

Mental health and sentencing: How are judicial decisions made in light of the judgement of *R v Vowles*?

George Baldwin^{a,1,2}, Samantha Young^{a,3,2}, Lucy Fitton^a, Ian Edwards^b, Michael Butler^a, Peter Beazley^{a,*}

^a Norwich Medical School, University of East Anglia, Norwich NR4 7TJ, UK

^b Law School, University of East Anglia, Norwich NR4 7TJ, UK

ARTICLE INFO

Keywords:
Vowles
Hybrid order
Sentencing
Mental disorder
Diagnosis
Decision making
Experimental

ABSTRACT

In England and Wales, s.45A of the Mental Health Act 1983 allows a judge to pass a sentence including both an immediate direction to hospital as well as a punitive custodial element. *R v Vowles* provides four specific considerations for judges to attend to when considering such sentences (referred to as the 'Vowles statements'). The section, however, remains infrequently used. The present study adopted an online experimental methodology to explore decision-making in relation to the Vowles statements. We used a proxy judicial sample who made decisions about the same criminal case vignette. The experimental manipulation meant that participants were exposed to the same information except for the clinical diagnosis: 'complex mental health condition', 'Emotionally Unstable Personality Disorder' (EUPD) or 'Schizophrenia'. Participants were asked to decide which sentence they were most likely to give and rate their agreement with the each of the Vowles statements. Analysis considered relationships between Vowles statements, differences between experimental conditions, and the extent to which different factors (including the Vowles statements) predicted the overall sentencing decision. Results identified that s.45A was, by far, the most common sentencing decision, and that agreement on the different Vowles statements was variable. There was limited evidence of an impact of diagnosis on decision-making, except for some weak evidence that an EUPD diagnosis was associated with marginally higher rates of prison sentences. Most importantly, not all of the Vowles statements were predictive of the final sentence, with attitudes towards the need for punishment having the clearest relationship with the final sentencing decision.

1. Introduction

When an offender with a 'mental disorder' is convicted of a serious crime in England and Wales, the two primary potential sentences are a custodial sentence or a hospital order under s.37 of the Mental Health Act 1983. The two sentences have drastically different outcomes; a custodial sentence means the offender is required to serve a (typically) determinate length of time within a prison, and, if necessary, would be offered treatment for their mental disorder within the prison system (if they required compulsory treatment for their mental disorder, they could be transferred to a psychiatric hospital whilst still serving their custodial sentence). Upon completion of the custodial part of their

sentence, they would be released, with the remainder of the sentence spent on licence and supervision in the community and any ongoing risk to the public being managed by probation services.

On the other hand, a hospital order under s.37 means that the offender – now a 'patient' – is instead directed to a psychiatric hospital where they are detained 'without limit of time' to receive treatment. The process of discharge is governed primarily by considerations of the patient's risk, with community follow-up being organised by via NHS mental health services. An offender can only be given a hospital order if two medical practitioners provide evidence that the relevant criteria under s.37 of the Mental Health Act 1983 are met (primarily that the offender has a mental disorder of a 'nature or degree which makes it

* Corresponding author.

E-mail addresses: George.Baldwin@nsh.nhs.uk (G. Baldwin), L.Fitton@uea.ac.uk (L. Fitton), I.Edwards@uea.ac.uk (I. Edwards), M.Butler1@uea.ac.uk (M. Butler), p.beazley@uea.ac.uk (P. Beazley).

¹ Present address: Norfolk and Suffolk Mental Health Foundation NHS Trust, Drayton High Rd, Norwich, NR6 5BE, UK

² N.B. Baldwin and Young contributed equally and are to be credited as joint first authors.

³ Present Address: Great Ormond Street Hospital, Great Ormond Street London WC1N 3JH, UK

<https://doi.org/10.1016/j.ijlp.2025.102071>

Received 3 January 2025; Received in revised form 5 February 2025; Accepted 6 February 2025

Available online 10 February 2025

0160-2527/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

appropriate for him to be detained in hospital for medical treatment' (s.37(2)(a)(i)) and that 'the court is of the opinion having regard to all the circumstances including the nature of the offence and the character and antecedents of the offender, and to the other available methods of dealing with him, that the most suitable method of disposing of the case is by means of an order under this section' (s.37(2)(b)). A further decision for the court, where a s.37 disposal is being considered, is whether to additionally impose a 'Restriction Order' under s.41 of the Mental Health Act 1983. Such an order is imposed if it is considered necessary 'for the protection of the public from serious harm so to do'. If imposed, this results in restrictions on the powers of the 'Responsible Clinician' to discharge the patient or grant leave, with the Secretary of State playing a more direct role in approval of such decisions.

In practice, these different pathways mean that an offender could spend vastly different periods of time in custody or hospital depending on the type of sentence imposed. However, these two sentencing options were the only possible outcomes until the Crime (Sentences) Act 1997 inserted s.45A – the 'Hospital and Limitation Direction' – into the Mental Health Act. Such a disposal, known more commonly as a 'hybrid order', appears to present the courts with an intermediate option between prison and hospital: immediate transfer to hospital for treatment, but additional imposition of a custodial sentence. Once the patient no longer requires treatment in hospital, they would then return to prison to serve the remainder of their sentence as a prisoner. [Laing \(1996\)](#) and [Delmage et al. \(2015\)](#) provide excellent coverage of the historical context of the introduction of the s.45A, and for present purposes it is sufficient to explain that the context at the time was a political concern about risky or dangerous patients in the community, with neither hospital orders nor sentences of imprisonment providing a satisfactory sentencing outcome. Notably, at this time, s.45A orders were only available to those with 'psychopathic disorder' (outdated legal shorthand for personality disorder).

However, the s.45A option was infrequently used, with only a handful of cases each year, and subsequent efforts that might have increased its use have not done so. For instance, significant changes to the Mental Health Act 1983 came into force as a result of the Mental Health Act 2007. In theory, these amendments widened the potential scope of the Hybrid Order by removing its restriction to 'Psychopathic Disorder' and making it available to anybody who was eligible for a s.37. Moreover, even prior to this legislative change, developments in case-law had led to some potential widening of scope of s.45A ([Whyte & Gupta, 2007](#)). However, the order remained infrequently used: although there was some evidence that these changes led to an increase in usage, yearly figures remained less than 20 such disposals in the years 2009–2013 ([Delmage et al., 2015](#)).

Since this time, whilst there have been no further changes to statute concerning s.45A, developments in case-law have again potentially widened its scope. One of the key pieces of case-law, and the focus of the present article, was the Court of Appeal's judgement in [R v Vowles and others \(2015\)](#). The wider context and impact of this judgement has been discussed in more detail by others ([Peay, 2016](#)), but for practical purposes the judgement was of particular import as it appeared to create an expectation that a s.45A order was to be preferred to a hospital order when sentencing offenders with a mental disorder.

According to the Vowles judgement (§54), in cases where a court was considering imposing a hospital order (whether with or without a restriction order), it should first 'consider whether the mental disorder can appropriately be dealt with by a hospital and limitation direction under s.45A' instead, and – crucially – if it concluded that it could impose such an order, the court should go ahead and do so ('If it can, then the judge should make such a direction under s.45A(1)').

The judgement in Vowles also laid out four "matters to which a judge will invariably have to have regard" (§51), when deciding on the appropriate disposal in cases where a hospital order was being contemplated. These are as follows:

(1) *the extent to which the offender needs treatment for the mental*

disorder from which the offender suffers, (2) the extent to which the offending is attributable to the mental disorder, (3) the extent to which punishment is required and (4) the protection of the public including the regime for deciding release and the regime after release.

When setting out these four matters, the judgement emphasised that judges should not "feel circumscribed by the psychiatric opinion" as well as the importance of considering a penal element to the sentence: "[t] here must always be sound reasons for departing from the usual course of imposing a penal sentence and the judge must set these out" (§51).

The present article considers the application of these "matters to which the judge will invariably have to have regard" which are, for present purposes, referred to as the four **Vowles statements**.

It should be noted that the perceived presumption in Vowles in favour of a s.45A order over a hospital order when sentencing mentally disordered offenders led to further guidance from the Court of Appeal being issued in [R v Edwards, 2018](#). In this further guidance, the court noted that a 'level of misunderstanding of the guidance offered in Vowles appears to have arisen as to the order in which a sentencing judge should approach the making of a s.37 or a s.45A order'. According to the judgement in [R v Edwards](#), the correct sentencing starting point was to consider if a hospital order may be appropriate. As part of that determination, a sentencing court was required to *consider* 'all the powers at its disposal including a s.45A order', but, contrary to how Vowles had been interpreted, it was not required to *prefer* a 45A order to a hospital order. There was no 'default setting of imprisonment' (§12).

The judgement in Edwards was nevertheless clear on the need for 'sound reasons' for departing from the usual course of imposing a sentence with a penal element and, whatever their differences of emphasis, therefore, the judgements in Vowles and Edwards agreed on the importance of considering the need for a penal element when sentencing and, therefore, the need for an assessment of the culpability of the mentally disordered offender as part of the sentencing exercise.

In the 2020 case of [R v Nelson, 2020](#), the Court of Appeal, following Vowles and Edwards, noted the need for a careful assessment of the culpability of the offender, an issue which it said would be factored in to answering numbers 2 and 3 of the Vowles Statements. In [Nelson](#), the court also noted the helpful guidance on assessing culpability published by the Sentencing Council ([Sentencing Council, 2020](#)).

It may be reasonable to imagine that these judicial developments might have led to a significant widening of use of the s.45A, but rates of its use appear⁴ to have remained low (NHS Digital, 2023). Very limited research exists to explain this; [Beech et al. \(2019\)](#) conducted a qualitative study which highlighted attitudes about its use within consultant psychiatrists, and the Royal College of Psychiatrists has indicated that it is "ethically problematic for experts to explicitly recommend a s.45A hybrid order because this amounts to recommending punishment" ([Royal College of Psychiatrists, 2023](#)). Whilst multiple authors have undertaken a legal analysis of case law (with Jill Peay's analysis ([Peay, 2015](#)) arguably particularly important, illuminating the importance of 'partial culpability' within the legal decision-making process) there appears to be no research considering the attitudes of judges and legal professionals towards the use of s.45A or indeed the Vowles statements which now underpin it. Ultimately, investigating judicial attitudes towards the s.45A is important, since judges, not psychiatrists, are the primary decision makers.

The four 'Vowles Statements' reflect some of the most crucial and difficult questions in the intersection between mental health and criminal sentencing; we note Peay's description of the hybrid order as representing the "peak of complexity" ([Peay, 2015](#)). As Peay points out, judges themselves are not immune from human cognitive bias, a fact

⁴ We use the word 'appear' deliberately. The NHS dataset indicates a notable change in the frequencies of different types of detentions after 2015–16 but enquiries with NHS Digital have indicated there is likely to be unreliability in the later statistics.

that may be especially important in the context of the complexity of the matters being weighed, and the relative simplicity of the final decision to be made (custody, s.37 or s.45A). Indeed, it is far from obvious *how* a judge is expected to weigh the apparently competing aims of punishment and treatment that underpin *Vowles*; for example, the ‘extent to which the offending is attributable to the mental disorder’ is presumably a judgement that could be heavily influenced by expert clinical opinion, but it is not something that the relevant medical practitioners are required to address in a report, and possibly a factor which clinicians may be reluctant to address in expert evidence. Similarly, the fourth statement, concerning the anticipated release regime, arguably requires an impossible degree of fortune telling; anticipating whether a particular person’s risk may be better managed by probation services or mental health services potentially many years in advance is an inherently difficult task. Moreover, because judges can only make a hybrid order when medical evidence has confirmed that the conditions for a s.37 order are met, then if such an order is recommended, there must, ipso facto, be a need for treatment (Vowles statement 1). Within this context, perhaps it is not surprising that judges have tended to follow the standard medical recommendation of a s.37 disposal when it is given.

The *Vowles* judgement indicates that some of these issues may be resolved through consideration of the particular mental health condition with which the defendant presents; the *Vowles* judgement reviewed (and appeared to accept) psychiatric evidence that “[a] hospital and restriction order under s.37/41 is more likely to be appropriate in a case where the mental disorder is a severe mental illness (particularly a psychotic illness or an organic brain disorder) rather than a personality disorder” (§50(iii)). However, the challenge with this argument is that the distinction between ‘severe mental illness’ and ‘personality disorder’ is not always entirely clear, and there is a significant degree of symptomatic overlap. For instance, auditory hallucinations occur in borderline personality disorder as well as schizophrenia, and clinical literature is moving towards an understanding of shared symptoms between conditions, which are, in turn, underpinned by complex and overlapping causal mechanisms. Moreover, as the cases of the individual appellants in *Vowles* demonstrates, diagnoses themselves are not necessarily always terribly reliable; in *Irving*’s case, ‘his learning disability was underestimated’; *McDougall*’s appeal was based on a changed diagnosis from depression to ‘schizophrenia or schizoaffective disorder’; *Coleman* had initially been diagnosed with personality disorder, but was subsequently diagnosed as having had schizophrenia ‘at the time of sentence’ and *Barnes*’ diagnosis changed from mild depression to learning disability and dissociative personality disorder. All of this is to make the general point that the questions contained within the *Vowles* statements are inherently difficult ones to answer, and thus open to uncertainty and inconsistency. However, in the context of a general indication in *Vowles* that indicates that psychotic conditions such as schizophrenia are more likely to be associated with treatment needs that can be met in hospital, it is reasonable to consider whether similar symptoms, described by different diagnoses, might be subject to different decisions against the *Vowles* statements.

More generally, the complexity of the analysis that judges must undertake leads to a context in which there is potential scope for inconsistency in judgements made about the *Vowles* statements; one judge may assess the ‘need for treatment’ to be high and another may assess it to be low, based on the same facts. It is unknown how judges might (differently) make use of agreement or disagreement with the individual *Vowles* statements in weighing their decision. Are some criteria weighted more highly than others? Are others altogether ignored? There is no existing research which provides an answer to these questions.

Unfortunately, conducting research with qualified judges is a practical challenge. The present study had initially sought to do so, and appropriate permissions were sought, but unfortunately pragmatically this was not possible and thus a proxy judicial sample was sought. This involved recruitment of an online sample of participants with legal qualifications or working in legal settings. Whilst the authors recognise the limits of such a sampling approach, we argue that in the absence of

almost any other published research considering decision-making against the *Vowles* statements, such a ‘proxy judicial’ sample provides a platform on which further research can build, hopefully with actual judicial samples.

1.1. Aims

The present study adopts an experimental vignette-based design to attempt to answer some of the following questions:

1. To what extent do judges / participants agree in relation to judgements made on the *Vowles* statements?
2. Do judges / participants give make different decisions about the *Vowles* statements, or about sentencing, if similar mental health symptoms are described with different diagnostic terms?
3. To what extent to the ratings on the *Vowles* statements relate to each other?
4. Does mental health information or decisions about the *Vowles* statements influence sentencing recommendations?

When originally conceived, the study also sought to consider whether the judge’s attitudes towards people with mental health conditions, and causality beliefs concerning the origin of mental health conditions, were also related to decision-making. However, subsequent analysis indicated that the measures used to assess these attitudes and beliefs were not reliable (see Method section) and this element is thus omitted.

2. Method

2.1. Participants

The ‘proxy judicial’ sample was recruited via the online recruitment platform Prolific (prolific.com). This platform has the advantage of checking prospective participants’ government approved ID, thus reducing the risk of bots influencing responses. Moreover, Prolific allows the recruitment of participants by professional background. We used this option to recruit participants from England and Wales who were classified within a superordinate category of ‘legal professionals’. As identified below, this included professionals who were both legally and not legally qualified, working or studying in and legal profession.

285 participants were recruited in the initial sample. Of these, 88 were excluded on the basis that they failed at least one of the ‘knowledge check’ questions (requiring the name of the defendant and the offence to be correctly chosen out of three options). A further 17 participants were excluded from the analysis because they had taken less than nine minutes to complete the study, which was not a feasible time in which to review the video and complete the questionnaires. One further participant was excluded because time taken was inordinately long (>24 h). In total therefore this left a sample of 179 for final analysis.

Based on survey responses, the ‘legal professional’ umbrella term included barristers and solicitors (together 26.8 % of the sample, $n = 48$), ‘other legal professionals’ which included legal executives (38 %, $n = 68$) and law students (8.4 %, $n = 15$), as well as professionals with non-legal expertise such as administrators working in the legal sector (26.8 %, $n = 48$). Participants worked in a diverse range of legal areas.

2.2. Materials

2.2.1. Video vignette

In the English and Welsh Crown court, once the defendant’s guilt is established, the prosecution barrister starts with an opening statement summarising the case for the prosecution. A defence barrister will subsequently have opportunity to submit a plea in mitigation, and there will additionally be the opportunity for psychiatric evidence to be presented. In order to simplify the vignette material, we produced a composite

vignette which include the key prosecution arguments and also considerations of relevance for the defence. Vignette material was used to produce a video, described below, which depicted the ‘summing up’ of legal arguments. The video was the only material shown to participants.

All vignettes used to create the video were identical except that references to the mental disorder varied by diagnosis: ‘Schizophrenia’, ‘Emotionally Unstable Personality Disorder’ (henceforth EUPD), or ‘Complex Mental Health Problem’. The first two terms were chosen to reflect potential diagnoses which the wider judgement in Vowles suggests should be subject to different judicial responses. The ‘Complex Mental Health Problem’ condition was designed to act as a ‘control’ diagnosis which had the characteristics of a formal diagnosis without a specific name. This was included to control for any potential specific labelling effects. There were three such references to the diagnosis within the vignette.

The full version of the vignette is found in Appendix 1. The offence depicted was one of Grievous Bodily Harm (GBH), contrary to s.18 of the Offences Against the Person Act 1861, to which the defendant pleaded guilty. This meant that there had been no trial, and so mental health considerations were being raised for the first time at sentencing. In the vignette, the defendant had been described as having arrived an hour late for work, which led to him losing his job as a labourer. The following day, the defendant returned to the construction site at which he worked and used a steel pole to seriously injure the victim, his previous boss. There were reports that he had been talking ‘quickly and incoherently’. Two psychiatrists had assessed him and confirmed the diagnosis alongside a range of symptoms (unstable emotions, paranoid thoughts, auditory hallucinations and impulsive behaviours). The selection of symptoms was intended to represent those which could be identified either in EUPD or Schizophrenia. Disruption in his early life history was summarised. Evidence of ‘odd behaviour, rushed speech and paranoid beliefs’ was corroborated by evidence from an employment advisor. Specific psychiatric opinion spoke to the key issues in the Vowles statements; the psychiatric evidence indicated that the mental health condition was relevant to the offence but that the connection was not entirely clear, and the defendant could ‘benefit from a period of treatment within a hospital environment’.

Three videos were recorded based on the three versions of the vignettes. The actor was a professional academic (Author 6) dressed in judicial clothing against a neutral background. Each video was just under 7 min long.

2.3. Measures

2.3.1. Measures assessing individual attitudes and beliefs

The a-priori hypotheses for the study involved the consideration of the potential role of individual attitudes and beliefs on the sentencing decision. The Mental Health Locus of Origin scale (MHLO; Hill & Bale, 1980) was used to assess beliefs about the causal factors of mental health concerns, and the Perceived Devaluation and Discrimination Scale (PDDS; Link, 1987) was used to assess public stigma towards people with mental health concerns. We calculated the Cronbach’s Alpha and a Confirmatory Factor Analysis (CFA) for both measures. For the MHLO, Cronbach’s alpha was 0.63 and for the PDDS it was 0.91. However, a Confirmatory Factor Analysis (performed via the lavaan package in R studio 4.4.0) indicated a very poor fit according to conventionally used cut-offs for both of these scales (Hu & Bentler, 1999) (MHLO: Comparative Fit Index (CFI) = 0.78, Standardised Root Mean Squared Residual (SRMR) = 0.093, Root Mean Squared Error of Approximation (RMSEA) = 0.093; PDDS: CFI = 0.89, SRMR = 0.092, RMSEA = 0.204). Both of these measures were thus removed from further analysis. The authors intend to examine the utility of these measures in more detail in a separate publication.

2.3.2. Vowles ratings (individual elements)

Each of the Vowles Statements highlighted above was presented and

participants were asked to rate the extent to which they agreed with that rating. Ratings were made on a seven-point scale with the anchor points being ‘Strongly Disagree’, ‘Disagree’, ‘Somewhat Disagree’, ‘Neither Agree nor Disagree’, ‘Somewhat Agree’, ‘Agree’ and ‘Strongly Agree’. The four variables which were derived reflected the four Vowles Statements, identified as follows:

Vowles 1: “The offender’s mental health requires treatment”.

Vowles 2: “The offence is attributable to the mental health disorder”.

Vowles 3: “The offence requires punishment”.

Vowles 4: “The protection of the public is significant when deciding release & regime of release”.

2.3.3. Sentence: overall decision

Finally, participants were asked to identify the sentence that they would be “most inclined to give” based on the information available. They could only choose one option. Ratings were on a 3-level categorical nominal variable: s.37/41, s.45A or Prison.

2.4. Procedure

The study was hosted on Qualtrics (www.qualtrics.com) and advertised to suitable participants via Prolific identified as ‘legal professionals’, as described above. Participants were recompensed £2.67 each, which was classified as ‘good’ based on the hourly payment rate expected by Prolific. Participants were randomly allocated to one of the three conditions where the diagnosis associated with the video was manipulated. Beyond this difference, all participants completed the study in the same way. Fig. 1 shows the flow through the study.

2.5. Analysis

Analyses were conducted in line with the aims as follows:

1. To what extent do judges agree in relation to judgements made on the Vowles statements?

To answer this question, the endorsement of each of the four Vowles statements was broken down using a simple frequency analysis. The median response for each item was also calculated. We made no a-priori definition as to characteristics of the frequency distribution which would represent good or poor agreement.

2. Do judges give make different decisions about the Vowles statements, or about sentencing, if similar mental health symptoms are described with different diagnostic terms?

To answer this question, chi squared analyses were conducted using the experimental condition as the grouping variable.

3. To what extent to the ratings on the Vowles statements relate to each other?

To answer this question, we intended to calculate correlation coefficients between the four Vowles statements. Spearman’s Rho was used due to observed skew in the data.

4. Does mental health information or decisions about the Vowles statements influence sentencing recommendations?

To answer this question, we completed a regression analysis using ‘sentence’ as the dependent variable. The steps completed are explained in the Results section.

The mean time taken to complete the study was 16 min and 20 s.

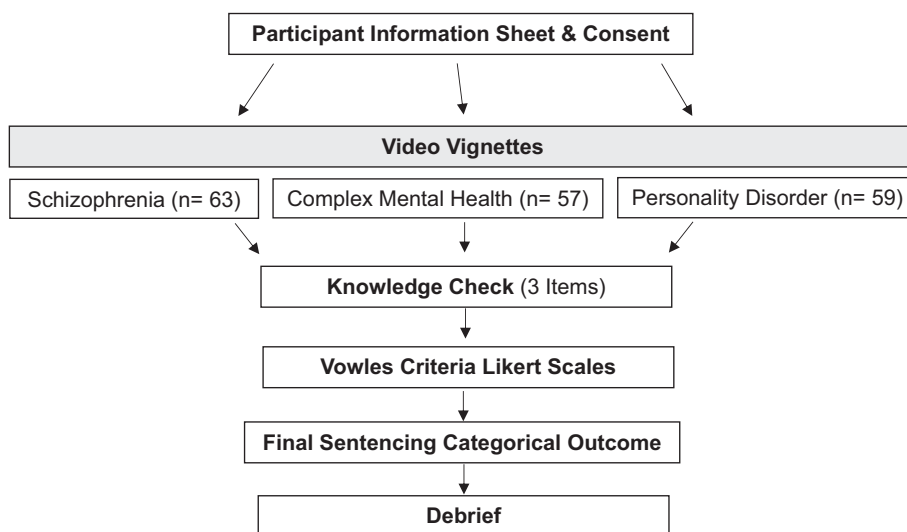


Fig. 1. Experimental study design flowchart (n = 179).

2.6. Ethics

The study was approved by the University of East Anglia, Faculty of Medicine and Health Sciences Ethics committee (reference 2020/21–067).

3. Results

Overall ratings against the four Vowles statements are indicated in Table 1. This suggests that, overall, there was a tendency towards endorsement for all four of the Vowles statements. For Vowles 1 (Requires Treatment), there was overall relatively strong agreement (Median response ‘Strongly Agree’) and very few participants showed evidence of disagreement. For Vowles 2 (Offending Attributable to MH) there was somewhat less clear agreement (Median response ‘Slightly Agree’). Vowles 3 (‘Requires Punishment’) and Vowles 4 (‘Protection of Public’) both had median responses of ‘agree’.

Vowles 1: “the extent to which the offender needed treatment for the mental disorder from which the offender suffered”; Vowles 2: “the extent to which the offending was attributable to the mental disorder”; Vowles 3: “the extent to which punishment was required; Vowles 4: “the protection of the public, including the regime for deciding release and the regime after release”.

Table 1 Overall response rates by Vowles condition, whole sample (n = 179).

Response	Vowles 1 (Requires Treatment)	Vowles 2 (Offending Attributable to MH)	Vowles 3 (Requires Punishment)	Vowles 4 (Protection of Public)
Strongly Disagree	4 (2.3 %)	2 (1.1 %)	1 (0.6 %)	2 (1.1 %)
Disagree	0 (0.0 %)	2 (1.1 %)	4 (2.3 %)	1 (0.6 %)
Somewhat Disagree	0 (0.0 %)	13 (7.4 %)	6 (3.4 %)	1 (0.6 %)
Neither Agree nor Disagree	2 (1.1 %)	19 (10.9 %)	8 (4.6 %)	2 (1.1 %)
Somewhat Agree	7 (4.0 %)	90 (51.4 %)	42 (24.0 %)	15 (8.6 %)
Agree	66 (37.7 %)	40 (22.9 %)	79 (45.1 %)	83 (47.4 %)
Strongly Agree	96 (54.9 %)	9 (5.1 %)	35 (20.0 %)	71 (40.6 %)

3.1. Differences in Vowles ratings between conditions

Frequencies of the responses to the Vowles statements, broken down by condition, are indicated in Table 2. Because of the skew evident in the distribution, and the associated small cell-sizes for some of the response options, for the purpose of this analysis, response options were collapsed into four categories; all disagree responses combined with the ‘Neither Agree nor Disagree’; ‘Slightly Agree’; ‘Agree’; and finally, ‘Strongly Agree’. None of the associated chi-squared tests were significant indicating that exposure to different diagnostic conditions did not have a detectable impact on Vowles ratings (Vowles 1: $\chi(4, N = 179) = 10.61, p = 0.10$; Vowles 2: $\chi(4, N = 179) = 3.86, p = 0.70$; Vowles 3: $\chi(4, N = 179) = 2.82, p = 0.83$; Vowles 4: $\chi(4, N = 179) = 9.01, p = 0.17$).

All questions preceded by “Please rate the extent to which you agree with the following statement...”: Vowles 1 Question: “The offender’s mental health requires treatment” (Legal wording: “the extent to which the offender needed treatment for the mental disorder from which the offender suffered”); Vowles 2 Question: “The offence is attributable to the mental health disorder” (Legal wording: “the extent to which the offending was attributable to the mental disorder”); Vowles 3 Question: “The offence requires punishment” (Legal wording: “the extent to which punishment was required”); Vowles 4 Question: “The protection of the public is significant when deciding release & regime of release” (Legal wording: “the protection of the public, including the regime for deciding release and the regime after release”).

3.2. Relationships between Vowles ratings

To establish the relationship between different Vowles ratings, correlation coefficients between all four Vowles criteria were calculated. Spearman’s rho (ρ) was used given the skewed distribution and use of ordinal categories. The correlation coefficients are shown in Table 3. This table indicates that there were three significant relationships at a conventional ($p < 0.05$) level of significance. The first was between Vowles 3 (Requires Punishment) and Vowles 4 (Protection of the Public; higher endorsement of Vowles 3 was associated with higher endorsement of Vowles 4). The second was between Vowles 2 (Offending Attributable to Mental Health) and Vowles 3 (Requires Punishment; higher endorsement of Vowles 2 was associated with lower endorsement of Vowles 3). The final was between Vowles 1 (Requires treatment) and Vowles 2 (Offending Attributable to Mental Health; higher endorsement of Vowles 1 was associated with higher endorsement of Vowles 2).

Table 2
Overall response rates by Vowles condition and exposure condition (% only).

Response	Vowles 1 (Requires Treatment)			Vowles 2 (Offending Attributable to MH)			Vowles 3 (Requires Punishment)			Vowles 4 (Protection of Public)		
	Schz	EUPD	CMH	Schz	EUPD	CMH	Schz	EUPD	CMH	Schz	EUPD	CMH
Strongly Disagree	4.8 %	0.0 %	1.8 %	1.6 %	0.0 %	1.8 %	0.0 %	0.0 %	1.8 %	3.2 %	0.0 %	0.0 %
Disagree	0.0 %	0.0 %	0.0 %	3.2 %	1.7 %	0.0 %	3.2 %	1.7 %	1.8 %	1.6 %	0.0 %	0.0 %
Somewhat Disagree	0.0 %	0.0 %	0.0 %	3.2 %	13.6 %	5.3 %	6.3 %	1.7 %	1.8 %	1.6 %	0.0 %	0.0 %
Neither Agree Nor Disagree	3.2 %	0.0 %	0.0 %	7.9 %	11.9 %	15.8 %	4.8 %	6.8 %	1.8 %	1.6 %	1.7 %	0.0 %
Somewhat Agree	4.8 %	3.4 %	3.5 %	54.0 %	49.2 %	49.1 %	22.2 %	23.7 %	24.6 %	11.1 %	10.2 %	5.3 %
Agree	28.6 %	49.2 %	40.4 %	22.2 %	20.3 %	24.6 %	46.0 %	40.7 %	47.4 %	41.3 %	44.1 %	54.4 %
Strongly Agree	58.7 %	47.5 %	54.4 %	7.9 %	3.4 %	3.5 %	17.5 %	25.4 %	21.1 %	39.7 %	44.1 %	40.4 %

Table 3
Spearman's Correlations between Vowles ratings.

	Vowles 1 (Requires Treatment)	Vowles 2 (Offending Attributable to MH)	Vowles 3 (Requires Punishment)
Vowles 2 (Offending Attributable to MH)	$\rho = 0.302, p < 0.001^*$		
Vowles 3 (Requires Punishment)	$\rho = -0.022, p = 0.733$	$\rho = -0.250, p < 0.001^*$	
Vowles 4 (Protection of Public)	$\rho = 0.061, p = 0.417$	$\rho = -0.132, p = 0.079$	$\rho = 0.341, p < 0.001^{**}$

3.3. Differences in sentences by condition

Table 4 shows the sentences given, broken down by condition. Overall, this indicates that there was a strong preference for s.45A disposals, and in total approximately three-quarters of sentences recommended were for s.45A. Whilst very few participants recommended a prison sentence, all of the four prison sentences given were in relation to exposure to the EUPD condition.

To establish whether there was a difference in sentencing behaviour by condition, a Chi Squared test was conducted, which was significant ($\chi(4, N = 179) = 9.48, p = 0.05$). However, because of the very small number of people who gave a prison sentence (and the resultant small cell counts) the Chi Squared test was repeated excluding respondents who gave a prison sentence. This was not significant ($\chi(2, N = 175) = 1.168, p = 0.56$). This analysis suggested that, overall, there was little evidence of a difference in sentences by condition, with some tentative evidence for slightly higher rates of prison sentences given to people in the EUPD condition.

3.4. Prediction of sentencing outcome by Vowles ratings and condition

To establish whether ratings against the four Vowles criteria and/or condition influenced sentencing decisions, a binary logistic regression was conducted. A multinomial logistic regression was not conducted due to the very small number of participants recommending a prison sentence, and instead these four cases were excluded from the regression analysis. The dependent variable thus represented whether the sentence was a s.37 or s.45A. The regression was conducted in three blocks, with 'condition' being entered in the first block, the four Vowles ratings being entered into the second block (these were entered as ordinal variables representing the full range of scores), and demographic variables age (continuous), gender (binary) and personal experience of mental health

Table 4
Sentence Outcomes by Exposure Condition.

Sentence	Complex Mental Health	EUPD	Schizophrenia
Prison	0 (0 %)	4 (6.8 %)	0 (0 %)
S37/41	14 (24.6 %)	15 (25.4 %)	12 (19.0 %)
S45A	43 (75.4 %)	40 (67.8 %)	51 (81.0 %)

difficulties (a four level ordinal variable rated 'none at all' (0), 'a little' (1), 'a moderate amount' (2), 'a lot' (3) or 'a great deal' (4)) entered in the final block.

The output of the Logistic regression is contained in Table 5. This indicates that in the first block, 'condition' was not associated with any significant effects, and the overall model explained a very small degree of variance in the dependent variable (Nagelkerke $R^2 = 0.007$). In the second block, Vowles 3 (Requires Punishment) was entered as a highly significant predictor, but none of the other Vowles items were significant predictors. The amount of variance explained by the model was much greater (Nagelkerke $R^2 = 0.22$). In the third block, the overall characteristics of the model were largely unchanged. None of the demographic variables were associated with individual significant effects, and Vowles 3 remained the only highly significant predictor. The amount of variance explained by the model was almost identical (Nagelkerke $R^2 = 0.23$).

The results, overall, suggest that the condition did not impact sentencing decisions, but that attitudes towards punishment (represented in Vowles 3) were strongly associated with sentencing decisions, even after controlling for all other variables, with greater endorsement of Vowles 3 being associated with a s.45A disposal as compared to a s.37

Table 5
Logistic Regression – Prediction of Sentence (s.37/41 or s.45A outcomes only).

	Model 1: Condition	Model 2: Model 1 + Vowles Terms	Model 3: Model 2 + Demographic Items
Nagelkerke R^2	0.007	0.222	0.225
Constant	$\beta = 1.075, SE = 0.310, p < 0.001$	$\beta = -3.903, SE = 2.313, p = 0.092$	$\beta = -3.293, SE = 2.499, p = 0.188$
Condition (Overall)	$p = 0.710$	$p = 0.360$	$p = 0.313$
Condition (CMH v EUPD)	$\beta = -0.178, SE = 0.474, p = 0.707$	$\beta = 0.266, SE = 0.513, p = 0.603$	$\beta = -0.296, SE = 0.517, p = 0.568$
Condition (CMH v Schizophrenia)	$\beta = 0.378, SE = 0.456, p = 0.408$	$\beta = 0.759, SE = 0.532, p = 0.154$	$\beta = 0.834, SE = 0.549, p = 0.128$
Vowles 1 (Requires Treatment)		$\beta = -0.225, SE = 0.245, p = 0.358$	$\beta = -0.220, SE = 0.244, p = 0.368$
Vowles 2 (Offending Attributable to MH)		$\beta = 0.046, SE = 0.213, p = 0.829$	$\beta = 0.023, SE = 0.217, p = 0.916$
Vowles 3 (Requires Punishment)		$\beta = 0.709, SE = 0.197, p < 0.001$	$\beta = 0.717, SE = 0.197, p < 0.001$
Vowles 4 (Protection of Public)		$\beta = 0.353, SE = 0.215, p = 0.100$	$\beta = 0.350, SE = 0.222, p = 0.115$
Age			$\beta = -0.009, SE = 0.023, p = 0.687$
Own Experience of Mental Health Concerns			$\beta = -0.124, SE = 0.225, p = 0.582$
Gender			$\beta = -0.128, SE = 0.501, p = 0.798$

disposal.

To control for potential multi-collinearity between Vowles statements, the model was repeated as described, but with significant Vowles predictors removed on successive steps. On the first occasion, Vowles 3 was removed. The resultant model was a much less good fit overall (Nagelkerke $R^2 = 0.09$), but Vowles 4 became a significant predictor ($B = 0.55$, $SE = 0.21$, $p = 0.008$). Vowles 4 was thus removed, and the model repeated, but on this occasion neither Vowles 1 nor Vowles 2 contributed to the model. Therefore, no further iterations were conducted.

The results overall suggest that ratings on Vowles 3 (Requires Punishment) appear to be the most important in determining a decision between a s.45A and a s.37 disposal. Vowles 4 (Protection of the Public) did seem to play a role in such decisions, but its effect was not significant after Vowles 3 was controlled for. Finally, the results indicated that ratings on Vowles 1 and 2 ('Need for Treatment', 'Offending Attributable to MH') were not important factors in deciding between s.37 or s.45A.

3.5. Did legally qualified participants make different decisions?

In order to consider the potential that legally qualified participants responded differently to other participants, we conducted post-hoc chi-squared tests exploring whether there were differences in ratings on the main Vowles statements, as well as the final sentencing decision, on the basis of the participant being legally qualified or not. None of these analyses showed any significant effects, suggesting that participants responded similarly irrespective of legal qualification (Output available at https://osf.io/cdqnj/?view_only=403bc025b3de47be9bd12ef56727bb11).

4. Discussion

The research sought to better understand how decisions are made against the Vowles statements which form a significant component of the case-law structuring decisions about criminal sentencing for defendants with mental health conditions. By using a vignette design with an experimental manipulation on diagnosis, we were able to assess whether different decisions would be made where the same mental health symptoms were labelled differently. Each of the main aims of the study is considered in turn.

4.1. Agreement between participants

The analysis of responses indicates that, in general, there was much more agreement than disagreement for each of the Vowles statements. Vowles 1 ('Requires Treatment') showed the clearest pattern of agreement, with almost all responses clustered around 'Agree'/'Strongly Agree'. This is arguably unsurprising in the context of the statement in the vignette that '[b]oth experts agree that Mr Smith could benefit from a period of treatment within a hospital environment'. As we have highlighted earlier, because a s.37 recommendation is required to consider a s.45A disposal, it is hard to see circumstances where a judge would not endorse the 'Requires Treatment' element of Vowles in practice. In fact, it is perhaps surprising that, in this context, there were any participants who did not endorse an 'agree' recommendation.

There was also a similar level of agreement apparent in the responses to Vowles 4, with again almost all responses clustered around 'Agree'/'Strongly Agree'. This perhaps reflects the focus on the wording of this statement which relates both to the protection of the public in general and the 'regime for deciding release and the regime after release'. It may have been clearer for participants to think about this more explicitly in terms of a choice between future management by probation vs mental health services. As worded, it is arguably unsurprising that participants agree that considerations of public protection are important at the sentencing stage, but it is much harder to see how this determination *can* be sensibly made potentially years in advance of the actual progression

to the community. Notably, the Vowles judgement includes reflection on this very issue, with Lord Thomas stating (§35) that "[i]t would be much better if there was a single judicial body which could decide at the time release is being considered the terms of release and the appropriate regime for supervision in the community, rather than this determination being made years earlier without the benefit of all the evidence gathered in the intervening years". We would agree that the present situation arguably does not well account for the complex and uncertain effect of treatment over the medium and long term on mitigating risk and supporting public protection, although would also note that follow-up by community mental health services for offenders released from secure psychiatric care has been associated with much lower reoffending risk than for prisoners followed up by probation services (Fazel et al., 2016). However, it is unclear whether this is because of the additional and specialist follow-up associated with the s.37/41, a difference in the characteristics of the people who are given a s.37/41 vs a custodial sentence, greater institutional control over the process of discharge/release, or a combination of these factors. Further research to differentiate between these possibilities might better inform the options for decision-making around potential sentencing options and aligning the best post-release conditions to those which are most likely to manage subsequent risk.

There was also reasonable agreement for Vowles 3 (requires punishment), although here a more sizable number of the participants (approx. 10 %) did not express agreement. Moreover, the 90 % who did select an agree option gave a wider selection of responses between 'somewhat agree' and 'strongly agree', suggesting the potential for more variance on this point. This variance however might well be explained by different responses to the individual facts of the case, but could also potentially be explained by variance in underlying attitudes towards punitiveness, which are noted to differ substantially in the wider population (Spiranovic et al., 2012). It is not known whether the same factors which influence attitudes towards punishment in the wider population also apply to judges, and if so, whether judges are better able to mitigate the effect of such attitudes in applying sentences in the individual case. However, on this point, Adriaenssen and Aertsen (2014) highlights that there is limited evidence suggesting that judicial attitudes towards punitiveness differ significantly from the general public, and that some research has highlighted more punitive attitudes in judges compared to the general population. Certainly, it would seem that there is a risk that judicial consideration of this item could be influenced by general attitudes towards punitiveness, and this may be an important point for future research to explore.

Finally, Vowles 2 ('Offending Attributable to Mental Health') showed a clearer pattern of disagreement between participants. Whilst the majority of responses tended towards 'agree', there were fewer 'Strongly Agree' responses and approximately 20 % of the sample selected a response suggesting uncertainty or disagreement. It is worth briefly reflecting that this statement probably reflects the core nexus of the Vowles judgement, and such uncertainty is therefore potentially rather important. Such uncertainty could arise for a number of reasons. First, despite the complexity of the potential question, judges are not provided with specific guidance as to how this is to be assessed. Is, for example, the offence 'attributable' to mental health simply by evidence that it occurred during a mental health episode? Or does there need to be a direct explanatory line between the person's mental health and the offence, e.g. at the symptom level. What if a mental health condition is confidently considered a partial explanation? However, the relative uncertainty could also reflect differences in attitudes or attributions at a between-subjects level. Variations in such attitudes in judges has been identified in the literature (Batastini et al., 2017) but future research is needed to identify whether such attitudes influence decision-making.

4.2. Do participants make different decisions about the Vowles statements, or about sentencing, if similar mental health symptoms are described with different diagnostic terms?

One of the clear findings from this research was that the diagnostic manipulation, overall, appeared to make little difference to decision-making, either in terms of the judgements made against the Vowles statements or in terms of the final sentence chosen. This is interesting in the context of the medical evidence cited in Vowles which implied that a Hospital Order may be a more appropriate for a 'severe mental illness... rather than a personality disorder'. This however suggests that participants, and consequently judges, may not make this decision based on the overall diagnosis presented. Beyond the lack of difference between schizophrenia and EUPD, it is perhaps particularly remarkable that abandoning the use of a formal diagnosis altogether and simply using the term 'Complex Mental Health Problem' seemed to make little difference to decision-making. This seems to run counter to the importance on diagnosis implied in the Vowles judgement as a whole (though see Hyman (2010) for an excellent review of the problems that arise from this emphasis), and also counter to wider research highlighting that judges express preferences for diagnostic information above wider clinical information in expert witness testimony (Redding et al., 2001). One possibility not tested by the argument was whether the presence of any diagnosis is important for decision making; a possibility for future research might be to include a condition with no diagnosis at all included. If such research also found no variation between conditions, it might highlight a gap between the information which judges state is important, and what they actually use in practice to make sentencing decisions.

Of course, the research does not illuminate which mental health information (e.g. symptoms, treatment types etc) may have been more important (or may become more important) in terms of decision-making; nor does it mean that diagnostic terms might not become important in different contexts. However, it does allow us to make a tentative recommendation that clinicians providing evidence at the sentencing stage should be careful not to depend on diagnostic terms in their explanation of evidence, and should instead ensure that they provide a fuller description including development and progression of symptoms, the functional impact of those symptoms, and the range of potential treatment options.

More widely, this finding may imply that courts and judges need to be better informed as to models of mental distress that extend beyond a classical diagnostic approach. As is, the Vowles judgement shows that conceptualisation of complexity in mental distress is at least partly seen in terms of co-morbidity, i.e. having more than one specific or discrete disorder at one time. For instance, at §50 the medical evidence accepted indicates that "[i]t is very rare for a person to have solely a psychotic illness such as schizophrenia or solely a personality disorder... It is usual for a person suffering from psychosis also to have a personality disorder and/or drug and alcohol problems". However, clinical literature and research is moving away from this form of conceptualisation; more recent approaches to taxonomy such as the Hierarchical Taxonomy of Psychopathology, or HiTOP (Kotov et al., 2017), and the description of Personality Disorder adopted in ICD-11, are emphasising the importance of dimensionality and focusing instead on understanding mental distress in terms of groups or clusters of symptoms using classification models that are better informed by research than our current taxonomies. Such an understanding is potentially an enormous change for judges and courts to grapple with, but ultimately may provide a more meaningful basis for judges to base complex decisions about issues such as culpability (Carroll et al., 2022; Mulay et al., 2024). Worryingly, recent data has suggested, however, that clinicians involved in evaluations for court are not terribly familiar with, and hardly ever use, such approaches (Mulay et al., 2024).

4.3. To what extent do the ratings on the Vowles statements relate to each other?

There were three significant relationships between ratings on the Vowles statements. Two of these were arguably quite predictable: the negative relationship between Vowles 2 ('Offending Attributable to MH') and Vowles 3 ('Requires Punishment') suggests that participants tended to ascribe less culpability where they attributed the offence to the person's mental health condition. Similarly, the positive correlation between Vowles 1 ('Requires Treatment') and Vowles 2 ('Offending Attributable to MH') is unsurprising. The relationship between Vowles 3 ('Requires Punishment') and Vowles 4 ('Protection of the Public') is arguably more interesting, potentially suggesting that participants at least partially conflated the aims of punishment and public protection. Such conflation would not be inconsistent with existing research which has highlighted the role of perceived dangerousness in sentencing decisions; presumably perceptions of dangerousness are closely related to perceptions of a need for public protection (Kruis et al., 2023; Sanderson et al., 2006). Interestingly, however, such conflation would be inconsistent with the spirit of the judgement given in a recent high profile appeal case (the appeal being brought by the solicitor general on the basis of being unduly lenient) where Lady Carr concluded that "[t]he need for punishment is a function of the seriousness of the consequences of the offending and the level of responsibility which the offender bears for that offending" (*R v Calcocane (Valdo)*, 2024, §80).

Notably, there was no correlation between Vowles 1 and 3, suggesting that participants consider issues of punishment and treatment independently of each other and not as dichotomies. The lack of correlation between Vowles 1 and 4 and Vowles 2 and 4 is however of note, since this suggests, as above, that mental health factors (including treatment) are not prioritised in the consideration of future risk management. It might be speculated that the release regime via community health services is perceived as a somewhat 'softer' regime, although clearly, where mental health factors drive ongoing risk, this regime is likely to be more suitable for management of that risk.

4.4. Does mental health information or decisions about the Vowles statements influence sentencing recommendations?

At the outset, it must be acknowledged that our ability to answer this question fully is somewhat limited since very few participants selected a prison sentence. Thus, the question that can be answered relates to the extent to which mental health information or decision about the Vowles statements influence a decision to impose a s.37 Hospital Order vs a s.45A hybrid order.

In this regard, it appears that the experimental condition of diagnosis does little to influence judgements. The only observation here is that the only participants who accorded a prison sentence were in the EUPD condition, but the numbers here were small and not statistically significant.

More interestingly, however, is the observation that Vowles 3 ('Requires Punishment') and to a lesser extent Vowles 4 ('Protection of the Public') were the only Vowles statements that were statistically significant in predicting sentencing decisions. This suggests that perceptions of culpability, or possibly perceptions of dangerousness (Kruis et al., 2023; Sanderson et al., 2006), are the primary basis on which judges may make a decision between s.45A or s.37. Whilst this is not outside the spirit of the Vowles judgement when taken as a whole, it is also surprising that these wider judgements about the person's mental health condition seemed to be unimportant in making a decision. Whilst recognising that generalisable conclusions are tricky given the limited data, this suggests that judges may be more likely to impose a s.37 when they feel that an offender does not require punishment, rather than because they perceive the offender to require treatment. It is possible that this behaviour broadly reflects the prioritisation of perceptions of dangerousness in the decision-making process.

4.5. Conclusions and recommendations

The present paper provides a starting point for future research considering how judges make decisions about defendants with mental health conditions. It provides, to the authors' knowledge, the first piece of research exploring such processes in relation to the judgement in *R v Vowles*. Nonetheless, it is important to directly acknowledge the inherent limitations of the sampling method used – including both the use of a non-judicial sample, and significant potential heterogeneity in the legal knowledge within the participants who were recruited. Taken together, this means that it is hard to make generalisations about how judges will make decisions in practice, and in particular, how different types of information – beyond that considered in the present study – may be important to judicial decision-making processes. In particular, it will be important to consider whether individual differences between judges (in terms of attitudes towards mental health conditions, stigma, causality, and treatment related beliefs) could influence judgements, something that was not possible in the present study. Moreover, in the real courtroom, numerous other variables – including the defendant's actual behaviour and presentation in the court setting – may well influence judgements. Nonetheless, we argue that in the context of the wider absence of research regarding the *Vowles* statements, the study makes a useful contribution, and highlight in particular the conclusion in a review by [Goodman-Delahunt and Sporer \(2010\)](#) that “comparisons of decisions by judges, juries and mock-judges and juries have revealed remarkably similar outcomes”. Indeed, we did not identify differences in ratings provided by the legally and non-legally qualified participants.

Regardless, it would seem that whilst there is reasonable agreement between participants in the vignette scenario in relation to three of the *Vowles* criteria, there is somewhat less agreement in relation to judgements made against *Vowles* 2 ('Offending related to MH'). This may simply reflect tension between the inherent clinical considerations in this question and the non-clinical expertise which judges bring, potentially combined with a focus in expert evidence on the presence of mental disorder (as required by s.37) rather than the nature of its linkage with the person's offending behaviour. Understanding the linkages between mental health concerns and offending is a complex task, even on a population level (e.g. [Nederlof et al., 2013](#); [Yu et al., 2012](#)). At the individual level required in a criminal case, and combined with the likely uncertainty in their own clinical judgements, the research also indicates that judges may be less reliant on the mental health elements of the *Vowles* statements when making their decision. Future research is important, but it is perhaps reasonable to conclude that to best assist judges in making decisions, clinicians providing expert evidence on sentencing matters should consider explicitly addressing the possible links between a person's mental health condition and their offending behaviour and future risk, and their degree of confidence in their opinions about this.

One final observation, however, is that in the present study, the s.45A outcome was particularly popular – overall around three-quarters of the sentencing recommendations were for a s.45A. This, of course, does not reflect the real situation in courts where the s.45A seems to be hardly used. There may be many reasons for this difference, but one intriguing possibility is that in the context of the research study, distanced from physical contact with an actual psychiatrist, and providing recommendations privately to a computer, is an inherently easier context for a judge to decide to make a s.45A, which inherently requires the judge to go 'above and beyond' a psychiatrist's medical recommendation. Another potential, of course, is that participants simply saw the s.45A as a 'middle option' that allowed them to resolve the ambiguity between the distinct aims of punishment and treatment, perhaps also influenced by the fact that our sample likely lacked expertise and confidence in judicial or criminal decision-making. Nonetheless, it does not seem impossible that judicial attitudes towards the s.45A are more positive than those of psychiatrists, but the professional position of psychiatrists to never recommend a s.45A

provides a context in which they will never or rarely do so.

These are undoubtedly preliminary findings and there are plenty of opportunities for future research studies to develop these preliminary findings further. Certainly, future research studies should consider not only the possibility of improving the representativeness of the sample being used (in particular seeking to recruit 'real judge' samples) but should also consider improving the representativeness of the decision-making process, for instance by assessing decision-making in an environment closer to that of a real courtroom. If further research adopts similar online recruitment methods, we recommend that particular attention is given to methodological enhancements which would support drawing a sample that would have less heterogeneity in legal knowledge, for example through recruiting only legally qualified participants, or making further efforts to recruit real judges. To achieve this, consideration would need to be given to development of the recruitment strategy, including recompense. Nonetheless, the broad lack of difference in ratings provided by the legally and non-legally qualified parts of our sample suggests that meaningful data can be obtained from samples recruited using less robust recruitment methods. Apart from improvements in recruitment and sampling, methodological quality could be improved through further detail given towards improvement of ecological validity of the materials: whilst video-vignette designs are convenient for experimental methodologies, they do not well replicate the wider courtroom environment and the many other factors that could influence decision-making in the real world.

Beyond the scope of the present study, there are a number of apparent lines of research which would be useful to explore. For instance, it would be helpful to consider research which addressed a wider range of factors which might influence judicial sentencing decisions in the context of mental health information. Beyond *Vowles*, judges are expected to give regard to a number of other pieces of case-law, formal sentencing guidelines, and of course apply these considerations to the individual facts of the case before them. More broadly, research to better understand judicial knowledge, awareness and beliefs about mental health concerns would be of enormous value in enhancing the fairness of sentencing for offenders with mental health concerns.

Finally, there would be scope to consider research to develop our understanding of the process of developing recommendations and decision-making at the psychiatric level; such research could potentially help understand differences in opinion and views within psychiatrists, but could also inform the understanding of the degree of alignment between psychiatric and judicial attitudes. This knowledge would be of vital importance in any subsequent efforts to reform sentencing procedures, whether by case-law or statute; as the present research itself suggests, if psychiatrists are unwilling to make a particular recommendation, and judges are reluctant to reach decisions that contrast psychiatric recommendations, developments in case-law or statute may not be followed by significant changes in practice.

CRedit authorship contribution statement

George Baldwin: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Samantha Young:** Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lucy Fitton:** Writing – review & editing. **Ian Edwards:** Writing – review & editing, Supervision, Conceptualization. **Michael Butler:** Writing – review & editing. **Peter Beazley:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijlp.2025.102071>.

References

- Adriaenssen, A., & Aertsens, I. (2014). Punitive attitudes: Towards an operationalization to measure individual punitivity in a multidimensional way. *European Journal of Criminology*, 12(1), 92–112. <https://doi.org/10.1177/1477370814535376>
- Batastini, A. B., Lester, M. E., & Thompson, R. A. (2017). Mental illness in the eyes of the law: Examining perceptions of stigma among judges and attorneys. *Psychology, Crime & Law*, 24(7), 673–686. <https://doi.org/10.1080/1068316x.2017.1406092>
- Beech, V., Marshall, C. M., Exworthy, T., Peay, J., & Blackwood, N. J. (2019). Forty-five revolutions per minute: A qualitative study of hybrid order use in forensic psychiatric practice. *The Journal of Forensic Psychiatry & Psychology*, 30(3), 429–447. <https://doi.org/10.1080/14789949.2019.1588910>
- Carroll, A., Walvisch, J., & Marsh, T. (2022). Personality disorders and forensic assessments: The benefits of ICD-11. *Medicine, Science, and the Law*, 62(4), 285–291. <https://doi.org/10.1177/00258024221094188>
- Delmage, E., Exworthy, T., & Blackwood, N. (2015). The 'hybrid order': Origins and usage. *The Journal of Forensic Psychiatry & Psychology*, 26(3), 325–336. <https://doi.org/10.1080/14789949.2015.1017591>
- Fazel, S., Fimińska, Z., Cocks, C., & Coid, J. (2016). Patient outcomes following discharge from secure psychiatric hospitals: Systematic review and meta-analysis. *British Journal of Psychiatry*, 208(1), 17–25. <https://doi.org/10.1192/bjp.bp.114.149997>
- Goodman-Delahunty, J., & Sporer, S. L. (2010). Unconscious influences in sentencing decisions: A research review of psychological sources of disparity. *Australian Journal of Forensic Sciences*, 42(1), 19–36. <https://doi.org/10.1080/00450610903391440>
- Hill, D. J., & Bale, R. M. (1980). Development of the mental health locus of control and mental health locus of origin scales. *Journal of Personality Assessment*, 44(2), 148–156. https://doi.org/10.1207/s15327752jpa4402_5
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology*, 6, 155–179. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091532>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Kruis, N. E., Ménard, K. S., Choi, J., Rowland, N. J., Frye, T., Kosaka, R., & Williams, A. (2023). Perceived dangerousness mediates punitive attitudes toward sex offenders: Results from a vignette experiment. *Crime & Delinquency*. <https://doi.org/10.1177/00111287231170106>
- Laing, J. M. (1996). The proposed hybrid order for mentally disordered offenders — A step in the right direction? *Liverpool Law Review*, 18(2), 127–146. <https://doi.org/10.1007/bf02486519>
- Link, B. G. (1987). Understanding labeling effects in the area of mental disorders: An assessment of the effects of expectations of rejection. *American Sociological Review*, 52(1). <https://doi.org/10.2307/2095395>
- Mulay, A. L., Gottfried, E. D., Ruchensky, J. R., Russell, T., Natoli, A. P., & Hopwood, C. J. (2024). The problem no one is talking about: Forensic evaluators' lack of familiarity with dimensional approaches to personality and psychopathology. *Journal of Personality Assessment*, 1–9. <https://doi.org/10.1080/00223891.2024.2420172>
- Nederlof, A. F., Muris, P., & Hovens, J. E. (2013). The epidemiology of violent behavior in patients with a psychotic disorder: A systematic review of studies since 1980. *Aggression and Violent Behavior*, 18(1), 183–189. <https://doi.org/10.1016/j.avb.2012.11.018>
- NHS Digital. (2023). Mental health act statistics, annual figures, 2022–23. <https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-act-statistics-annual-figures/2022-23-annual-figures#chapter-index>.
- Peay, J. (2015). Sentencing mentally disordered offenders: Conflicting objectives, perilous decisions and cognitive insights. *LSE Working Papers*. <https://eprints.lse.ac.uk/61581/1/Sentencing%20Mentally%20Disordered.pdf>.
- Peay, J. (2016). Responsibility, culpability and the sentencing of mentally disordered offenders: Objectives in conflict. *Criminal Law Review*, 3, 152–164.
- R v Calcoane (Valdo). (2024). *EWCA Crim 490*.
- R v Edwards. (2018). *EWCA Crim 595*.
- R v Nelson. (2020). (*EWCA Crim 1615*).
- R v Vowles, & others. (2015). *EWCA Crim 45 2015*.
- Redding, R. E., Floyd, M. Y., & Hawk, G. L. (2001). What judges and lawyers think about the testimony of mental health experts: A survey of the courts and bar. *Behavioral Sciences & the Law*, 19(4), 583–594. <https://doi.org/10.1002/bsl.455>
- Royal College of Psychiatrists. (2023). Responsibilities of psychiatrists who provide expert evidence to courts and tribunals. https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr193.pdf?sfvrsn=c0381b24_2.
- Sanderson, C. A., Zanna, A. S., & Darley, J. M. (2006). Making the punishment fit the crime and the criminal: Attributions of dangerousness as a mediator of Liability1. *Journal of Applied Social Psychology*, 30(6), 1137–1159. <https://doi.org/10.1111/j.1559-1816.2000.tb02514.x>
- Sentencing Council. (2020). Sentencing offenders with mental disorders, developmental disorders, or neurological impairments. <https://www.sentencingcouncil.org.uk/overarching-guides/magistrates-court/item/sentencing-offenders-with-mental-disorders-developmental-disorders-or-neurological-impairments/>.
- Spiranovic, C. A., Roberts, L. D., & Indermaur, D. (2012). What predicts punitiveness? An examination of predictors of punitive attitudes towards offenders in Australia. *Psychiatry, Psychology and Law*, 19(2), 249–261. <https://doi.org/10.1080/13218719.2011.561766>
- Whyte, S., & Gupta, J. (2007). Widening the scope of the hybrid order. *Medicine, Science, and the Law*, 47(1), 41–44. <https://doi.org/10.1258/rsmmsl.47.1.41>
- Yu, R., Geddes, J. R., & Fazel, S. (2012). Personality disorders, violence, and antisocial behavior: A systematic review and meta-regression analysis. *Journal of Personality Disorders*, 26(5), 775–792. <https://doi.org/10.1521/pedi.2012.26.5.775>