% |
| Theft | 23.9% | 4.3% | 9.5% |
| Robbery | 58.1% | 10.6% | 6.1% |

Table 5. ARIMA models

| Guideline | Assault | Drugs | Sex | Fraud | Burglary | Theft | Robbery |
|-----------|---------|---------|---------|---------|----------|---------|---------|
| ARIMA | (0,2,0) | (0,1,1) | (0,1,0) | (1,1,0) | (0,2,0) | (1,1,0) | (1,1,0) |

Figure 4. Time-series with forecast regions*



* Forecast region based on 80% and 95% confidence intervals.