Journal of Legal Research Methodology (JLRM)
ISSN 2752-3403
Montanez - 1547

https://doi.org/10.19164/jlrm.v4i1.1547

Published date: 20/10/2025

Article

Unified Search, Analysis, and Reporting Protocols in United States Policy Surveillance: A Guide and Call-to-Action

Julio Montanez
Department of Sociology, University of Central Florida

Abstract

Multi-jurisdictional legal research is an important area of study for understanding the United States's (U.S.) legal landscape, including the impact of this landscape on social issues (e.g., overdose response, violent victimization). However, underexplored within the extant literature is unified and systematic guidance on conducting such research. Accordingly, the goal of the current paper is to construct a guide and call to action on bringing policy surveillance methods into focus. First, a systematized review of the extant empirical literature on multi-jurisdictional domestic violence policy surveillance is employed by inputting a search phrase—(statut* OR legisl* OR law* OR "policy" OR "policies") AND "content analysis" AND "United States" AND (violen* OR abus*)—into three scholarly databases: Criminal Justice Abstracts, Academic Search Premier, and Applied Social Sciences Index & Abstracts. Second, a systematized review of the extant literature on policy surveillance methodology more broadly is employed by inputting a search phrase—"policy surveillance"—into the scholarly database, Web of Science. After inclusion/exclusion and data abstraction processes, as well as with the information gained from the systematized reviews more broadly, the current work (a) constructs a series of common methodological

Affiliation: Department of Sociology, University of Central Florida
Corresponding Author: Julio Montanez | julio.montanez@ucf.edu, jrichmontanez@gmail.com | 407-823-3744

practices in policy surveillance and (b) develops a call-to-action on necessary future steps to ensure wide usage of unified policy surveillance guidance. Overall, the importance of the current work is embodied in an empirically-informed set of options for searching, analysis, and reporting of multi-jurisdictional policy surveillance research.

Keywords: United States, Policy surveillance, Legislation, Geography.

Article Text

1. Introduction

Multi-jurisdictional legal research is an important area of study for understanding the United States's (U.S.) legal landscape—as well as its impact on extant social issues. However, this area of research is plagued by a lack of unified options for searching for, as well as reporting the details of, state-level statute and regulation analyses. This paper aims to serve as a guide and call to action on this matter. Particularly, I conduct two back-of-the-envelope systematized reviews by including as many major Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines as suitable. The systematized reviews looked at (a) state-level statutes in the U.S. dealing with domestic violence and (b) policy surveillance methodology more broadly. Systematized reviews encompass some, but not all, components of systematic reviews.[1] Using the systematized reviews to draw context and examples, the current paper then describes a call-to-action on unified procedures for the search, analysis, and reporting of state policy surveillance in the U.S. It then develops a call-to-action regarding how to progress the field of policy surveillance forward. I highlight granular methodological

[1] Marjia J. Grant and Andrew Booth, 'A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies' (2009) 26 Health Information and Libraries Journal.

details in policy surveillance (e.g., number of coders, databases, coding discrepancy resolution).

Policy surveillance can be defined as the systematic excavation, categorization, and presentation of laws for the purpose of tracking their geographic distribution cross-sectionally or longitudinally. The value associated with this call-to-action and guide is manifold. First, such guidance will help scholars identify study components that must suit their research questions. Second, the guide will assist with navigating the complex methodological landscape of statute research. Accordingly, it is hoped that this article can provide guidance and standards that can be adapted for other areas of legal research as well.

2. Methods

To create an underlying methodological/analytical framework through which an understanding of policy surveillance can be based, I conducted two systematized reviews, one at the narrative level (i.e., for background information on policy surveillance) and one at the quantitative level (i.e., for examples on methodology).

2.1. Review #1

To construct a search for extant guidance on policy surveillance studies, Review #1, one term was used: *"policy surveillance"*. To construct the search further, one database was used. Web of Science was used due to its generalness and wide scope of the literature. Pre-screening was conducted removing abstracts, meeting abstracts, corrections, editorials, reviews, and book chapters. At the title and abstract level, excluded records were those that applied policy

surveillance empirically to a topic. At the full-text level, excluded records were those that (a) had more of an applied than methodological focus (qualitatively measured) and (b) not enough discussion about methodology. This search also captured two domestic violence policy surveillance studies, which were included in the second review (see Section 2.2). A flow diagram of the inclusion and exclusion screening processes can be found in Figure 1.

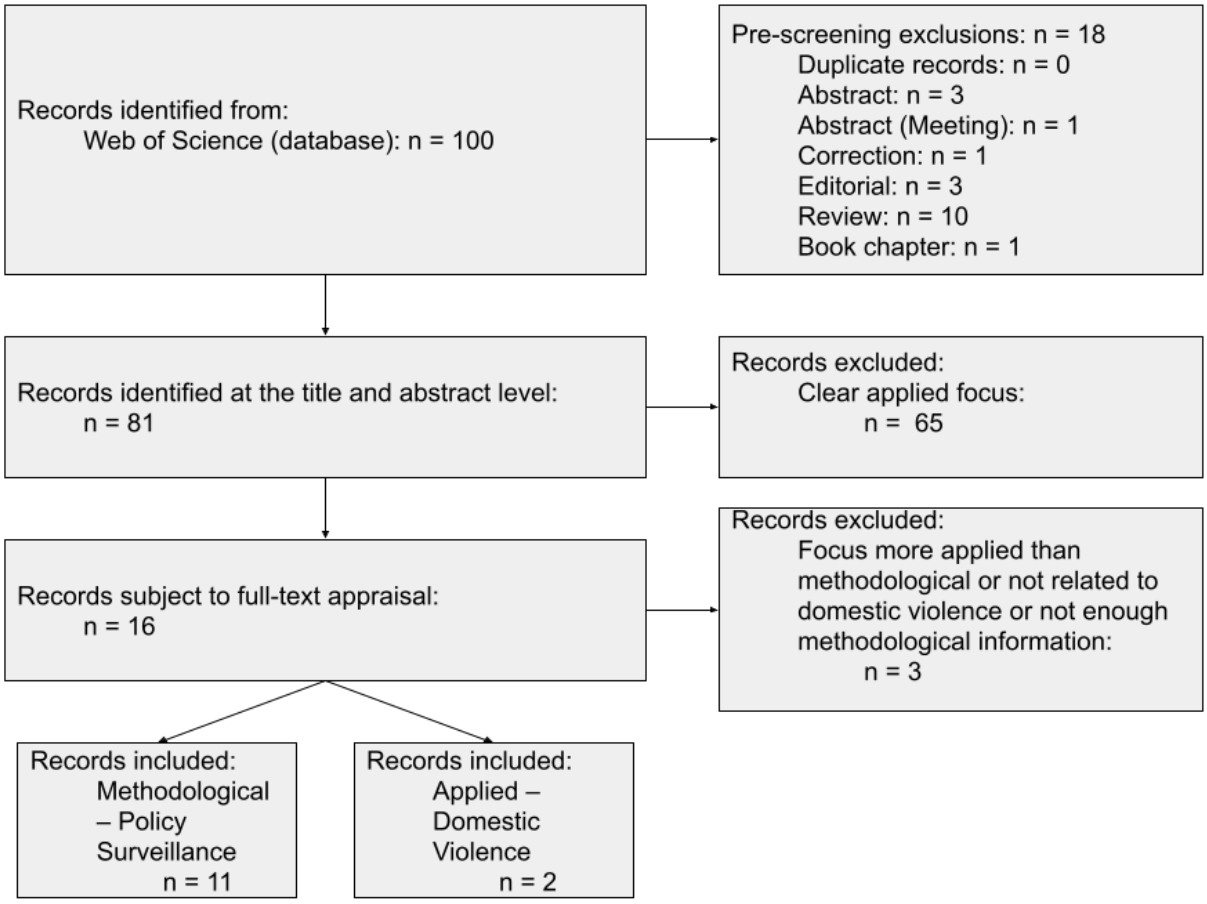

Figure 1. Flow diagram depicting inclusions and exclusions of the review.

Most of the results of Review #1 were not subject to data extraction. Instead, the 11 methodological papers (for a list, see Table 1) were interspersed throughout the remainder of the current study to create a baseline framework for understanding policy surveillance. These

11 methodological papers were integrated narratively through the current research to also give context for the examples and methodological intricacies found in Review #2.

Table 1. Reference list entries of articles found in Systematized Review #2.

#	Citation/Reference
1	Ross C. Brownson and others, 'Understanding Evidence-Based Public Health Policy' (2009) 99 American Journal of Public Health 1576.
2	Matthew Fifolt and others, 'Preliminary Findings of the Birmingham Policy Surveillance Initiative' (2023) 29 Journal of Public Health Management and Practice 2010.
3	James Hodge, 'The Promises (and Pitfalls) of Public Health Policy Surveillance' (2016) 41 Journal of Health Politics, Policy, and Law 1175.
4	Scott Burris and others, 'Policy Surveillance: A Vital Public Health Practice Comes of Age' (2016) 41 Journal of Health Politics, Policy, and Law 1151.
5	David Presley and others, 'Creating Legal Data for Public Health Monitoring and Evaluation: Delphi Standards for Policy Surveillance' [2015] Journal of Law, Medicine, and Ethics 27.
6	Abraham Gutman and others, 'Law as Data: Using Policy Surveillance to Advance Housing Studies' (2019) 21 Cityscape: A Journal of Policy Development and Research 203.
7	Jamie F. Chrique and others, 'What Gets Measured, Gets Changed: Evaluating Law and Policy for Maximum Impact. [2011] Journal of Law, Medicine, & Ethics 21.
8	Matthew Kavanaugh and others, 'Global Policy Surveillance: Creating and Using Comparative National Data on Health Law and Policy' (2020) 110 American Journal of Public Health 1805.
9	Lindsey Sanner and others, 'The Challenges of Conducting Intrastate Policy Surveillance: A Methods Note on County and City Laws' (2021) 111 American Journal of Public Health 1095.
10	Katie Moran-McCabe, Abraham Gutman, and Scott Burris, 'Public Health Implications of Housing Laws: Nuisance Evictions' (2010) 133 Public Health Reports 606.
11	Aila Hoss and others, 'Yes, You Need a Lawyer: Integrating Legal Epidemiology into Health Research' (2020) 135 Public Health Reports 856.

In Review #1, I found two domestic violence-related papers that empirically conducted policy surveillance (see Table 2). These two empirical works were transferred to Review #2 for data extraction.

Table 2. Reference list entries of articles found in Systematized Review #2.

#	Citation/Reference
1	Lindsay K. Cloud, Nadya Prood, and Jennifer Ibrahim, 'Disarming Intimate Partner Violence Offenders: An In-Depth Descriptive Analysis of Federal and State Firearm Prohibitor Laws in the United States, 1991-2016. (2023) 38 Journal of Interpersonal Violence 5164.
2	Avanti Adhia and others, 'Assessment of Variation in US State Laws Addressing the Prevention of and Response to Teen Dating Violence in Secondary Schools' (2022) 176 JAMA Pediatrics 797.

2.2. Review #2

Review #2 featured the conducting of a systematized review that focused on the previously-published empirical research on policy surveillance on domestic violence laws in the U.S. To construct a search for domestic violence studies that deal with state-level statutes, a series of search terms were developed into a search phrase and input into various databases. First, legal terminologies relevant to legislation were used: *statut* OR legisl* OR law* OR "policy" OR "policies"*. Then, a methodology term was included to narrow the focus: "content analysis." A geographic indicator was also included to exclude non-U.S. studies: "United States." Finally, two violence indicator words were used: *violen* OR abus**. Together the following search phrase was constructed: (*statut* OR legisl* OR law* OR "policy" OR "policies") AND "content analysis" AND "United States" AND (violen* OR abus*)*.

To construct the search further, three databases were used. Criminal Justice Abstracts was used to, in part, account for the criminological nature of domestic violence law. Academic Search Premier was used to represent a generalized scope of searching. Applied Social Sciences Index & Abstracts was used to, in part, nest social science literature within the catchment of the current paper's search strategy. The search phrase was entered into each database.

To narrow the number of records handled and focus the systematized review, a series of (pre-)screening steps were taken. First, duplicates were removed electronically through the spreadsheet processing program, Microsoft Excel. An initial title and abstract screening was then employed, excluding (a) symposia, (b) proceedings, (c) posters, (d) non-English works, (e) works that were not DV-related, and (f) works that were not policy-related. A follow-up full-text appraisal was then conducted, including only those records that met the following criteria: (a) U.S.-based, (b) policy-specific, (c) having policy coding, and (d) disaggregated analyses at the state level. The inclusion/exclusion screening process can be seen in Figure 2.

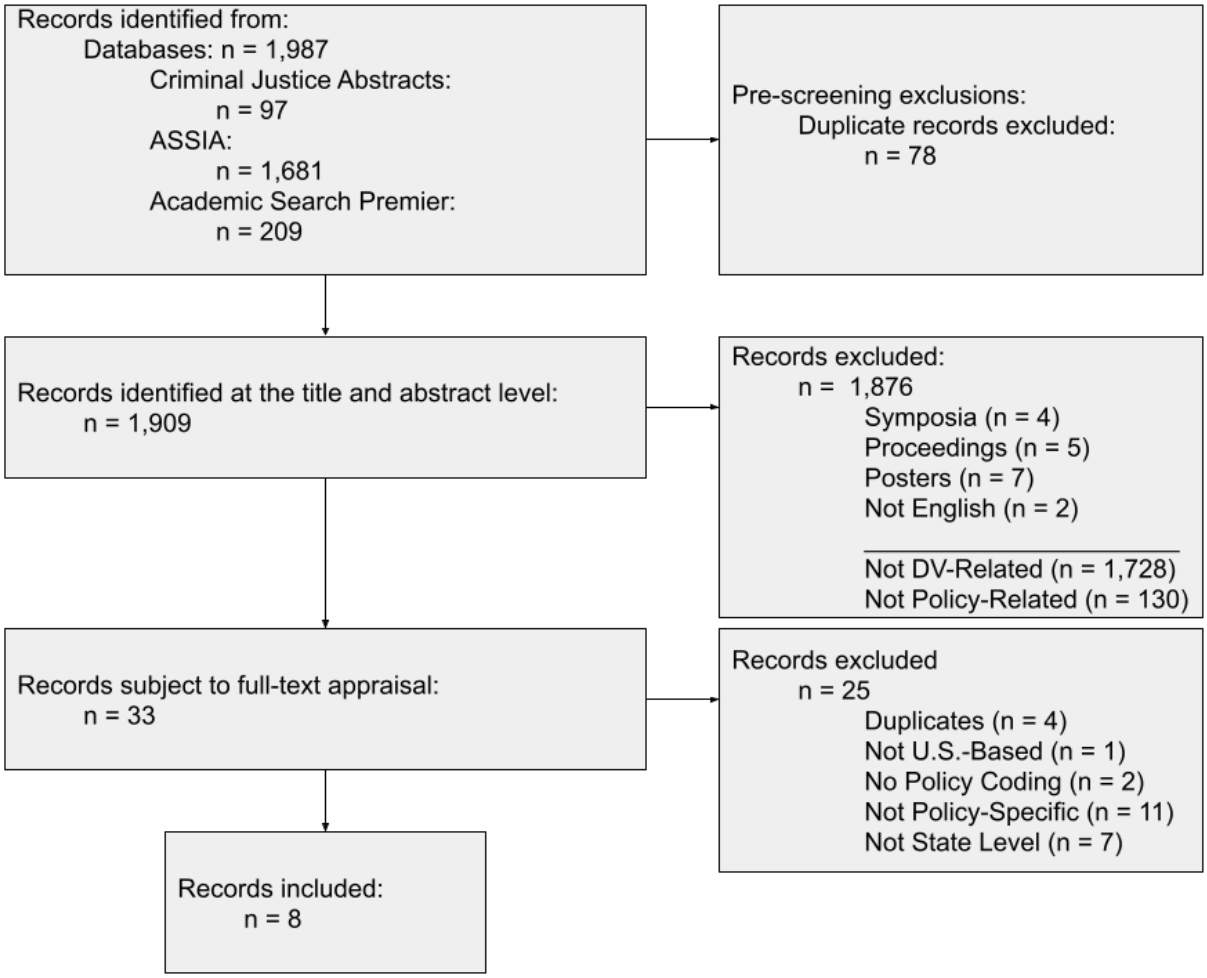

Figure 2. Flow diagram depicting inclusions and exclusions of the review. Note: ASSIA = *Applied Social Sciences Index & Abstracts*.

The eight records included from Review #2 can be found in Table 3. These articles empirically engaged in policy surveillance of several topics related to domestic violence. These include stalking, neglect, batterer intervention, and employment protections, among others. The two records from Table 2 were merged with the eight records from Table 3 to sum to 10 articles subject to data extraction.

Table 3. Reference list entries of articles found in Systematized Review #1.

#	Citation/Reference
1	Ethan C. Levine, 'Sexual Scripts and Criminal Statutes: Gender Restrictions, Spousal Allowances and Victim Accountability After Rape Law Reform.' (2018) 24 Violence Against Women 322.
2	Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment 303.
3	Caralin Branscrum and others, 'Stalking State Statutes: A Critical Content Analysis and Reflection on Social Science Research. [2021] Women & Criminal Justice 261.
4	Paulina Flasch and others, 'State Standards for Batterer Intervention Programs: A Content Analysis' (2021) 36 Violence and Victims 683.
5	Jennifer E. Swanberg, Mamta U. Ojha, and Caroline Macke, 'State Employment Protection Statutes for Victims of Domestic Violence: Public Policy's Response to Domestic Violence as an Employment Matter.' (2011) 27 Journal of Interpersonal Violence 587.
6	Emily M. Douglas and Sean C. McCarthy, 'Child Fatality Review Teams: A Content Analysis of Social Policy' (2011) 90 Child Welfare 91.
7	Hannah I. Rochford and others, 'United States' Teen Dating Violence Policies: Summary of Policy Element Variation' (2022) 43 Journal of Public Health Policy 503.
8	Michele Cascardi and others, 'School-Based Bullying and Teen Dating Violence Prevention Laws: Overlapping or Distinct?' (2018) 33 Journal of Interpersonal Violence 3267.

3. Quantitative Results/Data Extraction of Reviews

Table 4 presents quantitative data extraction results (n = 10) stemming from Review #1 (n = 2) and Review #2 (n = 8). Most studies focused on the 50 states without D.C. While some studies used the Westlaw database, others used Lexis Nexis. Several studies also used legislative/government websites. While most studies focused on statutes, others looked at administrative regulations. The use of more than one coder was common practice. Zero studies included a flow diagram for visually illustrating inclusion and exclusion processes.

Table 4. Data abstraction for systematized reviews (*N* = 10).

Characteristic	*n*	%
Jurisdictions under consideration		
50 states	5	50.0
50 states and D.C.	4	40.0
< 50 states	1	10.0
Search Strategy		
Westlaw database search		
Singular search terms used	0	0.0
Cross-tabulated search terms used	3	30.0
No search terms explicitly listed	1	10.0
No search terms used	6	60.0
Lexis Nexis / Nexis Uni database search		
Singular search terms used	0	0.0
Cross-tabulated search terms used	2	20.0
No search terms used	7	70.0
Other database used		
Singular search terms used	0	0.0
Cross-tabulated search terms used	1	10.0
No search terms used	9	90.0

Use of government documents/websites	6	60.0
Use of non-government documents/websites	4	40.0
Use of existing agencies - Phone calls	2	20.0
Use of existing agencies - Emails	2	20.0
Use of other sources - Not specified	1	10.0
Type of Law		
Statutes	9	90.0
"Policies"	1	10.0
Administrative Regulations	4	40.0
Number of Coders		
1	1	10.0
2	4	40.0
3+	4	40.0
Not specified	1	10.0
Statistical Computations		
% agreement	5	50.0
Krippendorff's alpha	2	20.0
Cohen's kappa	1	10.0
Discrepancy resolution		
Not mentioned	3	30.0
Discussion until 100% agreement	3	30.0

Pass-off to separate review to break ties	2	20.0
Pilot coding of subsample of laws		
Yes	6	60.0
Random number generation for subsample	2	20.0
Results presentation		
Narrative without examples	5	50.0
Narrative with examples	5	50.0
Mathematical	4	40.0
Tabular	10	100.0
Geospatial	2	20.0
Graphical	1	10.0
Inclusion/exclusion mentioned/described	6	60.0
Flowchart for inclusion/exclusion	0	0.0

4. Unified Protocols and Policy Surveillance

Unlike evidence-based synthesis methods used to understand extant literatures in research, the field of legal research's unified guidance on how to conduct policy-related projects lacks reach into the literature. Evidence of this matter can be traced to the wide variation in techniques used to search, code, analyze, and report on statute research across the U.S. Indeed, evidence-based synthesis methods have major, overarching technical guidance, such

as PRISMA,[2] JBI,[3] and the Cochrane Collaboration.[4] While uniform guidelines for policy collection and analysis exist (e.g., the Policy Surveillance Program),[5] scholarly reiteration and advancing of such guidelines are needed to:

- Provide clear options for learning about how to conduct multi-jurisdictional legal research.
- Foster uniformity across fields, and thus, more streamlined communication.

The policy surveillance methodology literature seems to be bisected into requirements and challenges. One requirement within the policy surveillance literature is that such studies should be systematic and should be able to be redone through a standardized methodology.[6] Indeed, documenting search processes and reporting them transparently is a hallmark of standardized methodology. Furthermore, this may involve keeping track of search terms/phrases, having inclusion/exclusion criteria, and a paper trail of coding for capturing the textual elements of law.[7]

[2] Matthew J. Page and others, 'The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews' [2021] 89 Systematic Reviews.
[3] Edoardo Aromataris and others (eds) 'JBI Manual for Evidence Synthesis' (JBI 2024) <https://synthesismanual.jbi.global> accessed 8 April 2024.
[4] JPT Higgins and others (eds), *Cochrane Handbook of Systematic Reviews of Interventions* (version 6.4, Cochrane 2023) <https://www.training.cochrane.org/handbook> accessed 8 April 2024.
[5] The Policy Surveillance Project, 'Learning Library' (LawAtlas) <https://lawatlas.org/page/lawatlas-learning-library> accessed 8 April 2024.
[6] Alia Hoss and others, 'Yes, You need a Lawyer: Integrating Legal Epidemiology into Health Research' (2020) 135 Public Health Reports; see also Matthew M. Kavanaugh and others, 'Global Policy Surveillance: Creating and Using Comparative National Data on Health Law and Policy' (2020) 110 American Journal of Public Health.
[7] Matthew Fifolt and others, 'Preliminary Findings of the Birmingham Policy Surveillance Initiative' (2023) 29 Journal of Public Health Management and Practice.

A challenge to the conduct of policy surveillance involves access to information at the multi-jurisdictional level.[8] For example, Jurisdiction A may have its own search platform for searching and browsing laws. Jurisdiction B may have its laws posted on Lexis Nexis. Jurisdiction C may not have their laws publicly available online at all. These differences create a challenge for the uniform application of systematic searching (e.g., using keywords) across platforms.

Another challenge is that studies may be conducted without staff who have adequate legal training. That is, having lawyers on a policy surveillance research team holds the promise of increasing the validity of the research. Extant research states that lawyers are needed on such teams.[9] Stacked on top of this challenge is the challenge of timing and updating once initial surveillance has been conducted.[10] For example, by the time a policy surveillance study has been published, laws may have already changed. Furthermore, some literature cautions not to "oversell" the potential impact and importance of policy surveillance.[11]

[8] Lindsey Sanner and others, 'The Challenges of Conducting Intrastate Policy Surveillance: A Methods Note on County and City Laws' (2021) 111 American Journal of Public Health; see also see also Matthew M. Kavanaugh and others, 'Global Policy Surveillance: Creating and Using Comparative National Data on Health Law and Policy' (2020) 110 American Journal of Public Health; Abraham Gutman and others, 'Law as Data: Using Policy Surveillance to Advance Housing Studies' (2019) 21 Cityscape: A Journal of Policy Development and Research; James Hodge, 'The Promises (and Pitfalls) of Public Health Policy Surveillance' (2016) 41 Journal of Health Politics, Policy, and Law.

[9] Scott Burris and others, 'Policy Surveillance: A Vital Public Health Practice Comes of Age' (2016) 41 Journal of Health Politics, Policy, and Law; Aila Hoss and others, 'Yes, You Need a Lawyer: Integrating Legal Epidemiology into Health Research' (2020) 135 Public Health Reports.

[10] Matthew Fifolt and others, 'Preliminary Findings of the Birmingham Policy Surveillance Initiative' (2023) 29 Journal of Public Health Management and Practice.

[11] James Hodge, 'The Promises (and Pitfalls) of Public Health Policy Surveillance' (2016) 41 Journal of Health Politics, Policy, and Law.

5. Guide to Common Policy Surveillance Methods

5.1. Search Strategies

One of the first choices of policy surveillance research is the overall search strategy, particularly in terms of search scope. The overall search strategy will depend on the goal of the study in terms of broadness and specificity and prior knowledge. There are two overarching options in this regard:

- *Searching by jurisdictio*n. When analyzing neglect and stalking statues across the U.S., some authors sift through the government and legislative websites.[12] One option for excavating statute information is venturing to each state's statute or legislative website (or host website) and looking for the codes manually. This option will be of best use if a researcher already has knowledge of which statutes they want to research. For example, if someone wants to search specifically for statutes on homicide, they may venture to the crime/criminal/criminal procedure chapters(s) of each state's statutes and pinpoint the relevant sections related to homicide.
- *Searching by legal database*. Another option for unearthing statutes is entering search terms in a legal database. This option will be of best use if a researcher is unaware of the breadth of statutes that exist across the United States. It is also a good option for exploratory research. For example, if someone wants to search specifically for statutes on officer-perpetrated domestic violence—but is unsure of where the statutes would be located—it would be best to gather some potential search terms and enter them into a legal database. For example, one study on dating violence statutes across the

[12] Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment.

U.S. used the Westlaw database, specifically by entering search terms like "domestic violence" and "education."[13]

5.2. Constructing Search Terms

Search terms can be constructed in various ways. These strategies can be subdivided into two categories:

- *Singular search term domains*. A singular search term strategy is one in which only one set of subject-specific terms is input into the database. When searching by jurisdiction, each state will have their own platform to which the search terms will need to be adapted. When searching by legal database, terms are generally entered as a string of words with the Boolean indicator, *OR* (i.e., the database's version of the indicator), for example: *immig* OR alien* OR undocumented*.
- *Cross-tabulated search term domains*. A cross-tabulated search term strategy involves combining search terms of two or more broader topics to construct a relevant search phrase. For example, if a researcher is looking at immigration and domestic violence, they may construct a search phrase that includes immigration terminologies and domestic violence terminologies, for example, "immigrant" AND "domestic violence."[14]

5.3. Defining the Scope of the Issue and Type of Law

[13] Hannah Rochford and others, 'United States' Teen Dating Violence Policies: Summary of Policy Element Variation' (2022) 43 Journal of Public Health Policy. See also Karisa Harland and others, 'State-Level Teen Dating Violence Education Laws and Teen Dating Violence Victimisation in the USA: A Cross-Sectional Analysis of 36 States' (2021) 27 Injury Prevention.
[14] Julio Montanez and others, *Between Systems and Violence: State-Level Policy Targeting Intimate Partner Violence in Immigrant and Refugee Lives* (Routledge 2022).

Studies collectively oscillate between jurisdictional focus. Particularly, some studies cover the 50 states of the U.S. Others cover the 50 states, as well as the District of Columbia (D.C.). D.C. is an important, nuanced consideration for both scopes because of its unique subnational position within the U.S.; it is (a) comparable to a state in terms of population; and (b) under the jurisdiction of the Congress of the U.S. For example, as Levine included D.C. in their analysis,[15] Cascardi and colleagues restricted their analysis to the 50 states without D.C.[16] Further disaggregating the nuances of law in the U.S., there are generally two levels of state-level law. The first are statutory codes. The second are administrative regulations. In terms of impacts on study methodology strength, the usage and non-usage of statutory and administrative codes forms two permutations that substantively impact results. First, choosing to analyze statutes instead of regulations (or regulations instead of statutes) allows researchers to more cleanly and clearly focus their studies, although gaps in understanding policy impacts may render the study incomplete. Second, in choosing to analyze both, coding for statutes can fill in the data gaps of regulation coding, just as coding for regulations can fill the data gaps of statute coding.

5.4. Data Abstraction

There are two levels of coding in the literature, inductive and deductive, as well as some back-and-forth between inductive and deductive. For example, Banscrum and colleagues' assessment of stalking statutes in the 50 U.S. states used grounded theory-oriented coding to

[15] Ethan C. Levine, 'Sexual Scripts and Criminal Statutes: Gender Restrictions, Spousal Allowances and Victim Accountability After Rape Law Reform.' (2018) 24 Violence Against Women.
[16] Michele Cascardi and others, 'School-Based Bullying and Teen Dating Violence Prevention Laws: Overlapping or Distinct?' (2018) 33 Journal of Interpersonal Violence.

construct a coding scheme—particularly, by inductively creating codes (e.g., 1 = "Some Characteristic; 0 = "Absence of Some Characteristic"; open coding) and then grouping such codes together into broader categories (axial coding).[17] In terms of deductive approaches to coding, Rebbe used questions from a national-level survey that featured neglect questions, applying the questions to each neglect statute and producing quantitative results.[18] "Emergent" categorization of text was used in Flash and Colleagues' work, such that coding categories were developed through previous literature (deductive) and the study documents themselves (inductive), simultaneously.[19] Other research has gone back-and-forth between inductive and deductive coding, particularly remaining open to new codes while simultaneously closed-coding, respectively.[20]

5.5. Ensuring Trustworthiness

One technique to fortify trustworthiness in multi-jurisdictional statute research is to embrace the use of multiple coders/reviewers/raters. Indeed, there does not seem to be a concrete rule for the number of coders that optimally fosters reliable coding. However, there seems to be an extent of agreement that at least two coders are necessary. There are two ways in which coders are employed. First, coders can discuss and resolve discrepancies and reach 100

[17] Caralin Branscrum and others, 'Stalking State Statutes: A Critical Content Analysis and Reflection on Social Science Research (2021) 31 Women & Criminal Justice.
[18] Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment.
[19] Paulina Flasch and others, 'State Standards for Batterer Intervention Programs: A Content Analysis' (2021) 36 Violence and Victims 683.
[20] Ethan Levine, 'Sexual Scripts and Criminal Statutes: Gender Restrictions, Spousal Allowances, and Victim Accountability after Rape Law Reform (2018) 24 Violence Against Women; Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment.

percent agreement after independently coding the text of the statues and comparing codes.[21] Second, a third-party coder may be recruited to break stalemates and resolve discrepancies between other independent coders. For example, a work on batterer intervention laws used two independent coders; after codes were compared and discrepant codes identified, all discrepancies were sent to a third independent coder to decide on the finalized codes per discrepancy.[22]

There are some techniques for understanding statistical measures for intercoder reliability. The first is the percentage of codes that are in agreement between two or more coders. The second is Krippendorf's alpha,[23] a measure of reliability in content analysis.[24] Cohen's kappa can also be used.[25] Other usages of statistical techniques include coding subsamples of the data (e.g., statutes) before final codes and numbers are produced.[26] One study used a random number generator to excavate the subsample for these pilot coding procedures.[27]

5.6. Enhancing Rigor

[21] Jennifer Swanberg, Mamta Ojha, and Caroline Macke, 'State Employment Protection Statutes for Victims of Domestic Violence: Public Policy's Response to Domestic Violence as an Employment Matter' (2012) 27 Journal of Interpersonal Violence.
[22] Paulina Flasch and others, 'State Standards for Batterer Intervention Programs: A Content Analysis' (2021) 36 Violence and Victims.
[23] Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment 303; Hannah Rochford and others, 'United States' Teen Dating Violence Policies: Summary of Policy Element Variation' (2022) 43 Journal of Public Health Policy.
[24] Klaus Krippendorff, 'Measuring the Reliability of Qualitative Text Analysis Data' (2004) 38 Quality & Quantity.
[25] Michele Cascardi and others, 'School-Based Bullying and Teen Dating Violence Prevention Laws: Overlapping or Distinct?' (2018) 33 Journal of Interpersonal Violence.
[26] Caralin Branscrum and others, 'Stalking State Statutes: A Critical Content Analysis and Reflection on Social Science Research. [2021] Women & Criminal Justice; see also Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment.
[27] Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment.

Triangulation is essentially mixed methodology, in which more than one method is used to either (a) see if findings converge (i.e., convergence), (b) see if findings diverge (i.e., divergence), and (c) see if findings work together to create a broader story (i.e., complementarity).[28] These can be integrated into policy analysis in different ways. Convergence and divergence can simultaneously be assessed through the inclusion of other data in addition to statutes and related legal mechanisms.[29] For example, Swanberg and Colleagues communicated with domestic violence agencies as a way to "cross-reference" and ensure the accuracy of the initial search for statutes (e.g., via databases, jurisdictions).[30] Complementarity can be integrated into the research by giving each data type a specific division of labor. For example, Crisafi's work triangulated statutes, court cases, and news reports to stitch together a story about race, gender, and the implications of stand-your-ground laws for intimate partner violence survivors.[31] The statutes, court cases, and newspaper articles each had a scaffolded role to play in shaping the findings of the work.

5.7. Presentation

There are several ways in which policy surveillance results are presented. The first involves presenting findings as a narrative, but without excerpts from the actual laws. The second is to

[28] David L. Morgan, 'Commentary—After Triangulation, What Next? (2019) 13 Journal of Mixed Methods Research.
[29] Ethan Levine, 'Sexual Scripts and Criminal Statutes: Gender Restrictions, Spousal Allowances, and Victim Accountability after Rape Law Reform (2018) 24 Violence Against Women; Rebecca Rebbe, 'What is Neglect? State Legal Definitions in the United States' (2018) 23 Child Maltreatment; Jennifer Swanberg, Mamta Ojha, and Caroline Macke, 'State Employment Protection Statutes for Victims of Domestic Violence: Public Policy's Response to Domestic Violence as an Employment Matter' (2012) 27 Journal of Interpersonal Violence.
[30] Jennifer Swanberg, Mamta Ojha, and Caroline Macke, 'State Employment Protection Statutes for Victims of Domestic Violence: Public Policy's Response to Domestic Violence as an Employment Matter' (2012) 27 Journal of Interpersonal Violence.
[31] Denise Crisafi, *No Ground to Stand Upon?: Exploring the Legal, Gender, and Racial Implications of Stand Your Ground Laws in Cases of Intimate Partner Violence* (doctoral dissertation, University of Central Florida 2016) <https://stars.library.ucf.edu/etd/4938/> accessed 9 April 2024.

present findings in paragraph format and, for example, block-quote excerpts of texts from the laws. The third presentation method and the most consistently used method involves using tables to parse out coding categories. Fourth, mathematical steps may be taken to understand the data such as creating broader indices that can be used to rank states based on a broad characteristic. Fifth, studies may use geospatial methods to map the distribution of law across geographies. The sixth means by which results may be presented involves graphical presentation, such as tracking the presence of some law type longitudinally.

6. A Call to Action

In light of the methodological exercises and options detailed above, a call to action on the matter of policy surveillance is necessary. Particularly, the following are needed:

- First, extant, reliable study guidance needs to be publicized beyond the field of public health. It seems that the most comprehensive, step-by-step guidance on policy surveillance is the Policy Surveillance Program: A LawAtlas Project and the Center for Public Health Law Research, both housed in Temple University's Beasley School of Law.[32] More effort is needed to help such comprehensive guidance escape the public health silo and permeate the boundaries of other fields, like criminology and criminal justice, among others. To make this possible, the current research proposes that strong collaborative and coalitional orientation between policy surveillance researchers and researchers in the field of evidence-based synthesis (e.g., PRISMA).

[32] Center for Public Health Law Research, 'Center for Public Health Law Research' (Beasley School of Law, Temple University) <https://www.phlr.org> accessed 21 April 2024.

- Second, the usage of differentiated search strategies may assist in further fortifying the rigor and trustworthiness of policy surveillance studies. For example, once a section of law (statute or administrative regulation) is identified, adjacency searching can be employed. This method of searching involves defining and searching a window of sections before and after an initially-identified section of law.[33] This additional step can help ensure that additional, relevant laws are also included in the sampling of laws beyond the initial search strategy.
- Third, studies may find relevance in venturing beyond analyzing the technical, enforceable aspects of statutes and regulations—particularly, by looking at discursive constructions of relevant topics. For example, Carson and Carter looked at abortion-related discourses in legislation across the U.S.[34] Learning about how things are said hints to the broader attitudinal climate in which such legislation is enacted and implemented. Indeed, how things are discussed may hint to how they are treated.[35] For example, the notorious "Ugly Laws" of Chicago, Illinois, U.S. used the terms "diseased, maimed, [and] mutilated" to refer to people with disabilities, enshrining a formal sanction to accompany stigmatizing language.[36] Fast-forwarding to the first decade of the 2000s, Rosa's Law was enacted in the U.S. This law imputed the term "mental retardation" with the term "intellectual disability" in various federal laws

[33] Julio Montanez and others, *Between Systems and Violence: State-Level Policy Targeting Intimate Partner Violence in Immigrant and Refugee Lives* (Routledge 2022).
[34] Saphronia Carson and Shannon K. Carter, 'Abortion as a Public Health Risk in COVID-19 Antiabortion Legislation' (2023) 48 Journal of Health Politics, Policy and Law.
[35] Barnett, Brian and Arron M. Bound, 'A Critical Discourse Analysis of *No Promo Homo* Policies in US Schools' (2015) 51 Educational Studies.
[36] Adrienne Phelps Coco, 'Diseased, Maimed, Mutilated: Categorizations of Disability and an Ugly Law in Late Nineteenth Century Chicago' (2010) 44 Journal of Social History.

purposed to increase accessibility. In these ways, discursive analyses can serve as a complementary component in policy surveillance.

- Fourth, the current call-to-action encourages the use of triangulation in data sources wherever possible. For concrete, technical research based on, for example, statutory and administrative law, this can take the form of examining court cases that cite such law.[37] For more discursive works, using non-legal data (e.g., newspaper text, qualitative interviews) could help as tests of multimethod convergence/divergence.[38]

7. Discussion

The current paper used systematized evidence-based syntheses to obtain a flavor of the policy surveillance literature, as well as how policy surveillance is conducted in the field of domestic violence research. Findings from the systematized reviews show that there are certain requirements (e.g., systematic searching, the need for lawyers)[39] that accompany the responsibility of conducting policy surveillance. Moreover, information from the systematized reviews shows that myriad methodological approaches (e.g., regarding the number of coders, coding discrepancy resolution) are used to build lists of statutes and administrative regulations. In these ways, there are many strategies (e.g., adjacency searching) that can be developed and used to increase rigor and ensure trustworthiness in such studies.

[37] Denise Crisafi, *No Ground to Stand Upon?: Exploring the Legal, Gender, and Racial Implications of Stand Your Ground Laws in Cases of Intimate Partner Violence* (doctoral dissertation, University of Central Florida 2016) <https://stars.library.ucf.edu/etd/4938/> accessed 9 April 2024.
[38] David L. Morgan, 'Commentary—After Triangulation, What Next? (2019) 13 Journal of Mixed Methods Research.
[39] Aila Hoss and others, 'Yes, You Need a Lawyer: Integrating Legal Epidemiology into Health Research' (2020) 135 Public Health Reports.

The current work is not without limitations. First, by employing systematized reviews of the extant literature, the current study does not meet the methodological threshold to claim that it is a more advanced type of review.[40] For example, the current study made use of one coder, the author. However, at the same time, goal was not to declare findings and generalize about the literature, but to galvanize a point of departure for developing unified protocols—across various fields—for searching, analyzing, and reporting on policy surveillance research. Moreover, the topical focus of one of the systematized reviews—the review on domestic violence policy surveillance—may not be perceived fully as standing up to the test of non-arbitrary-ness. However, the systematized review on domestic violence policy surveillance was based on the current study's author's major area of research interest (i.e., domestic violence). Accordingly, what is lost in not systematically identifying the topic of study, is consequently gained in the author's familiarity with and insight into domestic violence policy research.

The implications of the current work are manifold. First, while extant syntheses of knowledge exist on the matter,[41] the current work provides a piecemeal forward movement of the legal methodology literature by way of identifying specific components of policy surveillance (e.g., using random number generation for coding a subsample of statutes). Second, the current paper lays out the methodological components as a variety of potential options for policy surveillance researchers. Finally, its call-to-action attempts to galvanize the use legal-research

[40] Marjia J. Grant and Andrew Booth, 'A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies' (2009) 26 Health Information and Libraries Journal.
[41] Scott Burris and others, 'Policy Surveillance: A Vital Public Health Practice Comes of Age' (2016) 41 Journal of Health Politics, Policy, and Law.

coalitions, additional search strategies, discursive analyses, as well as triangulation. Overall, it is hoped that the current work provides a roadmap for publicizing and moving policy surveillance research forward.

8. Conclusion

Policy surveillance is an important means by which researchers can understand the impacts of public policy on the empirical world. Existing approaches on specifically "how" to conduct such research are diverse. This article explored these intricacies through the conducting of two systematized reviews, which provided a groundwork for a guide and call-to-action on the need for unified guidance on policy surveillance. First, extant unified guidance on policy surveillance may benefit policy studies more generally through greater reach into the literature. Second, integrating diverse strategies for establishing rigor and enhancing trustworthiness holds the promise of fortifying methodological strength across studies. Third, studies may benefit from also including a discursive focus in policy surveillance. Fourth, usage of multiple data sources can invoke the principles of convergence and complementarity to foster greater methodological strength within studies. In these ways, the field of policy surveillance can more easily disseminate and sharpen methodological techniques for understanding the relationship between law and the social world.

Journal of Legal Research Methodology (JLRM)
ISSN 2752-3403
Pina-Sánchez & Guilfoyle - 1661

https://doi.org/10.19164/jlrm.v4i1.1661

Published date: 20/10/2025

Article

Ethnic Disparities in Sentencing in England and Wales: Review of Recent Findings

Jose Pina-Sánchez
School of Law, University of Leeds
https://orcid.org/0000-0002-9416-6022

Eoin Guilfoyle
Brunel University of London
https://orcid.org/0000-0002-2497-7735

Abstract

Following the 2017 Lammy Review, research into ethnic disparities in sentencing in England and Wales has intensified. This article reviews the main findings from recent studies, focusing on the robustness of evidence, areas where disparities are most prevalent, gaps in the literature, and potential solutions.

Ethnic disparities are less severe and more offence-specific than previously reported. There are no substantial differences in custodial sentence length, while for the probability of receiving a custodial sentence, disparities are concentrated primarily among drug offences. However, such disparities cannot be fully explained by statistical bias, suggesting a degree of direct or indirect sentencing discrimination.

https://doi.org/10.19164/jlrm.v4i1.1661

Sentencing disparities appear consistent across most minority groups. However, intersectional analyses reveal nuanced patterns; for instance, white male offenders require over 50% longer criminal records than black male offenders before crossing the custody threshold, while no significant differences are observed between black and white female offenders. Notably, socioeconomic factors, such as area deprivation, do not seem directly linked to ethnic disparities, although deprivation independently influences sentencing outcomes.

Several gaps remain in the literature. Multivariate analyses focused on magistrates' courts, where most sentences are imposed, are lacking. Qualitative research is also needed to explore disparities in areas like drug offences, male ethnic minority offenders, and assessments of mitigating factors.

Current efforts to mitigate disparities should be expanded to include more structural solutions, such as increasing funding for legal aid, improving the quality of pre-sentence reports, and ensuring community services for addiction, mental health, and employment are universally accessible.

https://doi.org/10.19164/jlrm.v4i1.1661

Article Text

1. Background

In a recent survey of 373 legal professionals[1], 56% reported having witnessed one or more judges exhibit racial bias towards a defendant (Monteith et al., 2022). These subjective perceptions are congruent with the scientific evidence most widely shared. The Lammy Review (2017) documented ethnic disparities throughout the criminal justice system. In relation to sentencing, they referred to a study conducted by the Ministry of Justice (Hopkins, 2016), highlighting how the odds of receiving a custodial sentence are 240% higher for ethnic minority offenders compared to white offenders charged with the same drug offence and equivalent criminal histories and propensities to plead guilty.

Odds ratios are notoriously difficult to interpret, but a 240% disparity is simply huge. To put this in context, and assuming a baseline probability of custody of 0.50 for white offenders, 240 higher odds for ethnic minority offenders would represent a probability of custody close to 0.77, i.e., a custody rate differential of 27 percentage points. This shocking data point has been extremely influential in signifying the presence of vast ethnic disparities in sentencing in England and Wales. Specifically, the above odds ratio has been referred to in almost every discussion on the topic following the Lammy Review. See for example responses to the Lammy Review (Neilson, 2017), the House of Commons debates (2021) on the Police, Crime, Sentencing and Courts Bill, as well as other publications on ethnic disparities in England and Wales (Institute of Race Relations, 2024; Clinks, 2020).

[1] This survey was based on a combination of convenience sampling (the questionnaire was distributed to a variety of organisations and individuals within the legal profession) and voluntary response sampling (the questionnaire was also hosted on the research project's website and advertised via social media).

https://doi.org/10.19164/jlrm.v4i1.1661

Understandably, criminal justice agencies have been pressed to respond to this problem[2], while perceptions of discrimination have widened amongst the ethnic minority population. In reference to the 2015 Crime Survey for England and Wales, the Lammy Review indicated how 51% of the ethnic minority population believe 'the Criminal Justice system discriminates against particular groups and individuals. According to the 2023 version of the crime survey, that figure has risen to 65%.

The sense of urgency is clear. How to respond to the problem, however, is not. As correctly identified by the Lammy Review, the evidence base documenting the presence of sentencing disparities in England and Wales was practically non-existent. Ironically, to a great extent, the dearth of evidence available to criminal justice agencies to provide effective responses to the crisis stemmed from their own inability to share their data with researchers outside their organisations.[3] With the exception of Hopkins (2016), all we knew relied on Ministry of Justice aggregate statistics (Roberts & Bild, 2021), limiting the types of analyses that could be carried to explore the causal mechanisms behind the observed disparities, and to assess the robustness of the available evidence.[4]

[2] The Lammy Review advocated for the introduction of an 'explain or reform' principle i.e. if an evidence-based explanation for apparent disparities between ethnic groups cannot be provided, then reforms should be introduced to address any observed disparities.
[3] Following the publication of the Lammy Review the first author of this article submitted separate data access applications to the Judicial Office, the Ministry of Justice, HM Courts and Tribunals Service, and the Sentencing Council for England and Wales. All were either rejected or ignored.
[4] Notably, aggregate data cannot be used to conduct the type of multivariate analysis that allow conditioning on legal factors such as guilty plea or previous convictions. This type of analysis is key to approximate estimations of unwarranted disparities, and it is only possible through individual-level data.

https://doi.org/10.19164/jlrm.v4i1.1661

This impasse was overcome in 2021 through two momentous developments: i) the publication of the Sentencing Council for England and Wales first report on sentencing disparities, equality in their guidelines, and, subsequently, some of their underlying data; and ii) the release of the first magistrates' and Crown Court datasets from the Data First project.[5] A new wave of studies have followed the release of these official datasets, providing a fresh and more complete perspective about the prevalence and origin of ethnic disparities in England and Wales. Here, we review this nascent body of research and in so doing evaluate the robustness of the evidence base, highlight particularly problematic areas, identify gaps that should be addressed by future research, and speculate about the relative effectiveness of different policies that could be enacted to minimise the problem.

2. Setting the Record Right

Before reviewing the latest findings from the literature, it is crucial to address a common misconception. The claim of 240% higher odds of incarceration for ethnic minority offenders, cited in the Lammy Review and echoed in subsequent publications and debates, is factually incorrect. This figure does not appear in the original study referenced by the Lammy Review (Hopkins, 2016). Instead, the study reported an odds ratio of 2.37 for ethnic disparities among drug offenders - approximately 2.4 for simplicity - which equates to a 140% increase in odds, not 240%.

[5] Data First was created as a collaboration between the Office for National Statistics, the Ministry of Justice, and Administrative Data Research UK, tasked with making administrative government data available to accredited researchers through secure data access protocols. The first datasets released covered all sentences imposed since 2016 for the Crown Court, and 2011 from the magistrates' courts, capturing key case characteristics (such as the most serious offence, whether the defendant was placed in remand, or whether they plead guilty), but also important offender characteristics, crucially amongst them, their ethnic background.

https://doi.org/10.19164/jlrm.v4i1.1661

While these disparities are deeply concerning and merit further investigation, it is essential to recognise that the actual magnitude of the reported disparities is nearly half of what has been claimed. This error is likely an innocent mistake; however, we are less sympathetic to how this evidence has been framed in subsequent reports and official discussions. Most repeat the 240% higher odds for ethnic minority drug offenders but fail to mention that the same study found no significant disparities in sentencing for sex or violent offences. This omission appears to be an example of selective reporting, which is more problematic as it reflects a deliberate choice.

Accurate interpretation of these findings is vital for framing the issue correctly and guiding effective responses. Without delving further into our review, one of our key conclusions is already apparent: ethnic disparities in sentencing are not as profound, nor as widespread, as is often assumed. Furthermore, the original study referenced by the Lammy Review has notable limitations; most significantly, it does not adequately approximate 'like with like' comparisons, raising questions about whether the reported disparities are truly unwarranted.

3. Warranted or Unwarranted Disparities?

Disparities in sentencing do not necessarily indicate judicial prejudice. Variations in guilty plea rates or the types of offences committed by different ethnic groups can lead to disparities, which we might classify as warranted. To identify unwarranted disparities, researchers rely on statistical models that account for relevant legal factors. However, this process is challenging because many legal factors influencing sentence severity are either not recorded with sufficient precision or not recorded at all (Baumer, 2013; Halevy, 1995).

For example, while Hopkins (2016) controlled for guilty plea and offence category, the analysis overlooked key factors outlined in sentencing guidelines, such as harm, culpability, and aggravating or mitigating circumstances. Additionally, the legal factors that were considered were addressed only imprecisely. Critical details - like the timing of the guilty plea or specific offence types within broader offence categories - were excluded, further limiting the validity of her findings.

The Sentencing Council for England Wales set out to replicate the analysis in Hopkins (2016) using their own survey data, which was originally collected to assess the impact of their sentencing guidelines. Crucially, this dataset includes most factors listed in the sentencing guidelines, such as harm (e.g., the type and quantity of drugs supplied), culpability (e.g., the offender's role), and other aggravating or mitigating factors (e.g., whether the offence was committed while on license or whether the offender showed remorse).

As far as we know, the resulting study by Isaac (2021) is unmatched in terms of the number of legal factors controlled for; including harm, culpability, aggravating, and mitigating factors. Nonetheless, the study found that black offenders had 40% higher odds of receiving a custodial sentence compared to equivalent white offenders. While this disparity is much smaller than that reported by Hopkins (2016), it remains significant. For instance, assuming a custody rate of 50% for white offenders, this translates to a custody rate of 58.3% for black offenders charged with the same crime, criminal history, and personal circumstances.

However, the Sentencing Council cautioned that these findings might still reflect unobserved factors, such as recommendations recorded in pre-sentence reports, which could not be

controlled for. This leaves us with the question: are these unwarranted disparities or not? At present, it seems that the data can be interpreted to support differing positions. Unfortunately, open-ended conclusions on politically sensitive and complex issues like this tend to be prone to confirmation bias, leading readers to default to their pre-existing beliefs. To move beyond this impasse, we need to shed more light on the problem by delving deeper and adopting a more technical approach.

It is true that the Council's survey does not capture all potentially relevant factors considered by the judge. As pointed by Isaac (2021), many of these are recorded in the pre-sentence reports, but are not recorded in the Council's survey, e.g. risk assessment, or potential for rehabilitation. Furthermore, whether a pre-sentence report is available or not is itself a relevant factor, as is the quality of the pre-sentence report. Both are expected to affect sentence severity, and both are potentially unevenly distributed across ethnic groups (HM Inspectorate of Probations, 2021). Besides legal factors considered in pre-sentence reports, there are other relevant factors that were not captured in the survey for reasons of confidentiality, most importantly, whether the offender assisted with the prosecution of other cases. Lastly, some key factors are unduly oversimplified, the best example being the number of previous convictions, which is interval-censored.

However, failing to control for all relevant legal factors is not a sufficient condition to claim that the reported disparities are due to quantitative bias. For such a claim to hold, these unobserved factors must not only exist but also be disproportionately distributed across

ethnic groups and remain unexplained by the factors already controlled for in the model. In our view, both conditions are unlikely.

For instance, while the Sentencing Council study could not directly account for factors like the offender's dangerousness or rehabilitation potential (e.g., through the offender's risk assessment score in the pre-sentence report), it did control for many of their constitutive elements. These include the offender's criminal history, whether they targeted a vulnerable victim, their role in the offence, and whether the judge believes that the offender is addressing or has the potential to address their addiction/offending behaviour. As a result, aspects of dangerousness or rehabilitation potential are indirectly accounted for.[6]

Even if we disregard that point entirely, for the observed disparities to be nothing more than statistical bias we would still need those missing factors to be unevenly distributed across ethnic groups. To quantify the likelihood of such scenarios, Pina-Sánchez et al. (2023) conducted an analysis based on simulated data. Since the original dataset was unavailable, the study recreated a dataset approximating the key statistical properties of the one used by the Sentencing Council, as reported in Isaac (2021). These properties included metrics such as the custody rate for white and black offenders and the estimated effect of ethnicity on custody probability. It then tested for hypothetical factors strong enough to account for the observed disparities. The findings indicate that for a missing factor to eliminate the reported disparities,

[6] To further illustrate, Pina-Sánchez et al. (2024a) demonstrated that only a few major legal factors significantly influence sentence severity. Specifically, in custody decisions for shoplifting offences, sentence severity could be predicted with 80.7% accuracy using only the top 10 most important guideline factors. Expanding to the top 20 factors only marginally increased predictive accuracy to 81.2%. For context, Isaac (2021) controlled for over 30 legal factors in their analysis.

https://doi.org/10.19164/jlrm.v4i1.1661

it would need to meet a set of stringent conditions. For example, in one of the four scenarios identified, the unobserved legal factors would need to: i) increase the probability of custody by at least 10%; ii) be present in at least 70% of the reference group (white offenders); and iii) be at least 20% more prevalent in the treatment group (i.e., present in at least 84% of Black offenders). While we cannot entirely rule out the existence of such a factor, it is highly improbable.[7] Therefore, and despite the complexity of the question[8], we conclude that the reported ethnic disparities among drug offenders are unwarranted.

Acknowledging this reality means we cannot ignore the problem. Even if sentencing disparities are highly localised and smaller in size than previously considered, they still appear to indicate a violation of the principle of equality under the law – a breach of a fundamental expectation in liberal democracies that simply cannot be accepted. In Section 5, we propose potential solutions. However, to implement them effectively, it is critical to understand the exact sources of these disparities. In the following section, we examine where these disparities are most pronounced, aiming to uncover their causal mechanisms and better inform the design of targeted and effective interventions.

[7] For example, Isaac (2021) did not control for whether the defendant was placed on remand, which could be taken as a proxy for offenders' dangerousness. However, according to the latest statistics on ethnicity and the criminal justice (Ministry of Justice, 2024), the remand rate in the Crown Court is 53%, and in the most extreme comparison only 13.5% more prevalent in black than white offenders. Hence, it is not possible that having failed to control for remand is on its own exerting a strong enough bias to explain away the estimated ethnic disparities in custody.

[8] It is worth nothing that we have not considered further assumptions commonly violated that are likely affecting the validity of our findings, such as when missing data is not missing at random (Stockton et al., 2024), or the fact that most studies misclassify the reference group in their measures of ethnicity by adding ethnic minority categories such as gypsy travellers in the UK, or white Hispanics in the US.

https://doi.org/10.19164/jlrm.v4i1.1661

4. Where Are Disparities Coming From?

As noted, Hopkins (2016) only found disparities in one of three offence groups considered. Using more recent data covering sentences imposed in the Crown Court from 2018 to 2020, Pina-Sánchez et al. (2025) corroborated that finding. Specifically, it explored eight offence groups but only found evidence of substantive disparities amongst drug offenders. In that category, and after conditioning on guilty plea, previous convictions, age and gender, ethnic minority offenders have an average 0.65 probability of custody compared to 0.58 for white offenders. The second largest disparities were for assault, where the study observed only a three-percentage point difference against ethnic minority offenders. For all other six offence groups considered the disparities were even smaller, failing to meet the threshold of statistical significance.

These findings reinforce the idea that ethnic disparities in sentencing in England and Wales are not widespread but heavily concentrated amongst drug offenders, and consequently, had we focused our analysis on that offence group we might have contributed to perpetuate a view of widespread disparities.

That same study also set out to test whether ethnic disparities in sentencing might in fact be reflecting class disparities. It did not find evidence to support that claim, the estimates of ethnic disparities were not affected by the level of deprivation of the neighbourhood of residence of the offender. Nonetheless, it did find important disparities in account of that factor. For instance, after controlling for offence and offender characteristics, it estimates that the probability of receiving a custodial sentence for a breach offence in the Crown Court is

https://doi.org/10.19164/jlrm.v4i1.1661

0.51 for offenders from the top 10% of the most affluent neighbourhoods, compared to 0.63 for those from the bottom 10%. That is, class disparities appear to be as significant as ethnic disparities. However, they appear to be independent of each other since the former does not explain the latter.

In another study using the new Crown Court administrative datasets, Lymperopoulou (2024) explored disparities across a wider range of ethnic minority groups (fifteen in total). Her findings reveal relatively consistent ethnic disparities, with no single ethnic minority group being markedly worse off. The only notable exceptions are white Irish and white-Asian mixed offenders, who do not appear to be sentenced differently from their white British counterparts. However, adopting an intersectional approach to enquire the same dataset, Sorsby (2023) noted that ethnic disparities can vary substantially by gender. Specifically, she found that white male offenders are allowed over 50% longer criminal records than black male offenders before being sentenced to custody, whereas no significant differences were detected between black and white female offenders. This is an interesting finding that helps us identify male offenders as a trait where ethnic disparities are most prevalent.

It is worth noting that all the disparities reported so far relate to differences in the probability of receiving a custodial sentence. When we consider differences in custodial sentence length we find much scarcer evidence of ethnic disparities. For example, in her study of drug offenders, Isaac (2021) found that the average sentence length after controlling for legal factors listed in the sentencing guidelines was only 4% longer for Asian than for white offenders, and not significantly different for offenders from a black or other ethnic

background. Similarly, Lymperopoulou (2024), found no significant disparities in sentence length across most ethnic minority groups after controlling for offence characteristics. The main exceptions being Bangladeshi, and other white offenders, who receive 11.6% longer, and 14.8% shorter sentences than white British respectively.

To understand the potential mechanisms behind the observed disparities we should also scrutinise disparities in the legal factors that determine the final sentence. The latest Ministry of Justice (2024) 'Statistics on Ethnicity and the Criminal Justice System' report, indicates how disparities in guilty plea rates remain relatively wide. For example, in 2022, the guilty plea rate in the Crown Court is 68% for white offenders, reaching only 60% amongst Asian offenders. All studies mentioned so far in this review condition on guilty plea, however, they do not always capture the timing of the guilty plea, or the specific reduction allowed by the judge, which might be explaining some of the observed disparities. The same report showed the percentage of sentences corresponding with pre-sentence recommendations of custody appeared to be equally distributed across ethnic groups. However, that report did not indicate whether custody was a more common recommendation for one ethnic group over another nor the coverage of pre-sentence reports across ethnic groups. A recent inspection by HM Inspectorate of Probation (2021) found that the quality of pre-sentence reports on ethnic minority individuals were insufficient in 21 of the 51 reports inspected, with not enough consideration of the service user's diversity, concluding: '*poorer quality reports that fail to consider all relevant factors run the risk of service users receiving more punitive sentences*' (p29).

https://doi.org/10.19164/jlrm.v4i1.1661

Relying on Sentencing Council data obtained through a 'freedom of information' request, Guilfoyle and Pina-Sánchez (2024) showed how most factors listed in the sentencing guidelines are markedly uniformly distributed across ethnic groups. The main exception being personal mitigating factors. For example, considering drug offences, white offenders where 45% more likely to be found of good character and 47% more likely to be deemed remorseful than black offenders. Therefore, it is possible that some of the observed disparities noted in the literature might be stemming from upstream decisions relating to how the case was constructed by the probation officer and the magistrate/judge.

In summary, reviewing the latest evidence on the subject we have learnt that ethnic disparities are mostly present in: i) male offenders; ii) drug offences; iii) decisions of custody and; iv) personal mitigating factors. Similarly, ethnic disparities *do not* appear to vary substantially across ethnic minority offenders, be the result of deprivation related disparities, or be present amongst: i) female offenders; ii) non-drug related offences; iii) decisions of sentence length, iv) objective sentencing factors listed in the guidelines. In the next and final section, we reconcile this new information with known gaps in our understanding to review ongoing initiatives seeking to redress the problem and additional ones that could be cautiously conceived.

5. Policy Solutions

Ethnic disparities in sentencing are not as widespread as previously thought. A series of new studies made possible by the release of sentencing datasets from the Ministry of Justice and

the Sentencing Council have shown how unwarranted disparities are far smaller in size, and heavily concentrated around drug offences.

Despite their narrower scope, these disparities should be interpreted as genuinely unwarranted as it is unlikely they are simply the result of differences in legal characteristics defining criminal cases committed by different ethnic groups. As such, all of the ongoing initiatives seeking to redress them, such as those carried out so far by the Judicial College (unconscious bias training) and the Sentencing Council (reminders in sentencing guidelines to consult the equal treatment handbook, amending the expanded explanations in sentencing guidelines for certain mitigating factors, as well as the inclusion of a new mitigating factors of 'difficult and/or deprived background or personal circumstances')[9], are well justified. Recent findings might be used to further enhance some of these efforts. For example, the reminder/note in drug offence sentencing guidelines highlighting there is evidence of ethnic disparities in sentencing outcomes could be updated to specifically state that the evidence of ethnic disparities relates to decisions of custody for male offenders being sentencing for drug offences. This additional specificity might help focus attention on cases where the research evidence has shown disparities are mostly likely to arise.

In addition, based on recent findings, we argue there is scope to go even further and broaden these efforts. Increasing the availability and quality of pre-sentence reports is one policy option that could significantly mitigate disparities. Ensuring the availability of comprehensive

[9] Many of these recent reforms stem from a review commissioned by the Sentencing Council to assess equality and diversity within its work (Chen et al., 2022). The final report presented a range of recommendations, a significant number of which were accepted and acted upon by the Sentencing Council.

https://doi.org/10.19164/jlrm.v4i1.1661

pre-sentence reports would allow judges to consider personal mitigating factors more thoughtfully. It would also allow judges to better consider and impose non-custodial sanctions for ethnic minority offenders. As recent findings have shown, ethnic disparities likely stem, at least in part, from the judicial assessment of personal mitigating factors and disparities are mostly present in decisions of custody rather than sentence length. It is the decision regarding custody where pre-sentence reports are most influential. As well as improving the availability and quality of pre-sentence reports, it is also important that suitable programmes and services are available in the community for ethnic minority groups. Research has shown that ethnic minority groups face additional barriers in accessing drug and alcohol treatment (Fountain 2009). This can include language barriers, stigma and cultural differences, mistrust of mainstream services and experiences of racial prejudice. These barriers can prevent ethnic minority groups from engaging with treatment services which in turn can impact risk assessments and rehabilitative potential.

In 2021, HM Inspectorate of Probation found that there were very few treatment programmes available specifically for ethnic minority users (p29). It found there was no coherent national approach to assess the differing needs/barriers for ethnic minority users, to identify any disproportionality in service delivery or to improve how services are delivered to these groups (p8). The inspectorate recommended the development of a national race equality strategy for probation service delivery, supplemented by strategic needs assessments in each probation region, to ensure that ethnic minority service users are not disadvantaged (p12). At the time of writing, a national race equality strategy for probation service delivery has yet to be published. It is unclear if one has been developed. HM Inspectorate of Probation in a follow

up review in 2023 stated that race equality in probation remains 'a work in progress' (2023). Developing a national race equality strategy addressing the issues identified by HM Inspectorate of Probation, in our view is a policy option worth pursuing.

Another policy option would be to restore legal aid to their pre-austerity levels. Improving legal aid access could help enhance ethnic minority defendants' trust in the justice system, increase guilty pleas where appropriate, and address potential issues of courtroom demeanour. Currently, there remains a relatively wide disparity in guilty pleas across ethnic groups (Ministry of Justice, 2024). Additionally, the expanded explanations in sentencing guidelines could be used to warn about potential biases in attributing 'good character' or 'remorse' to ethnic minority offenders, with similar reminders given to probation officers writing pre-sentence reports. Doing so could help to reduce the uneven distribution of these impactful mitigating factors across ethnic groups which was observed by Guilfoyle and Pina-Sánchez (2024) and identified as likely sources of ethnic disparities in sentencing outcomes.

5.1. Imposition of Community and Custodial Sentences

Since the pre-print of this article, the Sentencing Council has published a revised Sentencing Guideline on the '*Imposition of Community and Custodial Sentences*', set to take effect on 1 April 2025 (Sentencing Council 2025). One of the objectives of the revised guideline is to increase the use of pre-sentence reports as a mechanism for reducing disparities. To achieve this, the guideline introduces a new section specifically listing cohorts of offenders for whom a pre-sentence report should normally be considered necessary. This non-exhaustive list

includes offenders from an *'ethnic minority, cultural minority, and/or faith minority community'.*

The revised guideline has been met with strong political opposition. Shadow Justice Secretary Robert Jenrick branded it '*two-tier justice'*. In response, Justice Secretary Shabana Mahmood stated her intention to write to the Sentencing Council to '*register [her] displeasure'* and recommend reversing the change (BBC, 2025).

Addressing disparities through an increased use of pre-sentence reports broadly aligns with our proposed policy solutions. Ideally, high-quality pre-sentence reports would be available for all offenders, ensuring judicial decisions are informed by comprehensive assessments. However, given current state of the criminal justice system, it could be argued that this is an unrealistic *short-term* goal. In this context, prioritising pre-sentence reports for specific cohorts - such as ethnic and cultural minorities - has merit, as it equips judges with critical insights into cultural backgrounds, structural disadvantages, and potential barriers to accessing rehabilitative services, and in doing so, should help to reduce ethnic disparities in sentencing outcomes. However, if this is the approach taken, then in our view, it is important for 'deprived backgrounds' to also be specified in the guideline as a cohort for whom a pre-sentence report should normally be requested. This would address criticisms that the guideline is seeking to reduce some documented disparities (ethnic/cultural) but not others (social class), while it will also contribute to redress the equally important deprivation-related disparities that have been recently detected (Pina-Sánchez et al., 2025).

https://doi.org/10.19164/jlrm.v4i1.1661

Furthermore, it would be beneficial to explicitly frame the approach the Sentencing Council is taking in the revised guideline as a necessary short-term measure, with the long-term goal remaining the universal provision of high-quality pre-sentence reports. Clarifying this ambition would go some way towards counteracting perceptions of unfairness and reinforce the principle that sentencing should be based on a fully informed assessment of each individual case.

6. Research Gaps and a Plea for Open Data

Despite these insights, evidence remains incomplete. Most notably, no multivariate analyses have addressed disparities in magistrates' courts, where over 90% of sentences are imposed. This is a major gap in our current knowledge and understanding of ethnic disparities in sentencing. Differences in offence types and sentencing procedures between the Crown Court and magistrates' courts caution against generalising Crown Court findings.

Moreover, some of the explanations for disparities we have put forward are tenuous. Hypotheses linking the availability or quality of pre-sentence reports to disadvantages for ethnic minority in risk assessments and in evaluations of rehabilitative potential ought to be tested empirically as should the potential that courtroom demeanour might be impacting the assessment of personal mitigating factors. Additional further insights into the reasons behind the large disparities in guilty pleas would be of benefit too. It could also be possible that we are being completely misled and the reason behind the observed disparities is simply that some judicial decisions are consciously or subconsciously biased. For example, recent research carried out by Pina-Sánchez and Lewis for the Crown Prosecution Service (CPS, 2023),

https://doi.org/10.19164/jlrm.v4i1.1661

identified regional disparities in charge rates, with rural areas displaying greater ethnic biases compared to urban centres like London or Manchester. Similar insights are needed for sentencing disparities. If similar dynamics exist in sentencing, then diversifying the judiciary's ethnic composition should be further encouraged (Veiga et al., 2022).

Intersectional dimensions also demand attention. For instance, Sorsby (2023) found ethnic disparities specific to male offenders, while Pina-Sánchez et al. (2025) highlighted deprivation-related disparities alongside ethnic ones. Additionally, U.S.-based analyses (Pina-Sánchez & Tura, 2024c) suggest factors like nationality and education level may further exacerbate disparities. Addressing these issues will require a combination of qualitative research and better use of existing datasets.

Many of these questions such as evaluations of demeanour in court require qualitative designs, which inevitably will be costly and take a long time to be carried out. They also require the cooperation of the Judicial Office and members of the judiciary. Other questions like those related to offender's nationality, level of education, or legal aid, represent factual information that could be easily incorporated to future surveys of the Sentencing Council or administrative records from the Ministry of Justice, making them available in the near future at no added substantial cost. While a third group of key questions, such as estimating disparities in the magistrates' courts, the extent to which disparities are concentrated in certain court locations, or - to some degree - the effect of sentencers failing to receive a pre-sentence report, could be explored with data already available.

https://doi.org/10.19164/jlrm.v4i1.1661

A key reason we lack answers to many of these questions, despite the availability of relevant data, lies in the numerous hurdles researchers face in accessing and using it. These include: (i) lengthy application processes; (ii) mandatory research accreditation requiring a data security course and examination; (iii) limited access to IT infrastructure, available only at a few selected universities; (iv) reliance on slow servers for analysis; (v) re-application requirements to address research questions not outlined in the original plan; and most frustratingly, (vi) publication clearance from three separate gatekeepers - Office for National Statistics, Ministry of Justice, and Judicial Office - causing unnecessary delays and uncertainty.

To advance our understanding of fairness in sentencing, these stringent data security protocols must be reconsidered. Arguments for maintaining them typically cite confidentiality concerns, but we do not think these are valid. Sentencing occurs publicly, and information about legal and extra-legal factors influencing decisions, as well as the sentencing outcomes themselves, should be treated as public domain. In fact, sentencing's public nature is a feature meant to communicate the consequences of wrongdoing transparently. Other jurisdictions demonstrate that greater transparency is feasible. U.S. Federal Courts and states like Minnesota, Florida, and Pennsylvania have openly shared detailed sentencing data for decades, often including information about the judges involved. More recently, countries like China, the Czech Republic, Slovakia, France, Germany, Russia, Poland, the Netherlands and Finland have adopted similar practices[10]. England and Wales should not remain an exception.

[10] The availability of open individual-level sentencing data across European countries was recently discussed in an online event hosted by the Empirical Research on Sentencing Network. Slides from national correspondents can be accessed here: https://empiricalresearchonsentencing.wordpress.com/presentations/ and the full recording of the meeting is available here: https://owncloud.cesnet.cz/index.php/s/jRyF1RHARSSay2P.

https://doi.org/10.19164/jlrm.v4i1.1661

The current barriers severely limit the quantity of researchers equipped to work with these datasets and the quality of their work. In practice, only those with sufficient funding to dedicate their time exclusively to navigating these obstacles have used the data. This not only slows the research process but undermines its integrity, as replication of analyses becomes prohibitively time-consuming or practically infeasible.

The case for facilitating access to sentencing data cannot be overstated. Better data access will lead to a fairer sentencing process by enabling deeper understanding of the disparities and more effective policies to address them. Drawing on our three years of collaboration with the Sentencing Council, Judicial Office, magistrates, and Crown Court, we have observed a genuine commitment among stakeholders to tackle unwarranted disparities. All of them have shown a keen interest in learning about the latest findings on the subject to implement solutions to tackle the issue effectively. We are therefore convinced that by identifying the specific mechanisms through which disparities come to be, we will be able to redress them.

Acknowledgements/Funding

This work was supported by the Secondary Data Analysis Initiative of the Economic and Social Research Council (Grant Ref: ES/W00738X/1). We also thank participants of 'Ethnic Disparities in Sentencing in England and Wales: Taking Stock and Moving Forward', a symposium held in November 2024 at the LSE, which inspired this review.

References

Baumer, E. P. (2013). Reassessing and redirecting research on race and sentencing. *Justice Quarterly* 30(2):231–261. URL: https://www.tandfonline.com/doi/full/10.1080/07418825.2012.682602

BBC (2025) Call for ministers to overrule 'two-tier' sentencing changes. URL: https://www.bbc.co.uk/news/articles/cvg19gx7vl4o

Chen, Q., Vuk, M., Kuppuswamy, C., & Kirsch, D. (2023). Equality and diversity in the work of the Sentencing Council. *Sentencing Council for England and Wales.* URL: https://www.sentencingcouncil.org.uk/wp-content/uploads/Equality-and-Diversity-Report-FINAL.pdf

Clinks (2020) Future Sentencing Policy – Entrenching Inequality. URL: https://www.clinks.org/community/blog-posts/future-sentencing-policy-entrenching-inequality

Crown Prosecution Service (2023). CPS charging decisions - examining demographic disparities in the outcomes of our decision making. URL: https://www.cps.gov.uk/publication/cps-charging-decisions-examining-demographic-disparities-outcomes-our-decision-making

Guilfoyle, E., & Pina-Sánchez, J. (2024). Racially determined case characteristics: Exploring disparities in the use of sentencing guideline factors in England and Wales. *The British Journal of Criminology.* URL: https://academic.oup.com/bjc/advance-article/doi/10.1093/bjc/azae039/7708313

Halevy, T. (1995). Racial discrimination in sentencing: a study with dubious conclusions. *Criminal Law Review*, pp. 267–271.

HM Inspectorate of Probation (2021). Race Equality in Probation: The Experiences of Black, Asian and Minority Ethnic Probation Service Users and Staff: A Thematic Inspection by HM Inspectorate of Probation. *HM Inspectorate of Probation.* URL: https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2021/03/Race-Equality-in-Probation-thematic-inspection-report-v1.0.pdf

HM Inspectorate of Probation (2023). Race equality in probation follow-up: A work in progress, HM Inspectorate of Probations.URL: https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2023/09/Race-equality-in-probation-follow-up-thematic-inspection-v1.0.pdf

Hopkins K., Uhrig, N., & Colahan, M. (2016). Associations between ethnic background and being sentenced to prison in the Crown Court in England and Wales in 2015. *Ministry of Justice.* URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/639261/bame-disproportionality-in-the-cjs.pdf

House of Commons Debate (2021). Police, Crime, Sentencing and Courts Bill: Twelfth Sitting. *Hansard.* URL: https://hansard.parliament.uk/Commons/2021-06-10/debates/fc94f669-16a7-4e5d-94b0-eea72d258033/PoliceCrimeSentencingAndCourtsBill%28TwelfthSitting%29

Institute of Race Relation (2024). Criminal Justice System Statistics. *Institute of Race Relation.* URL: https://irr.org.uk/research/statistics/criminal-justice/

Isaac, A. (2020). Investigating the association between an offender's sex and ethnicity and the sentence imposed at the Crown Court for drug offences. *Sentencing Council for England*

and Wales. URL: https://www.sentencingcouncil.org.uk/wp-content/uploads/Sex-and-ethnicity-analysis-final-1.pdf

Lammy, D. (2017). The Lammy Review: An independent review into the treatment of, and outcomes for, Black, Asian and Minority Ethnic individuals in the Criminal Justice system. Her Majesty's Government. URL: https://www.gov.uk/government/publications/lammy-review-final-report

Lymperopoulou, K. (2024). Ethnic inequalities in sentencing: Evidence from the Crown Court in England and Wales. *The British Journal of Criminology*, 64(5), 1189-1210. URL: https://academic.oup.com/bjc/article/64/5/1189/7612940

Ministry of Justice (2024). Statistics on Ethnicity and the Criminal Justice System, 2022. URL: https://www.gov.uk/government/statistics/ethnicity-and-the-criminal-justice-system-2022/statistics-on-ethnicity-and-the-criminal-justice-system-2022-html

Monteith, K., Quinn, E., Dennis, A., Joseph-Salisbury, R., Kane, E., Addo, F., & McGourlay, C. (2022). Racial Bias and the Bench: A response to the Judicial Diversity and Inclusion Strategy (2020-2025). *University of Manchester.* URL: https://documents.manchester.ac.uk/display.aspx?DocID=64125

Nielson, A. (2017). 'Explain or reform': Howard League responds to Lammy Review. *Howard League*. URL: https://howardleague.org/news/explain-or-reform-howard-league-responds-to-lammy-review/

Pina-Sánchez, J., Dhami, M., & Gosling, J. P. (2024a). Which are the main characteristics determining sentence severity? An empirical exploration of shoplifting offences using spike-and-slab models. In Fix, M. P. and Montgomery, M., editors, *Research Handbook on Judicial Politics*, pages 450–464. Edward Elgar. URL: https://www.e-elgar.com/shop/gbp/research-

handbook-on-judicial-politics-9781035309313.html?srsltid=AfmBOoqWhcMQP1PxgW0jVICArFQ875-KvayYthP176uIUxEkCCyQU_Tw

Pina-Sánchez, J., & Hamilton, M. (2024b). Modelling unwarranted disparities in sentencing: Distinguishing between good and bad controls. *SocArXiv*. URL: https://osf.io/preprints/socarxiv/ymzsv

Pina-Sánchez, J., Morales, A., Guilfoyle, E., Veiga, A., & Geneletti, S. (2025). Testing the interrelationship between area deprivation and ethnic disparities in sentencing. *Analyses of Social Issues and Public Policy*. URL: https://spssi.onlinelibrary.wiley.com/doi/10.1111/asap.12446

Pina-Sánchez, J., & Tura, F. (2024c). Intersectional disparities in sentencing: An application of MAIHDA to Federal Courts data. *CrimRxiv*. URL: https://www.crimrxiv.com/pub/9pmllz74/release/1

Roberts, J.V., & Bild, J. (2021). Ethnicity and Custodial Sentencing: A review of the trends, 2009-2019. *Sentencing Academy*. URL: https://www.sentencingacademy.org.uk/wp-content/uploads/2023/09/Ethnicity-and-Custodial-Sentencing-1.pdf

Sentencing Council (2025) Imposition of Community and Custodial Sentences. URL: https://www.sentencingcouncil.org.uk/overarching-guides/magistrates-court/item/imposition-of-community-and-custodial-sentences-overarching-guideline/#3.%20Pre-sentence%20reports%20(PSRs)

Stockton, B., Strange, C.C., & Harel, O. (2024). Now you see it, now you don't: A simulation and illustration of the importance of treating incomplete data in estimating race effects in

sentencing. *Journal of Quantitative Criminology*, 40, 563–590. URL: https://link.springer.com/article/10.1007/s10940-023-09577-w

Sorsby, A. (2023). The number of convictions before the imposition of a short sentence of immediate custody: ethnicity and gender difference. *Administrative Data Research UK.* URL: https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Insights/Data_Insight_The_number_of_convictions_before_the_imposition_of_a_short_sentence_of_immediate_custody_ethnicity_and_gender_differences_August_2022.pdf

Veiga, A., Pina-Sánchez, J., & Lewis, S. (2023). Racial and ethnic disparities in sentencing: What do we know, and where should we go? *The Howard Journal of Crime and Justice*, 62(2):167–182. URL: https://onlinelibrary.wiley.com/doi/10.1111/hojo.12496

Journal of Legal Research Methodology (JLRM)
ISSN 2752-3403
Alshehri - 1699

https://doi.org/10.19164/jlrm.v4i1.1699

Published date: 20/10/2025

Article

Reviewing Grievances of Automated Decisions in UK Administrative Justice: Qualitative Documentary Analysis Methodology

Aysha Alshehri
PhD student at University of York
Aama556@york.ac.uk

Abstract

AI and other advanced technologies are increasingly deployed in governmental decision-making, including for fundamentally important decisions. Traditional methods of redress for grievances, such as ombudsmen and judicial review, were designed to focus on processes of human decision making, which might not be applicable in cases involving components or whole decisions made by automated processes. There is a dearth of legal precedents for such issues, and theoretical implications of law in this area are typically lagging behind rapid technological and governmental developments. More timely and comprehensive insights are needed to understand emerging administrative justice issues. This paper explores the utilisation of empirical qualitative documentary analysis as a viable methodology to categorise the challenges in reviewing administrative automated decisions grievances, demonstrating the application of systematic review and thematic analysis to derive insights for legal development.

Keywords: Administrative justice, Artificial Intelligence (AI), Automated decision-making, Grievances, Legal analysis, Qualitative documentary analysis.

https://doi.org/10.19164/jlrm.v4i1.1699

Article Text

1 Introduction

The use of automatic algorithms to make decisions based on historical data is increasingly common. This may be relatively innocuous, such as private companies performing customer credit scoring with appropriate permission from consumers, or it may involve serious and profound administrative decisions by governments (e.g., entitlement to government benefits, such as for disability allowances). In traditional cases of administrative decisions being made by humans, the avenues of review (e.g., by ombudsmen, tribunals, or courts) were obvious, but this is less defined with regard to automated decisions.[1] This paper evaluates the scope of empirical qualitative documentary analysis to study automated decision-making (ADM) administrative justice (AJ). Documentary analysis method is commonly used in traditional legal research concerning behaviour, problem sources, and policy formation,[2] typically to overcome the limitations of legal doctrine method with regard to highly contextualised issues and practical effects. [3]

This paper presents qualitative documentary analysis in order to suggest a comprehensive typology of the problems and challenges encountered with regard to reviewing ADM grievances. The method is used to extract data from texts and organise them according to

[1] Jennifer Cobbe, Michelle Seng Ah Lee and Jatinder Singh, 'Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems' [2021] FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 598.
[2] Glenn A Bowen, 'Document Analysis as a Qualitative Research Method' (2009) 9 *Qualitative Research Journal* 27.
[3] Terry Hutchinson, 'Doctrinal Research : Researching the Jury' in Dawn Watkins and Mandy Burton (eds), *Research Methods in Law* (2nd edn, Routledge 2017).

https://doi.org/10.19164/jlrm.v4i1.1699

their types.[4] This facilitates comprehensive categorising of the AJ challenges associated with reviewing ADM grievances. The analysis concludes with a suggested typology of ADM challenges of review, derived from the thematic analysis work that incorporates the perspectives of diverse stakeholders, including government policymakers as well as legal analysts, to help inform real-world solutions.

A clear typology is important because it breaks a broad issue into specific, useful, categories. This also allows for the development of distinct legal and technical responses. Classifying these challenges helps in building theories by highlighting patterns that enable future research to create hypotheses about causes, institutional weaknesses, and possible judicial pathways. For policymakers and judges, a clear typology highlights the most frequent legal or regulatory gaps that need reform. Additionally, by providing a shared vocabulary that connects legal issues to technical details, a typology encourages effective collaboration among legal scholars, social scientists, and technologists. This collaboration enhances both the empirical strength and practical relevance of research on ADM in administrative justice.

2 Methodology

2.1 Overview

The project aims to assess the capacity of Administrative Justice Institutions (AJIs) (e.g. courts, tribunal and ombudsman) to scrutinize and resolve complaints related to automated decisions. In order to achieve this aim, it investigates the challenges faced by AJIs in

[4] Patrick Ngulube, 'Qualitative Data Analysis and Interpretation: Systematic Search for Meaning' in ER Mathipa and MT Gumbo (eds), *Addressing research challenges: making headway for developing researchers* (Mosala-MASEDI Publishers & Booksellers 2015) p.131-156.

https://doi.org/10.19164/jlrm.v4i1.1699

reviewing ADM decisions and explores potential solutions to overcome these issues by conducting a qualitative document analysis research.

Qualitative researchers may generate data (e.g., from interviews) or use pre-existing data (i.e., documents), the latter of which is obviously more practically expedient.[5] Documentary analysis is an obvious and time-honoured way to collect and analyse voluminous data from diverse sources in detail.[6] Documentary analysis is suitable for legal research on automation as a respected qualitative methodology,[7] generating diverse information, including official, corporate, and personal content, and textual, visual, and audio data.[8]

This paper concerns data sourced from diverse sources. These include websites, such as policy documents, expert evidence reports, and government responses regarding ADM on Gov.UK. Similarly, responses to access to information requests, publications by private institutions and NGOs such as the Public Law Project (PLP) and Alan Turning Institution, and surveys are freely available online. Government sources were also used, including parliamentary documents, reports from the Information Commissioner Office (ICO).

Judgments from courts, tribunals, and ombudsmen; judges' opinions on ADM issues, and expert evidence in cases against automated decisions comprised the main substantive legal

[5] Hani Morgan, 'Conducting a Qualitative Document Analysis' (2022) 27 *The Qualitative Report* 64.
[6] Sharan B Merriam and Elizabeth J Tisdell, *Qualitative Research : A Guide to Design and Implementation* (4th edn, John Wiley & Son 2015) p 175.
[7] Michael Quinn Patton, *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (4th edn, SAGE Publications 2015) 84-169.
[8] Lisa Webley, 'Qualitative Approaches to Empirical Legal Research' [2010] in Peter Cane and Herbert M. Kritzer '*The Oxford Handbook of Empirical Legal Research*' 2010, (Oxford University Press) p 938.

sources, while other diverse documents pertaining to real cases were included from newspapers published documents regarding public ADM cases and litigations, published radio podcasts and interviews (such as the Public Law Project's interview with Professor Tomlinson), and videos of some published trials and litigations, such as the Post Office Horizon scandal in the UK, and legal experts and law firms' published videos relating to experience of ADM. These documents have reported on, studied, or discussed reality-based issues that affect the UK judicial system on the role of review and address ADM cases.

Embracing methods less commonly used in legal research is justified on the grounds of the scarcity of empirical studies germane to exploratory analysis of dynamic and emerging issues, and the inadequacy of traditional theoretical and library-based methods and case law. Furthermore, there is a general lack of literature on challenges of ADM appeals and judicial review in the UK, rending more original approaches suitable to analyse legal issues and rules or policies, and to advise on legal reform. [9] Easily available and publicly accessible documents of varying multimedia types[10] can offer diverse and holistic perspectives on legal issues.[11] Digital records can be particularly insightful in troubleshooting required reforms, with regard to legal system flaws, best practices, proof of policy aims, and legislative considerations. Such advantages cannot be obtained from the relatively narrower scope of qualitative interviews, and can avoid types of bias associated with the personal dimension of the latter,[12] and ethical considerations associated with human research participants.[13]

[9] Michael McConville and Wing Hong Chui, *Research Methods for Law* (Edinburgh University Press 2017) p.20.
[10] Morgan (n 5).
[11] Merriam and Tisdell (n 6) pp.164-168.
[12] ibid pp.187-189.
[13] ibid.

https://doi.org/10.19164/jlrm.v4i1.1699

Challenges associated with documentary analysis include the total number and type of documents required before starting research being unknown,[14] and limited data *per se*.[15] For example, when selecting documents for analysis, it was found that there was insufficient data and documents in administrative law providing cases and information about ADM challenges of review. Consequently, documents from other areas of law were included that have documented the challenges of review of ADM. The broader scope of available documents from different aspects of law could identify the challenges of reviewing ADM and may provide solutions which could be helpful for administrative law. The process of conducting documentary analysis begins with selecting documents based on four elements identified by Brıd Dunne et al.: authenticity, credibility, representativeness, and meaning in conducting documentary analysis.[16]

'Authenticity' involves documents' consistency, including being free from linguistic or factual errors or changes,[17] and the provenance and context of sources.[18] 'Credibility' addresses the reliability of a document's source concerning biases.[19] 'Representativeness' intersects with the generalisability of the source.[20] 'Meaning' concerns the implications and interpretation of the text.[21]

[14] ibid.
[15] Morgan (n 5).
[16] Bríd Dunne, Judith Pettigrew and Katie Robinson, 'Using Historical Documentary Methods to Explore the History of Occupational Therapy' (2016) 79 *British Journal of Occupational Therapy* 376.
[17] ibid.
[18] Merriam and Tisdell (n 6).
[19] Dunne, Pettigrew and Robinson (n 16).
[20] ibid.
[21] ibid.

All of the documents selected for analysis in this study were authentic, reliable, representative, and clear, being selected from published sources and official channels (such as Gov.UK publications, professional associations, files of judgments, and published first-hand experts' reports). These sources are recognised as authentic and highly respectable, free from forgeries and other forms of bias.

2.2 Selecting and Categorising Documents

As described below, this documentary analysis began by searching academic databases and websites for relevant texts from a diverse range of relevant sources, including from official government documents, private institutions concerned with administrative justice and technological aspects of ADM, and legal cases and analyses. The resulting of texts underwent thematic analysis, with coding and thematic clustering of identified themes that categorise the challenges of reviewing ADM.

Due to the shortage of research and published cases on ADM grievances, this research uses public documents as a primary data source, including case judgments, experts' evidence to judges, official reports from governmental bodies and private institutions from various aspects of law (administrative, criminal, civil, business, consumer protection, technology). The inclusion of different legal sources can offer a comprehensive understanding of the associated challenges, by including perspectives from empirical evidence and case studies. The broader scope of analysing legal cases from different areas provides the study with a more complete analysis of the process of reviewing ADM. Such data strengthens the analysis

with real-world examples from different legal areas concerning similar issues.[22] This study collected 53 documents from the various channels mentioned above, applying the rule of 'Selecting Observations' mentioned by Lee Epstein and Gary King:

> '1- identify the population of interest; 2- collect as much data as is feasible; 3- record the process by which data come to be observed; and 4- collect data in a manner that avoids selection bias.'[23]

After applying these rules on the documents selected, the analysis can identify challenges in addressing ADM cases and grievances. The documents for this purpose are divided into two categories: (1) documents determining the problems of review faced by judges and other reviewers; and (2) documents including responses and suggested solutions about the problems.

Table 1 provides the criteria about the collection of the documents. In terms of selection criteria, only documents indicating ADM review challenges published on selected websites were chosen for data extraction in the documents collection set. To start, the selection of sources was initially done by looking at regulatory websites on where reports about AI should be submitted, such as government reports and parliamentary committee sessions with experts. From there, private institutions mentioned in government files as entities concerned with AI topics were searched, such as the Alan Turing Institute, PLP, and Information Commissioner's Office (ICO). These entities were found to fund reports and

[22] Bowen (n 2).
[23] Lee Epstein and Gary King, 'Empirical Research and The Goals of Legal Scholarship: A Response' (2002) 69 *University of Chicago Law Review*.

studies for the UK government. In addition, the cases that related to ADM were searched for by typing the names of ADM systems used by different bodies and known to have caused problems for the public. Search engines on websites (such as LexisNexis, Heinonline, and Casemine) were used as a search tool to limit the searching process of the related cases. Judges' statements and experts' evidence are crucial documents for determining ADM issues and challenges.

The main collected and analysed documents are adumbrated below.

- **Government collection:** Reports from the Gov.UK publications, the Centre for Data Ethics and Innovation, and the Government's Central Digital & Data Office highlight the ethical considerations and practical challenges of implementing and overseeing ADM systems within the public sector. Parliamentary documents, including those from Select Committees and the House of Commons, reveal concerns regarding accountability, transparency, and bias in ADM, particularly within the justice system.
- **Private institutions collection:** Advocacy groups such as the PLP, ICO, and the Ada Lovelace Institute and independent institutions (e.g., the Alan Turing Institute) contribute policy papers and reports that emphasise the potential harm due to lack of transparency and accountability of ADM, emphasising the technical complexities involved in auditing and reviewing ADM systems.
- **Legal cases:** Cases demonstrate real-world ADM implementations and the practical application of legal principles. They also reveal judges' experiences regarding ADM-related challenges concerning review, and their impacts on the rights of individuals. These cases often include expert evidence and responses to judges that highlight the

lack of transparency in ADM systems, with even technical experts struggling to understand ADM internal workings due to 'black box' issues (which refers to the algorithmic complexity and opacity of automated systems, which are typically understood only by computer experts or designers).[24]

A targeted typology is derived by empirically identifying related challenges from real-world cases and reports that reflect actual reality. The documents selected for the analysis need to demonstrate the types of the challenges that most adjudicative justice scholars are attempting to identify and solve, such as in reviewing evidence. Tomlinson and others identified key practical challenges limiting any effective judicial review functions, including the opacity of AI and algorithmic technologies, and time limitations for judicial review.[25] Cobbe also argued that producing evidence within the statutory three-month time limit would constitute a significant obstacle (due to the limited time specified), even if demand for judicial review seems set to increase.[26]

Aside from the identified challenges, some of the documents propose a range of potential solutions to address the challenges of ADM review from different areas of law. These solutions vary widely which reflects the complexity evolution of ADM systems, and the policy and legal debates surrounding them.

[24] Karen Yeung, 'Why Worry about Decision-Making by Machine?', *Algorithmic Regulation* (Oxford University Press 2019) P.21-48.
[25] Sarah Nason, 'Oversight of Administrative Justice Systems' in Marc Hertogh and others (eds), *The Oxford Handbook of Administrative Justice* (Oxford University Press 2021).
[26] Jennifer Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making' (2019) 39 Legal Studies 636.

https://doi.org/10.19164/jlrm.v4i1.1699

Some documents were excluded in the process of gathering the collected documents due to not being directly relevant to the scope of this study. These included studies relating to the challenges that faced by ADM decision makers themselves, rather than judges or other *ex post facto* reviewers of ADM decisions. They are related to the challenges associated with the **use** of ADM, such as errors, bias, and privacy issues in the applications of ADM. The document selection process in this study excluded these documents, opting instead to focus on documents and cases relevant to the research question, which was exploring the challenges associated with review ADM.

Thematic analysis was applied to analyse the gathered documents (as described below). After the initial data collection of the types of ADM challenges of review extracted from the documents and texts, information was presented in text and tables. The overall steps were not only to analyse the challenges of review, but to map out a typology from the sources of the highlighted texts. In thematic analysis, all identified challenge types were coded to inform thematic development of an applicable typology for document types, sources, areas of law, and challenges. The content and the codes were retrieved and organised in tables to determine the themes of the concerned issues. The following section describes how the thematic analysis was applied using numbered codes.

https://doi.org/10.19164/jlrm.v4i1.1699

Table 1: Criteria of collected documents and types of Challenges

A) Type of Document	B) Area of Law
1-Report	1-Administrative
2-Policy/bill/guide/strategy	2-Criminal
3-Study/research/survey	3-Civil
4-Governmental response	4-Employment
5-Judgement	5-Healthcare
6-Expert evidence	6-Data protection
7-Judge statement	7-Business
8-Video	8-Competition
9-Podcast	9-Consumer
	10- AI regulation
	11- Not specified
C) Source of Document	**D) Type of Challenge**
1-UK Government	1-Lack of transparency
2-Case	2-Delay
3-Expert evidence	3-Difficult to provide evidence
4-Institutions/organisation	4-Disclosure issue
E) Type of Information	5-Limited access to information
1-Challenges	6-Lack of explanation
2-Response	7-Regulatory gap
3-Both	8-Difficult holding accountability
	9-Lack of authority
	10-Limited redress
	11-Interpretation difficulties
	12-Expertise gap
	13-Cost
	14-Procedural issues
	15-Time limit in judicial review
	16-Litigation cost

https://doi.org/10.19164/jlrm.v4i1.1699

2.3 Thematic Analysis

The process of thematic analysis itself is unpacked in the presentation of the results (below), but a brief description of the method is a useful primer. After creating the documents collection, the next step is to analyse the contents of those documents based on the aim of the study. At this stage, researchers have to choose between the uses of *content* or *thematic* analysis. Content analysis is usually conducted for statistical analysis in quantitative studies,[27] while thematic analysis focuses on how people interpret contextually rich qualitative data.[28] Consequently, thematic analysis is more flexible and suitable for this study. It is typically the default used in deductive research using in-depth expert interviews, as well as in some types of systematic review,[29] albeit the latter differs from thematic analysis in terms of purpose, processes, and data resources. While thematic analysis aims to generate themes of collected information observed from qualitative data, systematic review is a method of comprehensive summarising the results of literature on a specific topic.[30]

Systematic review adheres to a strict protocol of predefined selection criteria to select relevant research and studies.[31] In contrast, the significance of the thematic analysis is that it provides a flexible technique of qualitative data analysis by in-depth examination and interpretation of patterns and themes of meaning in texts from documents.[32] It enables its

[27] Patton (n 7).
[28] Merriam and Tisdell (n 6).
[29] Mark Petticrew and Helen Roberts, *Systematic Reviews in the Social Sciences : A Practical Guide* (John Wiley & Sons 2006) p 87.
[30] William Baude, Adam Chilton and Anup Malani, 'Making Doctrinal Work More Rigorous:Lessons from Systematic Reviews' (2017) 84 University of Chicago Law Review.
[31] Karen Chapman, 'Characteristics of Systematic Reviews in the Social Sciences' (2021) 47 The Journal of Academic Librarianship 102396.
[32] Virginia Braun and Victoria Clarke, 'Conceptual and Design Thinking for Thematic Analysis' (2022) 9 *Qualitative Psychology* 3.

application across various theoretical frameworks, official documents, reports, videos, podcasts and research paradigms.[33]

The thematic patterns identified from reading and reviewing the documents were assigned codes, which were then counted and compared across a dataset (in this case, documentary evidence) in order to identify emergent themes (areas of themes among repetitive codes).[34] In legal research, this method is applicable in many contexts, such as analysing legal issues in published reports, policy documents, expert statements or judgments.[35]

Codes are the smallest analytical units that capture significant features of the data related to the research question; they serve as foundational elements for themes,[36] abstract entities that identify and unify texts under a common meaning, representing broader patterns of meaning based on a central organising concept or a unified core idea.[37] This research followed Braun and Clarke's six-step method, as shown in Table 2 and described below.[38]

[33] Virginia Braun and Victoria Clarke, 'Using Thematic Analysis in Psychology' (2006) 3 Tandfonline 77.
[34] Lisa Webley, 'Qualitative Approaches to Empirical Legal Research' in Peter Cane and Herbert M Kritzer (eds), *The Oxford Handbook of Empirical Legal Research* (Oxford University Press 2012) p 926–950.
[35] ibid.
[36] Victoria Clarke and Virginia Braun, 'Thematic Analysis' (2017) 12 The Journal of Positive Psychology 297.
[37] Lorelli S Nowell and others, '*Thematic Analysis: Striving to Meet the Trustworthiness Criteria*' (2017) 16 International Journal of Qualitative Methods.
[38] Cited by Gerald A Craver, 'Not Just for Beginners-A Review of Successful Qualitative Research: A Practical Guide for Beginners Research: A Practical Guide for Beginners' (2014) 19 DOAJ Directory of Open Access Journals 12.

https://doi.org/10.19164/jlrm.v4i1.1699

Table 2: Braun and Clarke's strategies of thematic analysis[39]

Stage	Thematic Analysis	Description
1	Transcription	Turning audio data into written text (or transcripts) by writing down what was said and how it was said so the data can be systematically coded and analyzed.
2	Reading & Familiarization	Reading and re-reading the data to become intimately familiar with the content (i.e., immersion); analysis begins by noticing things of interest that might be relevant to the research questions.
3	Coding (Selective & Complete)	Identifying aspects of the data that relate to the research questions; can involve *selective coding* where only material of interest is coded or *complete coding* where the entire dataset is coded.
4	Searching for Themes	Identifying salient features that capture something important about the data in relation to the research question; may represent some level of patterned response or meaning within the dataset.
5	Reviewing Themes	Determining whether candidate themes fit well with the coded data; themes should tell a story (not necessarily *the story*) that "rings true" with the data; essentially represents quality control in relation to the analysis.
6	Defining & Naming Themes	Defining themes by stating what is unique and specific about each one; useful because it forces researchers to define the focus and boundaries of the themes by distilling to a few short sentences what each theme is about.
7	Writing the Report	Writing the report by selecting compelling, vivid examples of data extracts, and relating them back to the research question and literature.

Braun and Clarke (2013), pp. 202–203.

3 Results

The outcomes of applying the described thematic analysis methods are presented and explained below.

3.1 Step 1 – Familiarisation with the Content

As the archetypal "Stage 1" displayed in Table 2 was not necessary in this study (as the documents were already in written form, rather than transcribing audio interview data), "Step 1" of *this* study was becoming familiar with the collection. This step involves reviewing the complete collection of documents multiple times, before beginning to record observations and assign codes to chosen texts.[40] The first screening of the documents

[39] Cited by ibid.
[40] Nowell and others (n 37).

https://doi.org/10.19164/jlrm.v4i1.1699

showed some repetitive ADM review challenges relevant to this research, which were helpful to develop the first stage of coding. Texts were selected due to explicitly mentioning direct and indirect challenges of review, determined by searching for the following keywords identified by the researcher as relevant to ADM review:

> Accountability; Ambiguity/ opacity/ vagueness/ vague; Black box; Challenge a decision/ contest; Clarity/ not clear; Cost; Court/ tribunal/ ombudsman; Difficult to understand/ I cannot understand/ understandable; Disclose/ disclosure; Evidence; Expertise/ expert; Explainability; Hard/ difficult to prove; Judge/ reviewers; Judicial review; Oversight/ assess; Transparency.

Subsequently, texts were selected for review according to direct and indirect types of challenges they mentioned. The relevant texts are shown in the data extraction table (Appendix 1). The selected texts illustrated a broad consensus focusing on ADM challenges of review, such as lack of transparency and accountability. For example, it was observed that most reported cases and challenges were about the opacity and lack of explainability of many algorithms (i.e., the black box issue). These issues are also the source of many additional challenges for reviewers to understand how decisions are made, and to assess them (e.g., disclosure problems). In addition, the cost and complexity of legal challenges are also frequently highlighted as barriers to access to justice, especially for individuals or groups with limited resources, who are likely to be victims of administrative injustice. Documents can be coded and categorized based on specific criteria for easy reference throughout the discussion, with unique serial numbers for government (GV), expert

https://doi.org/10.19164/jlrm.v4i1.1699

evidence (EX), cases and judgments (CJ), judges' statements (JS), and private institutions (IN).

3.2 Step 2 – Coding (All Mentioned Challenges)

In the second phase, repeated reading and note-taking for the texts fed into coding data manually according to review challenges, as shown in Table 3. The codes were developed by selecting the labelled challenges of review in the text previously identified in the first step. For example, in [CJ5] *Johnson and others v. Secretary of State for Work and Pensions* (2019), it was observed that the judges noted several challenges in addressing this case. One of which was the difficulty of providing redress for claimants. In the coding process, the text '*While the system was intended to be automated, as evidenced by Ms McMahon's testimony, this automation created complications in addressing specific issues that arose in the case'* comes under challenge code named '**limited redress**'. All codes and their descriptions are shown in Table 3.

Table 3: Coding for mentioned ADM challenges

No.	Code	Description
1	**Lack of transparency**	Black box and commercial sensitivity issues in algorithmic decision-making
2	**Explainability**	Opacity and lack of clarity in decisions and reasoning provided by tribunals/courts
3	**Accountability**	Difficult to hold accountability and determining the responsible party who can be blamed and sanctioned.
4	**Regulatory gaps**	Need to regulate AI and deficiencies in existing legal frameworks
5	**Expertise gap**	Lack of technical and legal expertise to adequately assess and regulate ADM in the AJIs
6	**Disclosure issue**	Issues related to provide the court with data quality, completeness, accuracy, and timeliness of data provided to courts/tribunals
7	**Access to information**	Difficulty accessing information, data, and relevant documentation
8	**Cost**	High costs associated with challenging ADM decisions
9	**Delay**	Delays and inefficiencies in the process of challenging ADM decisions
10	**Lack of authority**	Lack of component authority to regulate, assess and oversee AI
11	**Evidence**	Difficulties in providing evidence to support claims of unfairness or inaccuracy
12	**Time limit**	3-month time limit in judicial review
14	**Redress**	Lack effective redress mechanisms
14	**Interpretation difficulties**	Complexity of the ADM systems, making understanding and challenge extremely difficult

3.3 Step 3 – Development of Themes

At this stage, the identified codes were organised into themes, developed by grouping the codes that revealed most relevant and important ADM challenges of review.

Methodologically, grounded thematic analysis entails deriving deductive themes from codes pertinent to the study's question, as described by Braun and Clarke.[41] Practically, this was implemented by the researcher reading and re-reading the primary sources repeatedly, and

[41] Clarke and Braun (n 36).

considering them in the context of existing legal research. During this process, it was possible to identify recurrent concepts and patterns across sources. These were assigned codes, which were subsequently grouped under themes. The recurrent codes discerned from the documentary analysis thus led to identifying and validating the emergent themes.

This process led to the identification of three main themes, comprising 16 codes derived from the analysis of the primary data, as shown in Table 4. Some codes doubled as names for the overarching theme, such as “Lack of Transparency” as the ADM challenges of review theme, comprising “lack of transparency”, “explainability”, “accountability”, etc. Similarly, the “Expertise Gap” theme encompassed the codes “expertise gap”, “cost”, and “procedural delays”.

https://doi.org/10.19164/jlrm.v4i1.1699

Table 4: Emergent themes

No.	Code	Theme
1	Lack of transparency	**Lack of Transparency**
2	Explainability	
3	Accountability	
4	Disclosure issue	
5	Access to information	
6	Delay	
7	Expertise gap	**Expertise Gap**
8	Cost	
9	Delay	
10	Regulatory gaps	**Regulatory Gap**
11	Time limit	
12	Litigation cost	
13	Redress	
14	Lack of authority	
15	Access to Evidence	
16	Interpretation difficulties	

3.4 Stage 4: Reviewing the Theme Descriptions

Re-reading the texts under each code allowed the researcher to identify three specific patterns of challenges. This step is crucial for refining and clarifying themes to ensure that the themes accurately reflect the selected texts. This often involves splitting themes to achieve better clarity and representation of each group of challenges. Based on the codes, the prominent themes are as shown in Table 5.

https://doi.org/10.19164/jlrm.v4i1.1699

3.5 Step 5: Final Themes

After refining the themes, the final review confirms that all themes are well-defined and related to the research question.[42] This involves summarising each theme and giving it a name that accurately identifies a type of challenge. The final themes in this analysis, as summarised in Table 5, provide a thorough analysis of the ADM review challenges, serving as the basis for establishing the typology that this study intends to develop.

[42] Nowell and others (n 37).

https://doi.org/10.19164/jlrm.v4i1.1699

Table 5: Summary of themes

No.	Code	Theme	Examples
1	Lack of transparency	Lack of Transparency	[CV10] 'achieving full technical transparency is difficult, and possibly even impossible, for certain kinds of AI systems in use today.'
2	Explainability		[EX5] 'Despite the GDPR's intent for a 'right to explanation,' it practically offers a 'right to be informed,' which is limited by trade secret protections'.
4	Disclosure issue		[EX1] 'development and operation of ADM tools through requests under the Freedom of Information Act 2000 (FOIA), both the Home Office and the DWP often refuse disclosure.'
5	Access to information		[EX6] 'I cannot comment on whether AFR Locate has a discriminatory impact as I do not have access to the datasets"
6	Procedural delays		[CJ1] 'The Post Office disclosed crucial documents, including a large number of PEAKs (Problem Event Analysis and Knowledge) and KELs (Known Error Logs), very late in the process,'
7	Expertise gap	Expertise Gap	[CJ6] 'a tribunal's lack of technical expertise directly impacts its ability to assess the significance of the statistical data.'
8	Cost		[IN7] 'high Cost of contesting a decision that need to hire an expert and request for information.'
9	Procedural delays		[JS2] 'The judges spend significant time deciphering these regulations and their implications.'
10	Regulatory gaps	Regulatory Gap	[IN11] 'Existing law is unhelpful in assessing the procedural fairness of ADM/ASDM systems Under the common law of judicial review'.
11	Redress		[GV3] 'insufficient avenues for redress for individuals negatively impacted by algorithmic systems.'
12	Lack of authority		[IN2] 'Existing review bodies such as the Parliamentary and Health Service Ombudsman lack the powers to initiate investigations'
13	Evidence		[IN12] ''Claimants face a range of barriers, including being dissuaded from making a challenge, being required to provide documentation'
14	Interpretation difficulties		[CJ5] 'The primary challenge was determining the proper interpretation of the regulations,
15	Time limit		statutory three-month time limit in Judicial Review
16	Litigation cost		[GV4] 'high costs of litigation in seeking redress.'
3	Accountability		[GV1] 'Lack of clear accountability for who is legally responsible'

https://doi.org/10.19164/jlrm.v4i1.1699

3.6 Step 6: Report

Broadly speaking, UK Administrative Justice Institutions' (AJI) traditional models of redress (for administrative grievances) include mechanisms such as judicial review by courts, tribunal appeals, internal administrative review systems, and ombudsmen. The majority of reviewed cases and documents on AJI challenges reviewing or redressing ADM focused on the role of courts and judicial review rather than other institutions, and judges rarely self-assessed or acknowledged their limited capacities to review opaque technologies or assess expert evidence (e.g., algorithm training). While some relevant cases were reviewed by tribunals, none were investigated by the ombudsman. These non-judicial bodies can be effective in cases pertaining to quasi-regulatory or adjudicative bodies, due to the challenges of lack of judicial expertise in courts and regulatory guidelines, limited procedural times, and other barriers explained in this paper, while the ombudsman has the role of investigating claimed decisions from their internal process until issuing them. Similarly, one body that has yet to be fully explored in the collected documents is the UK ICO, which is responsible for overseeing information rights in the public interest, and data privacy and has quasi regulatory and adjudicative functions and may hear disputes.

Thematic analysis helped to deduce three main themes and identified the challenges that reviewers face while reviewing and addressing ADM issues and cases. As noted before, each theme is characterised by a specific type of challenge. This section provides a detailed description of all of the identified challenges of review in ADM for each emergent theme, to establish the typology of ADM challenges of review.

https://doi.org/10.19164/jlrm.v4i1.1699

Documents in the Government collection (in Appendix 1) seem to be more comprehensive than other groups in determining the ADM challenges of review. This may be because these unified reports are based on data, cases, and reports submitted in different areas of law, also including stakeholder consultations. For example, for the 'AI Barometer 2020' report [GV1], over 100 experts from across five key UK sectors (Criminal Justice, Financial Services, Health & Social Care, Education and Public Sector) informed the government about the most pressing opportunities, risks and governance challenges associated with AI and data.

All experts who submitted evidence in the governmental and case documents analysed in this study stressed their views that all ADM systems have been developed based on black box codes and data, which prevents accessing information in the internal design of the systems and their data. According to the cases collection, it was not possible for the experts to provide judges with answers to their questions about whether the system was wrong, or if the data was biased.

The documents collection also includes live videos and podcasts, which highlighted the ADM challenges of review discussed under the themes in this study. These types of documents enhance the credibility of the analysis, providing insights from reality and official live sources. One example is the live recorded video of *Pantellerisco & others v. Secretary of State for Work and Pensions (2020) [CJ4]*, published by the Court of Appeal on their YouTube channel.

https://doi.org/10.19164/jlrm.v4i1.1699

3.6.1 Theme 1: Lack of transparency

This theme refers to insufficient transparency about the information of the data used to train the AI algorithms that operate the processes in ADM systems, and the need for a clear and understandable explanation of decision-making rationale. It also refers to a lack of fair access to relevant and accurate data and explanation of how ADM systems work for all parties involved. Most collected documents from all sources in Appendix 1 mentioned the lack of transparency as a main source for other challenges, as shown in Table 5. The vast majority (90%) of the selected texts in the documents collections repeatedly cited and referenced this issue.

Based on the overall data from the documents collections for this paper the **limitations on access to information and transparency** significantly hinder the ability of reviewers to evaluate ADM. Without mandatory transparency, individuals and judges appear to face challenges in understanding how automated decisions are made, as well as how ADM affects the subjects of decisions and conventional legal formats. This lack of information not only restricts the public's ability to understand the systems but also limits the parties' capacity to **provide information for evidence** in judicial review.

Although automated decisions were not involved in the case, the issue of a lack of transparency is illustrated in *FO v Secretary of State for Work and Pensions (UC) [2022] UKUT 56 (AAC)*. It was ruled that the Department of Work and Pensions (DWP's) submission to the first tier tribunal was insufficient because of the absence of key documents, like the original

UC (Universal Credit) claim and the initial claimant commitment document.[43] The Social Entitlement Chamber first-tier tribunal (FTT) relied on the DWP's claims without sufficient corroborating evidence. The Upper Tribunal (UT) granted the appeal, stating the FTT make a mistake by not reviewing all relevant documents and misunderstanding the claimant's case. The UT also determined that the FTT incorrectly applied the law concerning the termination of UC awards in the presence of existing commitments. The DWP's representative agreed that the FTT's finding was flawed and that the DWP's response lacked crucial evidence, including the claimant commitment document and documentation showing the proper procedure for setting new commitments.[44]

Furthermore, during the proceedings captured in the video of *Pantellerisco & others v. The v Secretary of State for Work and Pensions (2020) [CJ4]*, the judge remarked on the challenges of **lack of transparency**. In the minutes from 29:00 to 34:00, the judge expresses a need to understand the UC system stating:

> "… I think it's confusing enough…"
>
> " …but I'm just trying to understand it…'
>
> '… that doesn't tell you enough in order to understand what the intention or what the purpose of the scheme should be for that sort of perceived income is throwing up…' [CJ4].[45]

[43] *FO v Secretary of State for Work and Pensions (UC)* (2022) 56.
[44] ibid.
[45] Court of Appeal, 'Pantellerisco & Others (Claimant/Resp) v Secretary Of State for Work and Pensions (Def/Appellant) - YouTube' (15 June 2021) Pantellerisco & others (claimant/resp) v Secretary Of State for Work and Pensions (def/appellant) - YouTube accessed on 19 November 2024.

The [GV1] report mentioned above concluded that:

> 'It is difficult for people to understand or challenge decisions made or informed by algorithms because of their 'black box' nature or commercial confidentiality regarding their functionality.' [GV1]
>
> 'It is difficult for supervisory bodies to interrogate the accuracy and robustness of AI and data-driven systems used within financial services (e.g., in credit decisions) due to lack of transparency and their 'black box' nature.' [GV1].[46]

3.6.2 Theme 2: Regulatory gap

This theme means insufficient existing legal frameworks to address unique challenges related to automated decisions. The main characteristic of the documents collected in this theme is the frequent reference to **lack of clear accountability**, **lack of competent authority**, **limited types of redress**, and **interpretation difficulties**. Accountability requires adequate avenues for people to challenge ADM systems, together with effective enforcement mechanisms and the possibility of sanctions.

Recently, increasing reference to the insufficiency of AI and ADM regulation can be observed in all types of the collected documents. Scholars and academics from different perspectives (including law as well as technological fields) have increasingly discussed the effects and challenges of the absence of clear AI regulations to the AJ review process. AJ scholars have indicated that:

[46] 'CDEI AI Barometer ' (*GOV.UK*, 23 June 2020) <https://www.gov.uk/government/publications/cdei-ai-barometer/cdei-ai-barometer> accessed 19 January 2025.

> 'ADM relies on personal data (which will be in most systems used in public administration to make decisions about individuals), the General Data Protection Regulation and the Data Protection Act 2018 require a variety of information to be disclosed. This includes the right for information to be provided.'[47]

Nevertheless, they stated that there are limitations to these protections, raising questions about whether it is feasible to receive a comprehensive explanation of an ADM system.[48] Under this theme, the analysis revealed the challenges that resulted from the regulatory gap as discussed below.

Ambiguous or unclear regulations led to **challenges in interpreting** the legal requirements. In Judge Wright's statement, repeatedly highlights the complexity and intricacy of the UC regulations, specifically concerning earned income calculations (regulations 54 and 61). The judges spent significant time deciphering these regulations and their implications, and the **difficulty in applying and interpreting** the law was clearly demonstrated. The court's own description of the reasoning as 'compressed' further highlights this issue.

Another example from the documents under the theme of regulatory gap is demonstrated in the report '*Auditing algorithms: The existing landscape, role of regulators and future outlook*' [GV2] published on the Gov.UK website, which focuses on the governance and

[47] Joe Tomlinson, Katy Sheridan and Adam Harkens, '*Judicial Review Evidence in the Era of the Digital State*' [2020] SSRN Electronic Journal 740-760.
[48] ibid.

auditing of algorithmic systems. While it does not specifically address algorithmic judicial review, it does extensively discuss the broader issues of algorithmic accountability, transparency, and the need for auditing processes. The report notes the **insufficient avenues for redress** for algorithmic systems, and that the current algorithmic audit landscape **is largely unregulated**.

3.6.3 Theme 3: Expertise gap

The collected documents show that this theme illustrates that AJIs lack the technical expertise to assess and review complex algorithmic evidence. It is helpful to begin analysing this theme by examining the real cases to determine the level of technological knowledge about ADM in AJIs. Therefore, this theme will initiate the discussion by looking at what judges and reviewers have expressed about the lack of knowledge concerning ADM systems. In this context, Jude Wright stated in [JS2] that the extensive analysis and discussion of technical legal issues presented challenges due to the complexity of the UC system and the ambiguities in both the initial decision and the Court of Appeal's declaration. He added that the court **needed to extensively research** and consider nuanced aspects of multiple legal precedents to arrive at its conclusions. His statement emphasises a type of challenge under this theme that the court needs '**specialised knowledge**':

> 'The court's detailed analysis of regulations, legal precedents (Johnson, NCCL, Majera, etc.), and the overall intricacies of the UC system and the calculation methods suggests a requirement for **specialised knowledge to fully grasp** the matter.'

https://doi.org/10.19164/jlrm.v4i1.1699

In addition, Lord Sales notes that challenges in the courts may arise from 'technical illiteracy,' as 'understanding algorithms **requires specialised skills**,' which most people do not have. He added that judges are not well equipped to assess or understand whether the system of automation relies on a defective methodology to analyse the data inputted into them, or which are too inflexible to account for differences in individual cases (such as in the *Johnson* decision and *HMRC v Tooth*).

Overall, from the documents collected it can be observed that the expertise gap caused other challenges for both the court and the affected individual. The expertise gap within AJIs exacerbates **delays** in the judicial review process. The need for expert testimony to understand complex algorithmic evidence significantly lengthens proceedings, contributing to substantial delays. Furthermore, this same expertise gap drives up the **cost** of litigation. The requirement for expert evidence and testimony adds a considerable financial issue to legal challenges, making access to justice more difficult and potentially deterring individuals from pursuing necessary reviews of ADM. This is supported by the Alan Turing Institute's written evidence, cautioning about 'the **financial burden** a citizen may have to undergo in hiring the right type of expert to support their challenge.' The BIIL report also raised the causes of the expertise gap while contesting automated decisions, and highlighted that the need for expertise and knowledge lead to the **high cost** of contesting a decision that needs to hire an expert and request for information.

In general, a main observation has been noted from the above discussion that the courts may not understand expert testimony and evidence even if they request it to fill the

expertise gap in ADM judicial review process due to combined challenges. These are the expertise gap within the courts beside the other challenges like lack of transparency in the ADM process and lack of explanation in responses by the respondents.

4 Suggested Typology of ADM Challenges of Review

Based on the thematic analysis, two fundamental aspects of ADM challenges of review typology have emerged. The first is challenges primarily affecting judges and administrative justice in general, including the themes transparency and explainability, legal and regulatory gap, technical expertise, and practical and procedural issues (Table 5). Inconsistent transparency practices exist in accessing information and dealing with varying levels of transparency across different organisations and cases. Judges also face the challenges of uncertainty in the legal basis while seeking to interpret and apply laws to ADM. The difficulty in identifying the responsible party for algorithmic bias or error in complex systems comes also under the regulatory gap challenges. Based on the above analysis, the most common issue in ADM judicial review is the lack of expertise in the judicial system and limited understanding of the technical aspects of ADM within the judiciary.

Secondly, there are challenges primarily facing people affected by automated decision making (Table 5). The **lack of transparency and accessing to information** due to a refusal to disclose information directly affect people's rights to have an explanation about the decisions. For example, in *Public Law Project V. The Information Commissioner (2023) [CJ2],* the PLP appealed the decision by the Information Commissioner which upheld the HO refusal of a freedom of information request. The complainant requested information about

https://doi.org/10.19164/jlrm.v4i1.1699

the model's criteria, including how nationality was considered in the “sham marriage” tool. However, the Home Office refused to fully disclose information related to a sham marriage triage model.

This issue also affects their ability to provide evidence in issuing judicial review proceedings. Similarly, many of the collected documents indicate that the people are struggling to understand the reasons and the rationale for automated decisions, because of their limited awareness about ADM technology; indeed, in some cases they do not even notice (or are not adequately informed) that a decision was issued by an ADM system. Furthermore, financial barriers to challenging ADM decisions (including hiring experts) potentially prevent justice. The basic typology arising from the analysis undertaken is shown in Table 6.

https://doi.org/10.19164/jlrm.v4i1.1699

Table 4: Overview of basic typology

Typology Category	Challenges Primarily Affecting Judges	Challenges Primarily Affecting Individuals	Overlapping Challenges
I. Transparency & Explainability	Problems of disclosure. Insufficient explainability. Access to information.	Limited access to information. Difficulties in obtaining evidence.	Lack of transparency, insufficient explainability.
II. Legal & Regulatory Gap	Uncertainty in the legal basis. Lack of clear accountability (determining responsible parties). Applying existing legal frameworks to ADM. Insufficient redress mechanisms.		
III. Technical Expertise	Lack of expertise (in the judicial system). Delay in understanding ADM. Cost to hire external experts.	Lack of technical awareness. Delay in providing evidence and waiting for outcomes. Cost in seeking legal support from experts.	Lack of technical expertise. Delay. Cost.

Most collected documents covered ADM application and substantive use in decision-making in relation to privacy rights, judicial review concerning legality, or rules of applicable ADM standards, while the ADM challenges of review is based on complex and opaque systems, indicating the need to develop the role of judicial review in ADM cases. Collected documents concurred that the magnitude of ADM warrants reasoning and transparency requirements, but ADM challenges of review in AJIs has received negligible consideration in all areas of law and policy, and is merely *inferred* from expert evidence in some cases.

https://doi.org/10.19164/jlrm.v4i1.1699

Most pertinent documents concern judicial review procedures, such as the lack of metrics for assessing and evaluating ADM, time limits, and the need for expert evidence.[49] Moreover, there are some challenges raised about how courts can determine competences for decisions based on specialist/ opaque codes and technologies.[50] Relatedly, some sources heavily focused on the challenge of the lack of AJIs' expertise and disclosure requirement issues,[51] associated with costs and delays.[52] By exploring academic sources from different areas of law, some have demonstrated problems such as data gap, limited competent authority and difficulties in holding accountability to ADM systems where there is no human intervention.[53]

Few analysts have addressed how ADM affects the role of AJIs in fields outside administrative law (e.g., civil and commercial law), given that judicial review is typically associated with public law. Therefore, addressing potential review challenges seems incomplete in UK law. However, there are few cases in criminal, civil, intellectual property, and business law that can be considered here for identifying the types of challenges and that fill the gap of the solutions needed in this study. For example, concerning the "Issues arising

[49] Rebecca Williams, 'Rethinking Administrative Law for Algorithmic Decision-making' (2021) 42 Oxford Journal of Legal Studies.
[50] Igor Gontarz, 'Judicial Review of Automated Administrative Decision-Making: The Role of Administrative Courts in the Evaluation of Unlawful Regimes' (2023) 2023 ELTE Law Journal 151.
[51] Michèle Finck, 'Automated Decision-Making and Administrative Law' in Peter Cane and Other (eds), *The Oxford Handbook of Comparative Administrative Law* (Oxford University Press 2020) P 655-676, see also, Richard Moorhead, Karen Nokes and Rebecca Helm, '*Post Office Scandal Project: Issues Arising in the Conduct of the Bates Litigation*' (2021) Evidence Based Justice Lab Avalaible at https://evidencebasedjustice.exeter.ac.uk/wp-content/uploads/2021/08/WP1-Conduct-of-the-Bates-Litigation-020821.pdf.
[52] Matt Davies and Michael Birtwistle, '*Regulating AI in the UK* ' Report (2023) Ada Lovelace Institute Available at https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/ .
[53] Abe Chauhan, '*Towards the Systemic Review of Automated Decision-Making Systems*' (2021) 25 Judicial Review 285.

https://doi.org/10.19164/jlrm.v4i1.1699

in the Conduct of the Bates Litigation" about *Bates v Post Office*, concerns were raised about the decision that affected more than 500 employees. Issues such as disclosure problems, costs, and expert delays and misconduct in evidence arose as serious concerns in reviews of decisions in civil courts.[54]

5 Conclusion

This paper gathered a diverse collections of texts, including from official government documents, private institutions concerned with administrative justice and technological aspects of ADM, and legal cases and analyses. Sourced from searches of academic databases and the internets, high-quality and important publications were selected that were within the real-life ADM decisions and concerns pertaining to review. The resultant texts were thematically analysed, with coding and thematic clustering of identified themes, in order to identify three emergent thematic categorisations: "lack of transparency," "*r*egulatory gap," and "expertise gap." Based on this analysis, the parameters of a new typology of ADM challenges was suggested, including "Transparency and Explainability," "Legal and Regulatory Gap," and "Technical Expertise." While the emergent typology incorporates authentic and relevant issues pertaining to the scope of ADM review, it should be noted that there are inherent limitations when qualitatively selecting texts and the subjective analysis of qualitative data, which is an inherent limitation of documentary analysis.

Nevertheless, incorporating a wide array of legal sources and empirical data, this paper has mapped out the multifaceted nature of grievances associated with ADM, highlighting the

[54] Moorhead, Nokes and Helm (n 51).

gaps in existing literature and case law. The qualitative documentary found that issues pertaining to the identified black box issue of ADM systems and the dearth of relevant technological expertise among most public servants and legal experts concerned with ADM redress can result in limited transparency and gaps in knowledge among justice institutions. Such issues can affect disadvantaged people, whether or not judges hear apportioned witnesses or experts. This is particularly exacerbated by the fact that legal proceedings in general are typically time-consuming, and prerequisite data cannot universally be rendered accessible on a timely basis.

This method has not only revealed the limitations of current administrative practices, but also emphasized the need for a comprehensive typology to better address these challenges in relation to fast-emerging technologies, based on relatively novel methods of systematic review and thematic analysis not commonly used to comprehend areas of law. By examining best practices from various fields, this research aims to propose viable solutions that can be adapted to enhance administrative law's responsiveness to ADM grievances.

Ultimately, advancing administrative justice in the era of technological transformation requires a concerted effort to bridge the existing legislative gaps and procedural solutions. Continued exploration and dialogue in this domain will be vital in fostering an administrative system that honours individual rights while navigating the complexities introduced by emerging technologies.

References

Abe Chauhan, 'Towards the Systemic Review of Automated Decision-Making Systems' (2021) 25 Judicial Review 285.

Bríd Dunne, Judith Pettigrew and Katie Robinson, 'Using Historical Documentary Methods to Explore the History of Occupational Therapy' (2016) 79 British Journal of Occupational Therapy 376.

'CDEI AI Barometer ' (GOV.UK, 23 June 2020) <https://www.gov.uk/government/publications/cdei-ai-barometer/cdei-ai-barometer> accessed 19 January 2025.

Court of Appeal, 'Pantellerisco & Others (Claimant/Resp) v Secretary Of State for Work and Pensions (Def/Appellant) - YouTube' (15 June 2021) Pantellerisco & others (claimant/resp) v Secretary Of State for Work and Pensions (def/appellant) - YouTube accessed on 19 November 2024.

FO v Secretary of State for Work and Pensions (UC) (2022) 56.

Gerald A Craver, 'Not Just for Beginners-A Review of Successful Qualitative Research: A Practical Guide for Beginners Research: A Practical Guide for Beginners' (2014) 19 DOAJ Directory of Open Access Journals 12.

Glenn A Bowen, 'Document Analysis as a Qualitative Research Method' (2009) 9 Qualitative Research Journal 27.

Hani Morgan, 'Conducting a Qualitative Document Analysis' (2022) 27 The Qualitative Report 64.

https://doi.org/10.19164/jlrm.v4i1.1699

Igor Gontarz, 'Judicial Review of Automated Administrative Decision-Making: The Role of Administrative Courts in the Evaluation of Unlawful Regimes' (2023) 2023 ELTE Law Journal 151.

Jennifer Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making' (2019) 39 Legal Studies 636.

Joe Tomlinson, Katy Sheridan and Adam Harkens, 'Judicial Review Evidence in the Era of the Digital State' [2020] SSRN Electronic Journal 740-760.

Karen Chapman, 'Characteristics of Systematic Reviews in the Social Sciences' (2021) 47 The Journal of Academic Librarianship 102396.

Lee Epstein and Gary King, 'Empirical Research and The Goals of Legal Scholarship: A Response' (2002) 69 University of Chicago Law Review.

Lisa Webley, 'Qualitative Approaches to Empirical Legal Research' in Peter Cane and Herbert M Kritzer (eds), The Oxford Handbook of Empirical Legal Research (Oxford University Press 2012) p 926–950.

Lorelli S Nowell and others, 'Thematic Analysis: Striving to Meet the Trustworthiness Criteria' (2017) 16 International Journal of Qualitative Methods.

Mark Petticrew and Helen Roberts, Systematic Reviews in the Social Sciences : A Practical Guide (John Wiley & Sons 2006) p 87.

Matt Davies and Michael Birtwistle, 'Regulating AI in the UK ' Report (2023) Ada Lovelace Institute Available at https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/ .

Michael McConville and Wing Hong Chui, Research Methods for Law (Edinburgh University Press 2017) p.20.

Michael Quinn Patton, Qualitative Research & Evaluation Methods: Integrating Theory and Practice (4th edn, SAGE Publications 2015) 84-169.

Michèle Finck, 'Automated Decision-Making and Administrative Law' in Peter Cane and Other (eds), The Oxford Handbook of Comparative Administrative Law (Oxford University Press 2020) P 655-676.

Patrick Ngulube, 'Qualitative Data Analysis and Interpretation: Systematic Search for Meaning' in ER Mathipa and MT Gumbo (eds), Addressing research challenges: making headway for developing researchers (Mosala-MASEDI Publishers & Booksellers 2015) p.131-156.

Rebecca Williams, 'Rethinking Administrative Law for Algorithmic Decision-making' (2021) 42 Oxford Journal of Legal Studies.

Richard Moorhead, Karen Nokes and Rebecca Helm, 'Post Office Scandal Project: Issues Arising in the Conduct of the Bates Litigation' (2021) Evidence Based Justice Lab Avalaible at https://evidencebasedjustice.exeter.ac.uk/wp-content/uploads/2021/08/WP1-Conduct-of-the-Bates-Litigation-020821.pdf.

Sarah Nason, 'Oversight of Administrative Justice Systems' in Marc Hertogh and others (eds), The Oxford Handbook of Administrative Justice (Oxford University Press 2021).

Sharan B Merriam and Elizabeth J Tisdell, Qualitative Research : A Guide to Design and Implementation (4th edn, John Wiley & Son 2015) p 175.

Terry Hutchinson, 'Doctrinal Research : Researching the Jury' in Dawn Watkins and Mandy Burton (eds), Research Methods in Law (2nd edn, Routledge 2017).

Victoria Clarke and Virginia Braun, 'Thematic Analysis' (2017) 12 The Journal of Positive Psychology 297.

https://doi.org/10.19164/jlrm.v4i1.1699

Virginia Braun and Victoria Clarke, 'Conceptual and Design Thinking for Thematic Analysis' (2022) 9 Qualitative Psychology 3.

Virginia Braun and Victoria Clarke, 'Using Thematic Analysis in Psychology' (2006) 3 Tandfonline 77.

William Baude, Adam Chilton and Anup Malani, 'Making Doctrinal Work More Rigorous:Lessons from Systematic Reviews' (2017) 84 University of Chicago Law Review.

https://doi.org/10.19164/jlrm.v4i1.1699

Appendix 1: Analysed documents

#	Code	Document	Source	Type of Document	Area of Law	Challenge	Type of Information
Government							
1	G1	AI Barometer Report	1	1	2,5,7	1, 6, 8, 12,	3
2	GV2	Auditing algorithms: the existing landscape, role of regulators and future outlook	1	3	8, 9	1, 7, 10	3
3	GV3	Ethics, Transparency and Accountability Framework for Automated Decision-Making	1	3	1, 6	1, 8	3
4	GV4	A pro-innovation approach to AI regulation: government response	1	4	10	1, 7, 13	3
5	GV5	Review into bias in algorithmic decision-making	1	1	1, 2, 4, 7	1, 3, 6, 7, 10, 12	3
6	GV6	Study on the Impact of Artificial Intelligence on Product Safety	1	4	3, 9	1, 7, 12	1
7	GV7	The King's Speech 2023	1	2	7		2
8	GV8	Predictive Policing- West Midlands Police. Response to request based on Freedom of Information Act (736A/22)	1	4	2		2
9	GV9	Artificial Intelligence and Public Standards A Review by the Committee on Standards in Public Life	1	4	1	1, 7, 12	3
10	GV10	AI in the UK: ready, willing and able (parliament.uk)	1	4	11	1, 6, 7	1

#	Code	Document	Source	Type of Document	Area of Law	Challenge	Type of Information
11	GV11	Automatic Computer-based Decisions: Legal Status, Volume 690: debated on Wednesday 10 March 2021	1	1	10	1, 3, 7	1
12	GV12	The governance of artificial intelligence: interim report	1	1	10	1, 5, 8	3
13	GV13	Artificial intelligence and employment law	1	3	4	1, 7, 8, 11	3
14	GV14	AI and Healthcare	1	1	5	1, 7, 8, 9	1
15	GV15	Policy implications of artificial intelligence (AI)	1	2	2, 5, 10		2
16	GV16	Public Authority Algorithmic and Automated Decision-Making Systems Bill [HL]	1	2	10		2
17	GV17	Potential impact of artificial intelligence (AI) on the labour market	1	1	4	1, 8	3
18	GV18	Interpretable machine learning	1	3			2
19	GV19	Technology rules? The advent of new technologies in the justice system	1	1	1,4,		
20	GV20	the (UK Judicial Attitude Survey England & Wales courts, coroners and UK tribunals 2024)	1	3		11	

https://doi.org/10.19164/jlrm.v4i1.1699

#	Code	Document	Source	Type of Document	Area of Law	Challenge	Type of Information
Cases							
21	CJ1	Bates v. Post Office (2019)	2	5	3	1, 2, 4, 5	1
22	CJ2	Public Law Project V. The Information Commissioner (2023)	2	5	1	4, 5	1
23	CJ3	Bridges v South Wales Police (2020)	2	5	2	7, 12	3
24	CJ4	Pantellerisco & others v. Secretary of State for Work and Pensions (2020)	2	8	1	1, 6	1
25	CJ5	Johnson and others v. Secretary of State for Work and Pensions (2019)	2	5	1	1, 7, 10, 11	3
26	CJ6	Ofqual v. ICO (2023)	2	5	1	4, 5, 7, 12	1
27	CJ7	PLP v. ICO (2022)	2	5	1	1, 4, 8	3
28	CJ8	Pa Edrissa Manjang & other v. Uber	2	5	4	2, 6	1

#	Code	Document	Source	Type of Document	Area of Law	Challenge	Type of Information
Institutional Reports							
29	IN1	Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges by Alexander Babuta and Dr Marion Oswald MBE	3	1	2	1, 6, 7	3
30	IN2	Developing AI regulation: findings from PLP's roundtable	3	1	10	7, 9, 10	3
31	IN3	Digital Immigration Status: A Monitoring Framework by PLP	3	1	1, 10	1, 5, 10	3
32	IN4	Machine Learning Used to Stop Universal Credit Payment by PLP	3	1	1	1, 4, 8	3
33	IN5	Transparency mechanisms for UK public-sector algorithmic decision-making systems	3	1	1	1, 5, 8	3
34	IN6	Findings from ICO consensual audits on Freedom of Information of police forces in England and Wales	3	1	2	1, 2, 5	1
35	IN7	Contesting AI explanations in the UK	3	1	1	1, 3, 7, 13	1
36	IN8	Contesting automated decision making in legal practice: Views from practitioners and researchers	3	1	2	1, 8	1
37	IN9	All You need to know about AI adoption in Criminal Justice by Manish Garg	3	1	1	1, 7, 9, 12	3
38	IN10	Legal and regulatory frameworks governing the use of automated decision making and assisted decision making by public sector bodies.	3	1	10	1, 6, 7, 10, 12, 13	1
39	IN11	Reforming the law around the use of automated and assisted decision making by public bodies.	3	1	1	2, 3, 5	3
40	IN12	Surveying Judges about artificial intelligence: profession, judicial adjudication, and legal principles	3	3		11	

https://doi.org/10.19164/jlrm.v4i1.1699

#	Code	Document	Source	Type of Document	Area of Law	Challenge	Type of Information
Judges' Statements							
41	JS1	Algorithms, Artificial Intelligence and the Law The Sir Henry Brooke Lecture for BAILII Freshfields Bruckhaus Deringer, London Lord Sales, Justice of the UK Supreme Court 12	4	7	2		
42	JS2	Judge Wright statement	4	7	1	2, 7, 11, 12	1
43	JS3	*Lord Sales 'Information Law and Automated Governance, Keynote address at the Information Law Conference Institute of Directors, 24 April 2023'*	4	7			

Journal of Legal Research Methodology

#	Code	Document	Source	Type of Document	Area of Law	Challenge	Type of Information
Expert Evidence					**Expert in**		
44	EX1	Written Evidence to the Parliament Submitted by Public Law Project	5	6	Legal	1, 3, 4, 7, 8, 10	3
44	EX1	Written Evidence to the Parliament Submitted by Public Law Project	5	6	Legal	1, 3, 4, 7, 8, 10	3
45	EX2	How can the Department of Work and Pensions operate more transparently, lawfully, and fairly?	5	6	Legal	4, 5, 6	1
46	EX3	Professor Andrew Le Sueur, University of Essex, Advisory Evidence on ADM reforms	5	6	Legal	7	3
47	EX4	Written evidence to the Parliament submitted by Dr Alison Powell	5	6	Legal	1	3
48	EX5	Written evidence submitted by The Alan Turing Institute on “algorithms in decision-making” to the House of Commons’ Science and Technology Committee	5	6	Computer and data scientists	1, 7, 13	3
49	EX6	Witness report in ED Bridge v. South Wales Police from Professor Anil Jain	5	6	Computer science and engineering	5	1
50	EX7	The case of transparency, Podcast (voice source) with Joe Tomlinson (https://publiclawproject.org.uk/latest/people-law-power-the-new-podcast-from-plp/)	5	9	Legal	1, 7	3
51	EX8	Witness statements by Carol Krahé in Pantellerisco & others v Secretary of State for Work and Pensions (https://www.judiciary.uk/wp-content/uploads/2020/07/R-Pantellerisco-v-SSWP-Final-Approved.pdf)	5	6	IT	5	3
52	EX9	Witness statements by Ms McMahon regarding the technical and administrative aspects of UC system	5	6	Computer science and technology	10	3