Consistent and Transparent? Personality Disorder Diagnosis, the Sentencing of Offenders with Mental Health Difficulties, and the Relevance of Stigma and Psychiatric Diagnosis.

Samantha Young

Candidate Registration Number: 100299501

Primary Supervisor: Dr Peter Beazley

Secondary Supervisor: Dr Ian Edwards

A thesis submitted in partial fulfilment of the degree of

Doctorate in Clinical Psychology

Faculty of Medicine and Health Sciences

Norwich Medical School, University of East Anglia

24th May 2022

Word Count: 32,559 (excluding appendices)

Thesis Abstract

Aims: This portfolio aims to aid the understanding of psychiatric diagnosis, mental illness stigma and the sentencing of offenders with mental health difficulties (OMHDs).

Method: The thesis portfolio contains two main papers; a systematic review and meta-analysis which reviews the interrater reliability (IRR) of the novel Alternative Model of Personality Disorder (AMPD; DSM-5) and an empirical online project investigating whether psychiatric diagnosis, stigmatic attitudes and guidance laid down in *R v Vowles* can predict sentencing outcomes for OMHDs.

Results: Fifteen studies were included in the review. Meta-analysis provided tentative support for Criterion A of the AMPD and its domains with pooled ICCs above acceptability levels for the DSM and previous estimates for the current diagnostic framework. Subgroup analysis suggests IRR could be improved by using a structured clinical interview designed for the AMPD. Variation in IRR was found across domains. However, heterogeneity was high across all analyses and results should be interpreted with caution. Few studies were found examining the IRR of Criterion B and overall PD diagnosis; however initial studies show mixed results. In the empirical project, only one *Vowles* Criteria, "the extent to which the offender requires punishment" was a significant predictor in the model. However, the overall model was not significant, therefore results should be interpreted with caution. The remaining *Vowles* Criteria, stigmatic attitudes and psychiatric diagnosis were not found to be significant predictors of sentencing outcome.

Conclusions: Overall the portfolio demonstrates tentative support for IRR of Criterion A of the AMPD. Further research is recommended for Criterion B and overall PD diagnosis alongside more ecologically valid methods using structured clinical interviews. Only one *Vowles* criteria was found to be a significant predictor of sentencing outcome, however

limitations of the study were identified and discussed. Replication with a judicial population and a further focus on expert witness testimony is advised.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my primary supervisor, Dr Peter Beazley. In addition to your wealth of knowledge, you have shown incredible warmth, support, and patience towards me over the past three years. If I can be half the psychologist you are, I will be twice the psychologist I ever thought I could be.

I would also like to thank my secondary supervisor, Dr Ian Edwards, for providing a much-needed insight into the unfamiliar and complex world of Mental Health Law.

Thank you also to my fellow trainee colleague and "thesis buddy", George Baldwin. We did it! It has been an absolute pleasure to share this thesis journey with you. I am forever grateful for all the friends I have made whilst completing this course. The interpretive dance said it all – you have all been the sunshine on this stormy, stormy journey! I can honestly say I've made friends for life, and I will never forget this experience.

Thank you also to the many participants who took part in our research. It has been encouraging to see so many people interested in forensic mental health.

Finally, thank you to the people I love the most. Thank you for bearing with me whilst I have been largely absent in your lives whilst writing this thesis. Normal life will (hopefully) resume soon.

To my parents, I am forever grateful for all the opportunities you have provided me with. All my success is down to you and the sacrifices you made to get me where I am today. Dad, I promise I will stop being a student now and get a "proper" job!

Most of all, thank you to my partner, Manoj. You have always believed in me and celebrated my success, even when embarking on doctoral training meant moving miles away from you. You know someone truly loves you when they're willing to spend their evening on a National Express coach, just to see you for a night. You have experienced every inch of this doctorate with me, and this thesis is as much yours, as it is mine.

Chapter 1: Introduction to the Thesis Portfolio

In England and Wales, it is estimated that 65% of sentenced male prisoners meet the criteria for a Personality Disorder (PD) diagnosis (NICE, 2017). Moreover, research suggests anywhere between 52% (HM Chief Inspector of Prisons, 2021) and 90% (Singleton et al., 1998) of offenders have a "diagnosable mental health condition". However, despite high prevalence rates, mental health provision within the prison estate is poor. In fact, in a recent investigation by the Justice Committee, both the Ministry of Justice and the National Health Service admitted to not knowing the nature or extent of these difficulties due to a shortage of data and lack of resources (House of Commons Justice Committee, 2021). This investigation also raised concerns around the treatment of offenders with a diagnosis of PD. They found many of these offenders were denied hospital admissions due to a belief that their displayed behaviours were not derived from "treatable mental health conditions" (House of Commons Justice Committee, 2021). This practice goes directly against current NICE Guidelines which recommend evidence-based treatment for PD (NICE, 2009). For those offenders that are deemed to have a "treatable mental health condition", transfer times from prison to mental health units remain high, despite statutory time limits, due to a shortage of mental health inpatient beds (NHS Benchmarking Network, 2018). Therefore, it is critical that offenders who require the input of secure mental health services at the point of sentencing, receive this sentence in a consistent and transparent way.

In England and Wales, the two main stakeholders in this sentencing process are judges who pass sentences and court appointed medical doctors who assess these offenders. Although the responsibility for sentencing lies solely with a judge when an offender is found guilty of a crime, if an offender is (or appears to be) suffering from a mental illness at the time of an offence or during sentencing, judges are legally obliged to obtain and consider a medical report before imposing a custodial sentence (s.232 of the Sentencing Act 2020). If

two medical doctors agree an offender has a mental illness (within the meaning of s.1 of the Mental Health Act 1983 [MHA]) a judge can deviate from usual penal sentences and provide important mental health treatment sentences including the non-custodial Hospital Treatment Order (s.37 MHA) with or without restrictions (s.41) and the Hospital and Limitation Direction (s.45A). Therefore, consistent, transparent and accurate clinical assessment, observation and diagnosis of offenders during the sentencing process is essential.

Psychiatric assessment and diagnosis is a widely debated topic that arguably extends beyond the scope of this doctoral thesis. The Diagnostic and Statistical Manual for Mental Disorders (DSM-I; American Psychiatric Association, 1952) was introduced 70 years ago. The manual aimed to improve the communication and comprehension of mental health problems by developing a shared language for psychiatrists to use (Garand et al., 2009). Arguably, by operationalising definitions of psychological distress, the DSM has stimulated research and led to more evidence-based practice within mental health care (Lam et al., 2016). However, despite decades of work and multiple revisions, psychiatric diagnosis remains a controversial construct. Diagnostic systems have been criticised for concerns regarding reliability, validity and clinical utility (Aboraya, 2007; Bentall, 2003). These concerns have increased over recent years following the introduction of the Fifth Edition of the manual (DSM-5; APA,2013) and reports of below acceptable levels of reliability in DSM field trials (Frances, 2012; Kinderman et al., 2013).

In 2013, the Division of Clinical Psychology released a position statement which critiqued diagnostic frameworks and cautioned against using the medical model to pathologise distress (British Psychological Society, 2013). They also highlighted diagnostic stigma, which can impact on clinical decision making (Lam et al., 2016) and act as a barrier to accessing services (Dale et al., 2017). Link and Phelan (2001) describe a four-component model of stigmatisation: (1) identifying and labelling differences from social norms, (2)

associating these differences with negative and discrediting stereotypes, (3) separating between stigmatiser and stigmatised (or "us and them") and (4) status loss and discrimination. This model of stigma is highly relevant to the process of psychiatric diagnosis, where this label is given by those in power to individuals who are often from minoritized groups, and thus already discriminated against within society. The medical model neglects to address these social inequalities, which often underlie mental health difficulties (Kinderman, 2013). Consequently, the Power, Threat, Meaning Framework (Johnstone & Boyle, 2018) was introduced which proposed an alternative to psychiatric diagnosis incorporating this broader social context. Whilst this alternative framework has gained much support, it is recognised that this view is not the mainstream opinion. Diagnostic systems such as the DSM-5 (APA, 2013) and the International Classification of Diseases and Related Health Problems (11th ed.; ICD-11; World Health Organisation, 2018), continue to exert significant influence in psychiatric practice and underlie the delivery of mental health services in the Western World. If they are to continue to be used, efforts to consider how, and under what situations they might be most reliably applied, remain helpful.

PD is arguably one of the most criticised psychiatric diagnoses in current practice (Lewis & Appleby, 1988; Tyrer et al., 2010). The worldwide pooled prevalence rate of any given PD has been estimated to be 7.8%, 95% CI [6.1-9.5%] in a recent meta-analysis (Winsper et al., 2019). The diagnosis is often associated with high levels of social and occupational impairment of functioning (Skodol et al., 2002; Newton-Howes et al., 2015). In England alone, the annual economic burden of this population has been estimated to be £7.9 billion (McCrone et al., 2008). Therefore, early identification and treatment is essential. However diagnostic systems for PD are notoriously unreliable; multiple diagnoses are common, diagnostic comorbidity is high and within-category heterogeneity means a single PD diagnosis can be met with a wide range of presentations (Widiger et al., 2009; Wright &

Zimmerman, 2015). Dimensional approaches to personality classification have been proposed to address these difficulties. These alternative models encourage clinicians to assess underlying aspects of functioning across personality traits, largely based on the empirically-derived Five Factor Model of Personality (Costa & Widiger, 1994; Bender et al., 2011). ICD-11 has replaced the categorical model of PD and its associated diagnoses with one PD diagnosis assessed to three levels of severity (mild, moderate, and severe) across the five personality trait domains (WHO, 2018). A similar model was proposed for the DSM-5, which will be introduced in more detail in the review element of this thesis. However, this Alternative Model of Personality Disorder (AMPD) within the DSM-5 (APA, 2013) received criticism around reliability and clinical utility, with many perceiving the model to be too complex to be used in clinical practice (Clarkin & Huprich, 2011; Tyrer, 2012; Pincus, 2011; Pilkonis et al., 2011). Thus, the categorical model remains the primary model of diagnosis in the current manual. Further research into the proposed dimensional model is essential to understand whether it has improved the reliability, validity, and utility of the diagnosis. Particularly given the power and stigma this diagnosis holds within systems (Lam et al., 2016).

This is particularly important to consider given that even with the effective assessment and diagnosis of offenders, judges remain solely responsible for sentencing decision making. Therefore, a judge can still provide a custodial sentence to OMHDs even if both medical doctors have recommended a Hospital Treatment Order (Sentencing Guidelines, 2020). Very little research has been carried out concerning the processes by which decisions are made to detain people under the MHA, particularly using the legislation in Part III of the MHA, which relates to “Patients concerned in Criminal Proceedings or Under Sentence”. Sentencing Guidelines (2020) exist to safeguard offenders and the public from excessively lenient or punitive sentences, however these are inherently vague and consequently

vulnerable to the influence of bias and heuristics (Tversky & Kahneman, 1974). Further guidance was provided regarding the sentencing of OMHDs in a recent ruling by the Court of Appeal during the judgement of *R v Vowles (2015).* It stated that in all cases where a s.37 non-punitive hospital order could be considered appropriate, a s.45A hybrid order must always be considered first. The Court of Appeal also encouraged sentencers to consider four criteria before a s.45A could be applied; (1) the extent to which the offender's mental health requires treatment, (2) the extent to which offending is attributable to the mental health disorder, (3) the extent to which offending requires punishment and (4) the protection of the public when deciding release and regime of release. This fundamentally changes the priority for OMHDs from one of treatment to one of punishment. This has significant consequences for OMHDs considering the aforementioned barriers to accessing mental health support in prison environments, therefore it is essential that the application of these sentences are consistent, transparent and not impacted by biases such as stigmatic attitudes towards mental illness or psychiatric diagnoses.

This thesis portfolio explores the sentencing of OMHDs by considering the two main stakeholders within this sentencing process, clinicians and the Judicial Office. Firstly, it focuses on the diagnostic process by which a diagnosable mental health condition is determined that would indicate application of the MHA (1983). The initial review explores the interrater reliability of the dimensional AMPD included in Section III of DSM-5 (APA, 2013). Secondly, it addresses the application of this clinical observation, assessment and diagnosis during the sentencing of OMHDs by the Judicial Office. The empirical paper explores whether sentencing outcome can be predicted by psychiatric diagnosis, mental illness stigma (as measured by the Perceived Devaluation and Discrimination Scale; Link, 1987) or the guidance provided in *R v Vowles*. The empirical project was a joint research project with another trainee, who has focused on a different set of research questions and the

impact of a separate topic, locus of control, on sentencing outcome (please see Appendix A for further information). Finally, an overall discussion and critical evaluation will be presented with strengths, limitations and reflections on completing the thesis presented.

Chapter 2: Meta-analysis

Interrater reliability of the Alternative Model for Personality Disorder (DSM 5 - Section III): A Meta-analysis

Samantha E. Young[a] & Peter Beazley [a]

Written for publication to Journal of Personality Assessment (Author guidelines for manuscript preparation - Appendix B)
Word Count: 9997

(No recommended word count for journal)

[a]Department of Clinical Psychology and Psychological Therapies, University of East Anglia,

Norwich Research Park, Norwich, Norfolk, NR4 7TJ, United Kingdom

*Corresponding author: Samantha.young@uea.ac.uk

This review did not receive a grant or funding and no conflicts of interest were reported.

Abstract

The Alternative Model of Personality Disorder (AMPD) is currently included in Section III of the Diagnostic and Statistical Manual for Mental Disorders – Fifth Edition. It was developed to address criticisms of the categorical system of Personality Disorder (PD) diagnosis, however received criticism around clinical utility and reliability due to the complexity of its approach. This review sought to summarise the literature concerning interrater reliability (IRR) of the AMPD. Despite some methodological limitations, meta-analysis provided tentative support for Criterion A of the AMPD. Pooled ICCs for the Level of Personality Functioning Scale (LPFS) and its domains fell above acceptability levels for DSM inclusion and previous estimates of IRR for the categorical system. Sub-group analysis of the LPFS suggested IRR scores could be improved by using a specific AMPD structured clinical interview. Variation in IRR was found across LPFS domains and may indicate differences in comprehension and application of these constructs. However, heterogeneity was high across all analyses and results should be interpreted with caution. Less research has been conducted around Criterion B and overall PD diagnosis; however initial studies show mixed results. Further research into these areas is recommended alongside more ecologically valid methods within clinical practice using structured clinical interviews.

Keywords: Personality Disorder, interrater reliability, Alternative Model of Personality Disorder

Introduction

The Diagnostic and Statistical Manual for Mental Disorders (DSM) was developed in 1952 to enable clinicians to share a common language for diagnosing psychiatric disorders (American Psychiatric Association, 1952). This classification system is distinct from the International Classification of Diseases 11th Revision (ICD-11; World Health Organisation, 2018), which has a broader scope than the DSM, including physical health as well as mental health. The ICD is the most commonly used diagnostic system clinically in the United Kingdom and outside of the United States of America. However, the DSM is arguably the preferred classification system for researchers globally, largely due to clear operational diagnostic criteria (Mezzich, 2002; Tyrer, 2014). Since their development, both systems have relied largely on a categorical approach to diagnosis. However, a dimensional approach to diagnosis has been increasingly preferred, particularly within the domain of personality (Widiger et al., 2009; Kotov, 2017). Dimensional approaches to personality classification – detached from psychopathology – have been widely accepted (e.g., Five Factor Model of Personality; FFM [Costa & Widiger, 1994] or the Eysenck Personality Questionnaire; EPQ [Eysenck & Eysenck,1975]). This has led many to question the appropriateness of the DSM-IV (APA, 1994) and the ICD-10 (WHO, 1992) categorical approach to Personality Disorder (PD) diagnosis.

PD is a complex, common and disabling diagnosis often associated with high levels of social and occupational impairment of functioning (Skodol et al., 2002; Newton-Howes et al., 2015). In England, significant barriers to accessing treatment for PD were highlighted by the "*Personality Disorder: No Longer a Diagnosis of Exclusion*" report (National Institute for Mental Health in England, 2003). The document highlighted a need for specialist PD treatment services and evidenced how generic mental health services were not meeting the needs of the PD population. Since its publication, an overhaul of PD service provision has

taken place. It has been estimated that there are now five times more dedicated NHS community and inpatient PD services across England (Dale et al., 2017) in addition to the development of the Offender Personality Disorder Pathway Programme (OPDPP) in the Criminal Justice System (National Offender Management Service, 2015). Early identification is important to improve treatment outcomes (Mulder, 2021) and within the NHS, a diagnosis of PD is required to access specialist treatment pathways and evidence-based treatment (Dale et al., 2017; NICE, 2009). However, rates of diagnosis are less than a quarter of estimated prevalence rates (Winsper et al., 2019; Mulder et al., 2011). Reluctance to diagnose PD may stem from limited available treatment options within psychopharmacology (Lewis & Appleby, 1988), poor provision of suitable psychotherapies (Dale et al., 2017) or high levels of associated stigma (Aviram et al., 2006; Markham & Trower, 2003). Furthermore, the reliability and clinical utility of the categorical model has been questioned, with "PD not otherwise specified" (PDNOS) one of the most commonly used diagnoses (Verheul & Widiger, 2004). Furthermore, multiple diagnoses are common, diagnostic comorbidity is high and within-category heterogeneity means a single PD diagnosis can be met with a wide range of presentations (Widiger et al., 2009; Wright & Zimmerman, 2015).

Criticisms of the categorical model have been attributed to weak empirical evidence for the approach (Clark, 2007; Haslam et al., 2012;2020). Arguably, many of the traits included in the categorical model can also be viewed as maladaptive variants of traits evident within the general population (Clarkin et al., 1993; Costa & Widiger, 1994). Moreover, perceptions of what constitutes "maladaptive" also vary depending on cultural contexts. PDs are diagnosed less in collectivistic cultures (Winsper et al., 2019). For example, Borderline PD has never been included in the Chinese Classification of Mental Disorders (CCMD-3; Chinese Psychiatry Association, 2000). Moreover, impairment of personality functioning has been suggested to be the largest predictor of both treatment outcome and distress (Crawford

et al., 2011; Bender et al., 2011). A structured assessment of this dysfunction would provide a more clinically useful focus for diagnosis and prevent unhelpful behavioural descriptions of PD (Mulder, 2021). However, this is not possible within the current categorical approach.

To address the failings of the categorical model, the Alternative Model of Personality Disorder (AMPD; DSM-5, Section III [APA, 2013]) was developed. The AMPD claims to have a clearer conceptual basis for PD pathology, more efficient and effective approach to assessment, and empirically based criteria (Krueger & Hobbs, 2020). PD is diagnosed by meeting seven criteria (A-G). Criterion A focuses on personality impairment and is operationalised by the Level of Personality Functioning Scale (LPFS; Bender et al., 2011). The dimensional constructs are categorised into "Self" (including *identity* and *self-direction*) and "Interpersonal" (including *intimacy* and *empathy*) domains and are rated on a five-point scale of impairment. Moderate impairment or greater meets clinical threshold. Criterion B identifies pathological personality traits. The five trait domains loosely map onto the Five Factor Model of Personality (Costa & Widiger, 1994; Krueger et al., 2012). Twenty-five trait facets are identified under the following categories: Negative Affectivity, Detachment, Antagonism, Disinhibition and Psychoticism. The facets are rated on a four-level scale of descriptivism.

The additional criteria (B-G) address the pervasiveness and stability of the difficulties, differential mental and medical diagnoses, and alternative explanations for the difficulties (including developmental stage, substance use and sociocultural context). Traits can be clustered together to provide final categorical diagnostic categories; however, these categories have been reduced from ten to six to reduce diagnostic overlap. Finally, "PDNOS" is replaced with "PD-trait specified" (PD-TS) which includes a summary of identified traits to encourage more clinically useful descriptions of PD (Clark et al., 2015).

Despite much support for the dimensional model, the categorical system used in DSM IV-TR (APA, 2000) was replicated as the primary model for PD diagnosis in DSM-5 (APA, 2013). Concerns were raised around the AMPD's clinical utility, and the model was deemed too complex and theory laden for clinical practice (Clarkin & Huprich, 2011; Tyrer, 2012; Pincus, 2011; Pilkonis et al., 2011). However, the AMPD was included in Section III of the manual for "Emerging Measures and Models". It was hoped this would encourage future research into the clinical utility and reliability of the model.

In contrast, the dimensional model was included in the most recent revision of the ICD (ICD-11; WHO, 2018). PD diagnosis in the ICD-11 replaces all previous PD diagnostic categories with one overarching PD (excluding the "Borderline Pattern" specifier). This PD is assessed to three levels of severity (mild, moderate and severe), across five personality trait domains; Negative Affectivity, Detachment, Disinhibition, Dissociality, and Anankastia (WHO, 2018). Limited research has been published examining the reliability, validity or clinical utility of this model, however there is significant alignment between this model and the DSM-5 AMPD (Mulder, 2021; Hemmati et al., 2021). Thus, it is hoped that the growing body of research surrounding the AMPD will be generalisable to the ICD-11 (Mulder, 2021; Mulder & Tyrer, 2019).

Interrater Reliability

Interrater Reliability (IRR) is a fundamental assessment tool for diagnostic accuracy (Kraemer et al., 2012). It enables researchers to quantify agreement between two or more raters and thus, measure the consistency with which diagnostic criteria is applied. IRR is of particular interest in PD diagnosis where within category heterogeneity and multiple diagnoses are common, and no "gold standard" of assessment is recommended (Widiger et al., 2009; Krueger, 2020). It is even more relevant to consider for the AMPD, where concerns have been raised around clinician's understanding of the model (Tyrer, 2012; Pincus, 2011;

Pilkonis et al., 2011). Arguably, diagnostic systems with higher IRR indicate a shared conceptual understanding of the diagnosis and its application and are less prone to errors caused by things such as human judgment, inexperience, or heuristics (Aboraya et al., 2006; Tversky & Kahneman, 1974). IRR of the DSM has unsurprisingly decreased in recent years, with comparisons of the DSM-5 to the DSM-III deemed unfair due to the addition of new diagnoses and specifically, sub-categories of these diagnoses which are classified in finer detail (Frances, 2012; Aboraya et al., 2006). A previous review conducted by Samuel (2015) found overall moderate levels of IRR for the categorical system of PD. They reported a median kappa value for the diagnosis of any PD of .52 and an individual PD of .40. Their findings suggested using a structured clinical assessment (the Shedler-Westen Assessment Procedure; Westen & Shedler, 1999) produced increased levels of IRR (k=.61), however variations in study design had little impact on the overall score (Samuel, 2015).

Clinical Assessment

Structured clinical interviews (SCIs) are considered the "gold standard" of clinical assessment and have been reported to improve reliability and reduce bias within diagnostic decision making (Samuel, 2015; Aboraya et al., 2006; Wood et al., 2002). It has been suggested that using different, or indeed no, instrument to guide PD assessment, will result in substantial differences to diagnostic outcomes (Clarkin & Huprich, 2011; Samuel, 2013; 2015). Multiple structured assessments have been produced for the categorical model (e.g., SCID-II; First et al., 2011)], however remain unpopular in clinical practice due to constraints on time and resource (Milton, 2000; Perry, 1992). Self-report measures have been developed for use with the AMPD (e.g., PID-5; Krueger et al., 2012), however rely on self-awareness and accurate reporting; both of which can be impacted by the very nature of these psychopathologies (Carlson & Ottomanns, 2015). For example, a lack of insight is inherent to the narcissistic presentation, and a tendency to deceive is common in anti-social types.

Informant measures may provide additional information, however due to fundamental difficulties with interpersonal relationships, these may also not be accurate or indeed possible to obtain (Balsis et al., 2015). Unstructured clinical interviews guided by DSM criteria may provide additional observed non-verbal information; however, rely on clinical judgment, which is often overestimated (Monahan, 1981; Kitamura & Kitamura, 2000). In these conditions, clinicians are more likely to rely on cognitive heuristics and biases for decision making (Tversky & Kahneman, 1974). For example, evidence suggests a clinician preference for diagnosing Borderline PD in females (Sansone, & Sansone, 2011). Furthermore, it is common for clinicians to neglect to assess additional diagnoses or traits once a single diagnosis has been met (Herkov & Blashfield, 1995). Therefore, it is important validated and reliable SCIs are developed alongside the AMPD to increase reliability in clinical decision making.

Aims

This review builds on the work of Samuel's (2015) review concerning the IRR of PD diagnoses. To the author's knowledge, no systematic review or meta-analysis exists summarising literature on IRR of the AMPD and its associated instruments. Therefore, this review aims to effectively summarise existing studies which report a statistical measure of IRR for the AMPD. It is hoped this will provide a significant contribution to the literature around the consistent application of the AMPD and its potential inclusion as the dominant model of PD diagnosis in the DSM.

Where possible, this data will be pooled through meta-analysis and reported in relation to DSM-5 acceptability thresholds for IRR (>.40; Kraemer et al., 2012) and reporting guideline thresholds for Intraclass Correlation Coefficients (ICCs); a dimensional measure of IRR (Koo & Li, 2016). High rates of heterogeneity are expected due to variation in measures

and methodological differences. To account for this, sensitivity analyses will be conducted alongside sub-group analysis of assessment method.

Method

Protocol and Registration

The review protocol was published on the PROSPERO international prospective register of systematic reviews (PROSPERO 2021 CRD42021254416; accessed via www.crd.york.ac.uk/PROSPERO).

Search Strategy

A systematic search of published literature was conducted using the electronic databases PsycINFO, MEDLINE, CINAHL, and Scopus. Search terms "personality disorder*" AND ("alternative model of personality disorder*" or "ampd") were used. Broad search terms were used due to the limited research available. Reference lists and citations of key review papers and accepted papers were hand searched. Searches were conducted on 4th June 2021.

Eligibility

Inclusion Criteria

This review sought to identify empirical research which aligned with the aims of this systematic review and previously set inclusion criteria. All criteria had to be met to be included. The inclusion criteria were:

(a) Empirical research studies which applied one or more sections of the AMPD and were rated by more than one person, including an appropriate statistical measure of inter-rater reliability (IRR). For the dimensional components (A and B), the appropriate statistical measure of IRR was judged to be the single-

rater ICC or an equivalent statistic. For overall PD diagnosis, a categorical statistical measure of IRR will be required e.g., Cohen's kappa coefficient.

(b) Studies taking place in clinical or non-clinical settings (no exclusion criteria will be applied for setting).

(c) Articles must have been published in peer-reviewed journals only

(d) Articles must have been written or translated into the English Language

Exclusion Criteria

In addition to the inclusion criteria, studies were excluded under additional specific circumstances:

(a) based on design (qualitative studies, theoretical, conceptual, or critical commentary, reviews or meta-analyses)

(b) population (under 18s not included)

(c) outcome (outcomes not conceptually relevant to IRR or insufficient reporting of statistical data).

Identification and Selection of Studies

Articles were identified, screened, and assessed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Moher et al., 2009) (Figure 1). Papers were extracted from electronic databases using Mendeley reference management software and duplicate articles were removed (Lorenzetti and Ghali,2013). Titles and abstracts of the remaining papers were screened by the primary author according to eligibility for the review. Ten percent of these papers were independently reviewed by a second rater, a Trainee Clinical Psychologist (Boland et al., 2017). The kappa value at title and abstract screening was $\kappa = 0.83$ (95% CI, [.65 to 1.01], $p < .001$) indicating a strong level of agreement between raters (McHugh, 2012; see Appendix C). Any disagreements were

discussed with the researcher's academic supervisor and a final decision on the papers was agreed upon.

Full articles of remaining papers were acquired and assessed for eligibility. The first author and a Trainee Clinical Psychologist read 14 randomly selected papers (ten percent of identified articles) and independently completed an electronic screening checklist to ensure accurate selection of the eligible papers (Boland et al., 2017). The kappa value at full article screening was κ =0.85 (95% CI, [.57 to 1.13], $p < .001$) indicating a strong level of agreement between raters (McHugh, 2012; see Appendix C). Discrepancies were discussed and resolved, and a third reviewer was available for consultation, however not required. The final studies were checked against eligibility criteria to reduce bias. Both raters agreed that all the selected studies met the eligibility criteria.

Data Extraction and Coding

A data extraction database was used to record the following information for accepted studies; (a) article details (for example, author, publication year, title, journal), (b) study design and setting (c) sample description (including sample size and demographics), (d) rater description (including number, profession, and training) (e) AMPD criterion assessed, and instruments used, and (f) IRR statistics type and outcome.

For Criterion A and B, IRR statistics were extracted for both domains and total scores. Studies which reported IRR statistics for measure facets only were excluded. Only single-rater ICCs were extracted from studies as this was the most reported and clinically relevant statistically. A decision was later taken to also include weighted kappa in the analysis due to statistical similarity with single-rater ICC (Schuster, 2004). Where only "mean of *k* raters" ICCs were reported, studies were excluded from final analysis. Data duplication was monitored throughout, and a further two studies were excluded as a result.

Quality Assessment and Risk of Bias

The quality and risk of bias of included studies was assessed using the Quality Appraisal of Reliability Studies (QAREL; Lucas et al., 2010). The QAREL is an 11-item checklist used to assess reliability studies. It explores seven principles representing the appropriateness of subjects, qualification of examiners, examiner blinding, order effects of examination, suitability of the time interval between repeated measurements, appropriate test application and interpretation, and statistical analysis. The scoring criteria for each item in each of the domains are *Yes*, *No*, *Unclear*, or *N/A*. Each item on the checklist is weighted equally. The total number of "yes" values are calculated as a percentage. A second reviewer was consulted where a judgement was unclear. High quality studies were defined as ≥60% "yes" on the QAREL (Cuchna et al., 2016).

Data synthesis

Meta-analyses were conducted using the Metafor package version 3.0-2 (Viechtbauer, 2010) in R Studio version 1.4.1717 (Wallace at al., 2012). Random effects models were used due to the presumed variance in effect sizes extracted from each study. Cuijpers (2016) recommends this as the most suitable model for mental health research. Moreover, this approach allows for differences in true effect sizes between studies, as it provides broader and more conservative 95% confidence intervals than fixed effects models. A large amount of variation in effect sizes was expected, given the varied methodologies in the included studies.

For each meta-analysis, ICC and sample size values were extracted from each study, transformed to Fisher's *z* scale, and combined using a random–effects models (Cuchna et al., 2016; Borenstein et al., 2009). Fisher's *z* transformations are important to account for the non-normal distribution in these types of statistics (Cuijpers, 2016). Heterogeneity within meta-analyses was assessed following the transformation to Fisher's *z* by inspecting forest plots as well as using Cochran's Q test (Cochran, 1954) and the I^2 statistic (Higgins & Thompson,

2002). The Q test is important to determine whether heterogeneity was significant and the I^2 statistic provides a percentage of variation across studies due to heterogeneity versus chance. To aid in the interpretation of the results, Fisher's *z* values were then converted back to ICC values after completing meta-analyses (Cuchna et al., 2016). A descriptive summary of the selected research studies will be provided in table format. Where meta-analysis is not indicated, IRR statistics from studies will be summarised in a table. The results from a quality assessment tool will also be reported in table format to provide a more detailed assessment of the studies reviewed.

Sub-group Analysis

Sub-group analysis for Criterion A - Total LPFS Score was conducted using random effects model for instrument type (instrument designed for use with the AMPD vs no AMPD instrument).

Sensitivity Analysis

Sensitivity analysis was conducted to examine whether results were skewed by studies with a high risk of bias. High risk studies were removed if this was judged to be the same. Publication bias was explored visually using a funnel plot and the "trim and fill" method was applied to estimate IRR after bias had been accounted for (Duval &Tweedie, 2000; Viechtbauer, 2010).

Results

The selection, exclusion and inclusion process of studies is outlined in the PRISMA diagram (Figure 1). The search yielded 719 studies of which 337 were excluded as duplicates resulting in 386 articles which were screened based on title and abstract. During this stage, 134 articles were subject to full eligibility screening, resulting in a total of 15 eligible studies. The most common reason for exclusion at the final stage was due to an AMPD criterion not being rated by two or more raters.

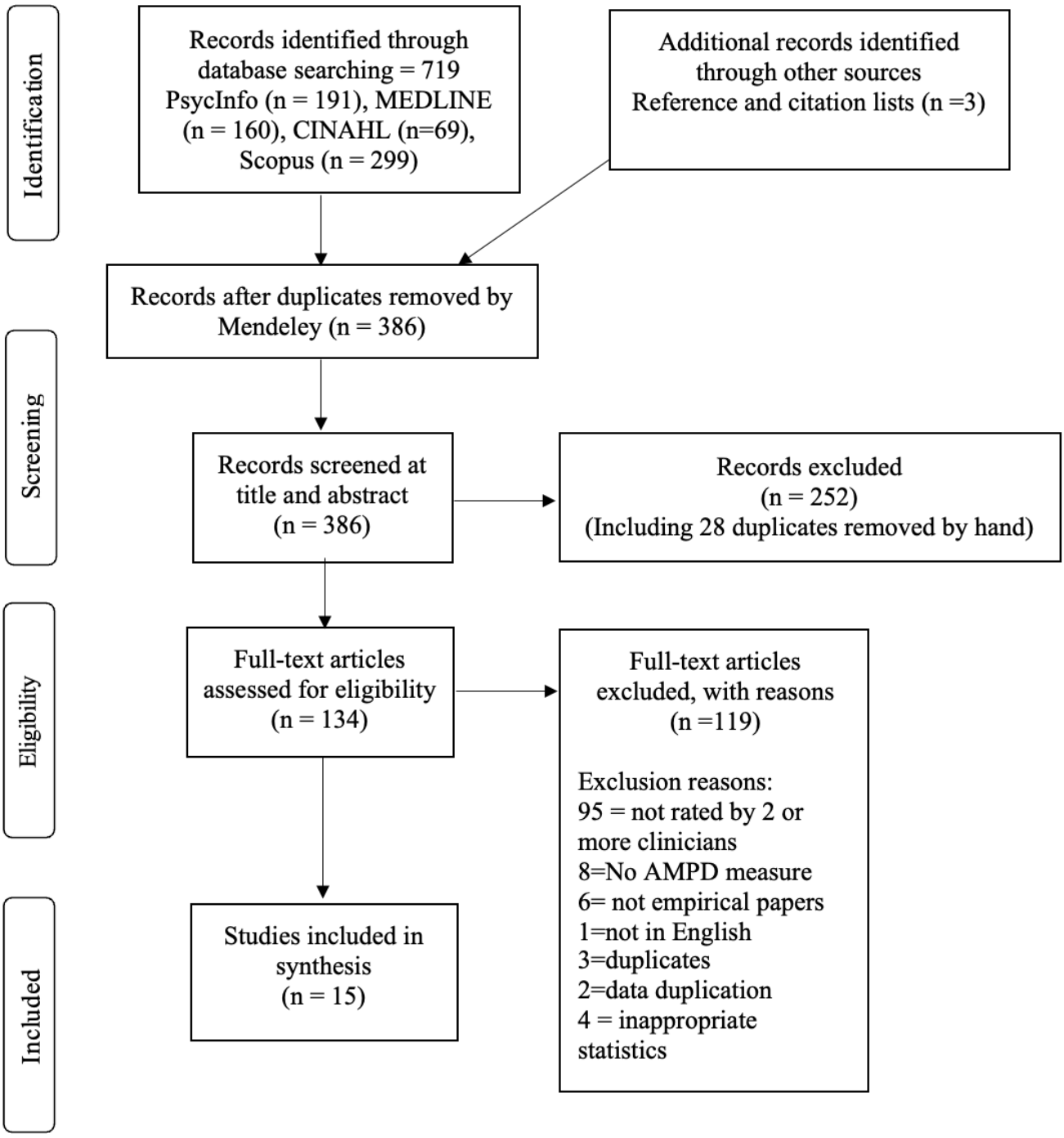

Figure 1 - PRISMA diagram outlining the searching and exclusion processes

For Criterion A, 14 studies reported IRR statistics for LPFS total score and 11 reported statistics for the individual domains. Three studies reported single rater ICCs for Criterion B and three studies reported reliability statistics for overall diagnosis. Total sample size across LPFS Total Score was 902 (range from six to 162), this reduced to 775 for the individual domains (range ten to 162). For Criterion B the sample size was 115 (range from 12 to 88). Sample sizes for Criterion B and overall diagnosis were not pooled but ranged from 12 to 88 from Criterion B and 12 to 120 for overall diagnosis.

Study Characteristics

Characteristics (location, instrument, methods, sample size, sample demographic [age and gender], population type and rater information [type and number]) for the 15 studies included in the review can be found in Table 1. Rater experience varied across studies from clinically inexperienced lay raters (Zimmerman et al., 2014) to qualified mental health professionals (Morey, 2019). Ethnicity of sample and raters was largely unreported across studies, therefore not included in Table 1.

Study Design

Studies used a variety of methods to assess IRR. Three studies used written vignettes or accounts of life stories. The remaining studies utilised a mix of live interviews and video recordings. Instruments used for assessments varied considerably. Three studies used SCIs designed for assessing the categorical system of PD diagnosis (Structured Interview of Personality Organization; STIPO [Clarkin et al., 2004]) and the SCID–II (First et al., 1995) and eight used a specific AMPD assessment tool (SCID–5 AMPD; Bender et al., 2018), The Clinical Assessment of the Level of Personality Functioning Scale (CALF; Thylstrup et al., 2016) and the Interview for Personality Functioning DSM–5 (STiP-5.1; Hutsebaut et al., 2014). Other studies did not use an instrument, but instead relied on clinical interview (or reviewing materials) using the information provided in the AMPD to guide assessments.

Risk of Bias Assessment

Outcome of the QAREL (Lucas et al., 2010) can be seen in Table 1. A detailed assessment can be found in Appendix D. All studies bar one reached the 60% cut off for low risk of bias and therefore high quality. Roche et al. (2018) scored 54.5% and was therefore rated as high risk of bias and low quality. This study was included in the initial analysis, however removed in sensitivity analysis.

Table 1.
Overview of final studies included in review

Study	Location	Instrument	Methods	Sample size	% Female	Sample age (years)		Population	Raters	No. of ratings pp	Risk of bias QAREL % of yes	Risk of Bias Rating
						Range	*M*±*SD*					
Buer Christensen et al. (2018)	Norway	SCID-5-AMPD	Live interviews and video recordings	17	65	19-59	31.6	Clinical and non-clinical	7 experienced raters	5	72.7	Low
Cruitt et al. (2019)	United States	LSI and LPFS	Video recordings of LSIs	162	56.2	55-64	NR	Non-clinical	9 undergraduate students	3	72.7	Low
Dereboy et al. (2018)	Turkey	SCID-II and LPFS	Live interviews	120	66.67	16-63	35.7± 12.5	Clinical (inpatient and outpatient)	Clinicians*, 3 CP students, 3 academics, 1 psychiatrist and 2 CPs	2	63.6	Low
Garcia et al. (2018)	United States	LPFS and PDLT–C	Written vignettes	15	NR	NR	NR	Clinical	13 CP doctorate students	13	63.6	Low
Hutsebaut et al. (2017)	Netherlands	STiP-5.1	Live interviews and video recordings	40**	66.3	16-61	33.6± 12	Clinical	12 CPs and 3 academics	2	90.9	Low
				18	4.2	18-60	39± 14.5	Community	(same as above)			
Hutsebaut et al. (2021)	Netherlands	STiP-5.1	Interview and observer	30	10.3	21-65	38.43± 11.70	Clinical (forensic inpatient)	CPs	2	81.8	Low

Kampe et al. (2018)	Germany	SCID-5-AMPD and STIPO	Interviews and video recordings	30	60	16-61	32.5± 9.77	Clinical	Author and a CP Masters student	2	63.6	Low
Morey (2019)	United States	DSM TEXT	Written vignettes	12	NR	NR	NR	Clinical (not all PD)	123 mental health professionals	40	72.7	Low
Preti et al. (2018)	Italy	STIPO	Live interviews and audio recordings	10	100	NR	36.6± 11.26	Clinical (inpatients)	10 clinically inexperienced undergraduate students	10	72.7	Low
Roche et al. (2018)	United States	LFPS	Written abbreviated LSIs.	70***	80	NR	20.52± 0.98	Non-clinical (students)	15 research assistants (3 teams of 5)	5	54.5	High
				85[a]	81	NR	19.92± 1.52					
				85 [a]	81	NR	19.92± 1.52					
Somma et al. (2019)	Italy	SCID-5-AMPD	Interview and observer	84	53.6	NR	36.42± 12.94	Clinical	10 Trainee CPs	2	72.7	Low
Somma et al. (2020)	Italy	SCID-5-AMPD	Interview and observer	88	54.5	NR	36.47± 14.04	Clinical	CPs with 1–3 years of experience	2	90.9	Low
Thylstrup et al. (2016)	Denmark	CALF	Interview and video recording	30[b]	47.22	18-56	36	Clinical	4 psychologists and 2 MDs (2 interviews conducted by a psychology student)	3	63.6	Low
				7	100	24-45	34	Non-clinical				

Zettl et al. (2020)	Germany	STiP-5.1	Interview and recordings	50[c]	57.27	NR	31.44± 11.74	Clinical and non-clinical	Trained researchers	2	63.6	Low
Zimmerman et al. (2014)	Germany	OPD and LFPS	Video recordings	10	100	NR	30.8± 9.6	Clinical (inpatients)	22 untrained and clinically inexperienced students	22	72.7	Low

Note. SCID-5-AMPD= SCI for the DSM-5 Alternative Model for Personality Disorders; LSI = Life Story Interviews; LPFS = Level of Personality Functioning Scale; SCID-II = SCI for the Diagnostic and Statistical Manual of Mental Disorders Axis II; CP=Clinical Psychologist; PDLT–C = Clinician Rating Personality Disorder Level and Traits; STiP-5.1= Semi-Structured Interview for Personality Functioning DSM–5; STIPO= Structured Interview of Personality Organisation ; CALF= Clinical Assessment of the Level of Personality Functioning Scale; OPD= Operationalized Psychodynamic Diagnosis system; *refers to intake clinicians as raters in addition to "second rater". Unclear how many clinicians took part, therefore this is coded as 2 raters to reduce error. **only 40 participants were included in IRR analysis, however demographics are reported for the total sample of 80. ***IRR sample of 70, however only 50 participants completed the initial demographic survey. [a]IRR sample of 85, demographics reported for total 110 sample of study. [b]IRR sample of 30, demographics reported for total sample of 37. [c]IRR sample of 50, demographics reported for total sample of 110.

Table 2

Interrater reliability statistics for Criterion A- Total LPFS scores and individual domains organised alphabetically by measure, then author.

				IRR (95% confidence intervals)				
Study	**N**	**Measure**	**IRR Stat**	**Total LPFS**	**Identity**	**Self-Direction**	**Empathy**	**Intimacy**
Hutsebaut et al. (2017)	58	STiP-5.1	ICC (1,1)	.71	.76	.64	.79	.80
Hutsebaut et al. (2021)	30	STiP-5.1	ICC (1,1)	.81	.54	.77	.69	.90
Zettl et al. (2020)	110	STiP-5.1	ICC (2,1)	.92	.93	.92	.89	.93
Kampe et al. (2018)	6	STiPO	ICC (2,1)	.78	-	-	-	-
Preti et al. (2018)	10	STiPO	ICC (2,1)	.42 (.21,.72)	.39 (.19,.71)	.35 (.16, .67)	.28 (.11,.60)	.42 (.21,.73)
Buer Christensen et al. (2018)	17	SCID-5-AMPD	ICC (2,1)	.96 (.92,.9)	.94 (.88, .98)	.94 (.87, .98)	.9 (.80, .96)	.89 (.80, .96)
Kampe et al. (2018)	30	SCID-5-AMPD	ICC (2,1)	.93 (.87,.90)	.89	.79	.92	.95
Somma et al. (2019)	84	SCID-5-AMPD	Kw	.87	-	-	-	-
Somma et al. (2020)	88	SCID-5-AMPD	ICC (1,1)	.87	.83	.87	.77	.88
Thyltsrup et al. (2016)	30	CALF	ICC (3,1)	.69 (.47-.83)	.59 (.33-.76)	.72 (.51-.84)	.42 (.11-.65)	.65 (.42-.80)
Cruitt et al. (2019)	162	LSI and DSM	ICC (1,1)	.56	.57	.50	.47	.37
Garcia et al (2018)	15	DSM	ICC (2,1)	.81				
Morey (2019)	12	DSM	ICC (1,1)	.50	-	-	-	-
Roche et al. (2018a)	70	DSM	ICC (2,1)	.58 (.48,.69)	.49 (.38,.61)	.44 (.33,.56)	.26 (.16,.38)	.29 (18,.41)
Roche et al. (2018b)	85	DSM	ICC (2,1)	.42 (.32,.53)	.41 (.31,.52)	.29 (.19,.40)	.23 (.13,.3)	.31 (.21,.42)
Roche et al. (2018c)	85	DSM	ICC (2,1)	.36 (.26,.47)	.41 (.31,.52)	.21 (.12,.32)	.14 (.06,.24)	.23 (.13,.33)
Zimmerman et al. (2014)	10	OPD and DSM	ICC (2,1)	.51 (.31,.70)	.41 (.23,.71)	.46 (.27, 75)	.25 (.12,.55)	.63 (.43, 85)

Table 3.
Meta-analysis outcomes for Interrater Reliability of Criterion A LPFS and its domains, including subgroup analysis.

Variable		*k*	ICC	95% Confidence LL	UL	SE	*p*	T^2	*z*	*Q*	*df*	*p*	I^2
TOTAL LPFS		17	0.75	0.63	0.84	0.12	<.001	0.44	8.27	171.18	16	<.001	90.10
Instrument used	AMPD Instrument	8	0.87	0.79	0.92	0.13	<.001	0.32	10.30	35.61	7	<.001	83.90
	No AMPD instrument	9	0.51	0.42	0.60	0.06	<.001	0.01	8.87	9.98	7	0.27	25.30
Domains	Identity	13	0.70	0.55	0.81	0.13	<.001	0.19	6.55	148.17	12	<.001	91.56
	Self-Direction	13	0.68	0.51	0.80	0.14	<.001	0.21	6.04	174.81	12	<.001	92.12
	Empathy	13	0.63	0.43	0.77	0.15	<.001	0.24	5.11	165.17	12	<.001	92.91
	Intimacy	13	0.73	0.54	0.85	0.16	<.001	0.31	5.68	247.33	12	<.001	94.46

Criterion A - LPFS

A total of 14 studies reported 17 interrater reliability scores using single-rater ICCs or equivalent for Total LPFS score (one study reported three IRR tests using independent pools of raters, and one study reported two separate analyses based on two separate instruments conducted in separate interviews, with separate samples). This resulted in a pooled ICC of .75 (95% CI .63 – .84), however this was significantly heterogeneous (Q (16) =171.18, p<.01, I^2 = 90.10%; see Figure 2). This is above the DSM-5 cut off for acceptable IRR for the DSM and would be categorised as good reliability under ICC reporting guidelines (Kraemer et al., 2012; Koo & Li, 2016). Table 3 provides additional information on overall agreement.

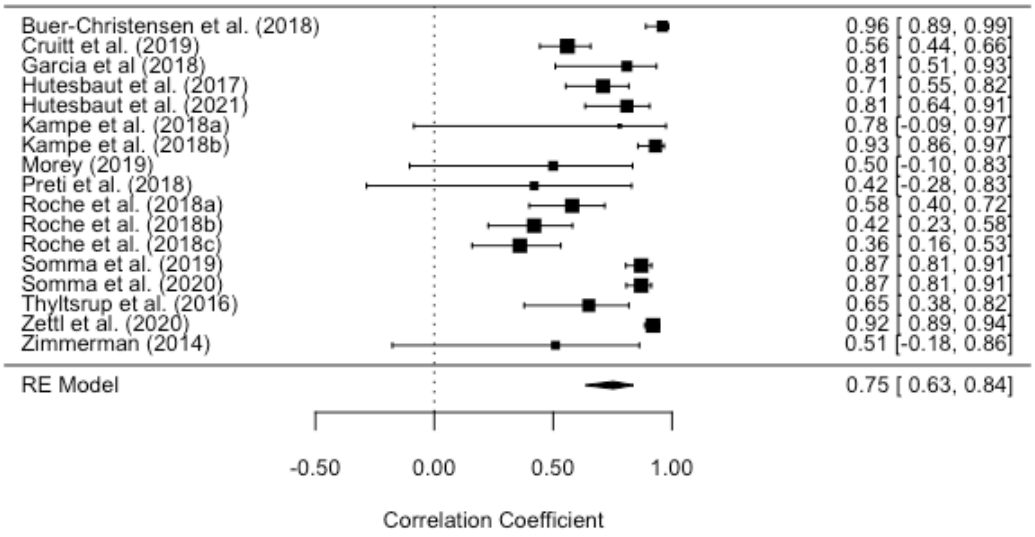

Figure 2. Forest plot of Total LPFS IRR meta-analysis.

Subgroup analysis of Total LPFS Score

Subgroup analysis of the assessment instrument can be found in Table 3. This separated studies into those which used an instrument designed for use with the AMPD (i.e., SCID-5-AMPD, STiP-5.1 and CALF) and those that either used an alternative structured assessment not recommended for the AMPD, or no assessment tool at all. Results showed the AMPD instruments yielded higher IRR correlations (*ICC*=.87, 95% CI [.79, .92]; Figure 3), however heterogeneity remained high for this group. Conversely, in the second group (*ICC*=.51, 95% CI [.42, .60]; Figure 4), heterogeneity was no longer significant (Q (8) =9.98, p=0.27, I^2 = 25.30%; see Table 2).

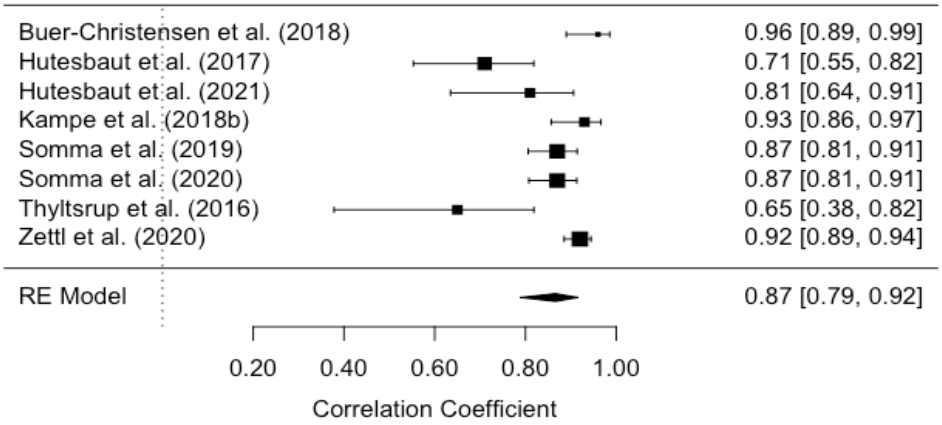

Figure 3. Forest plot of Total LPFS subgroup IRR meta-analysis including AMPD instruments only.

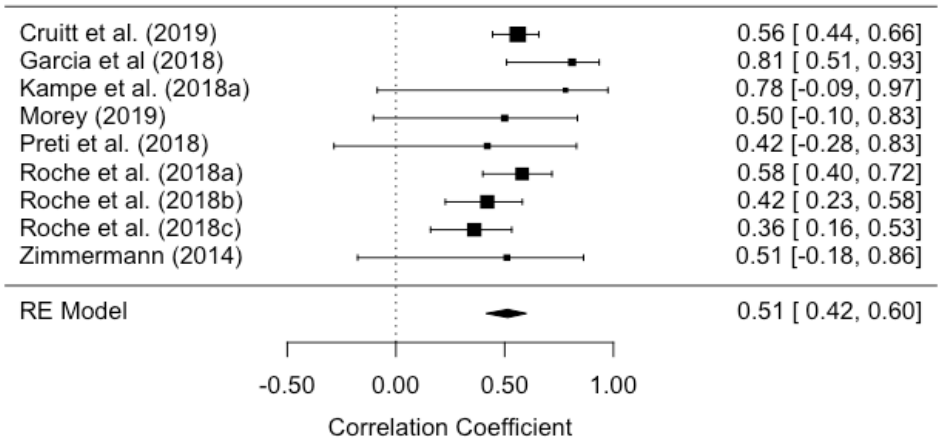

Figure 4. Forest plot of Total LPFS subgroup IRR meta-analysis including studies that did not use an AMPD instruments.

Sensitivity analysis of Total LPFS estimates

Sensitivity analysis was conducted by removing studies rated as having a high risk of bias (n=1) (Roche et al., 2018). This increased reliability ICC of .80 (95% CI .70 – .87) and decreased heterogeneity, however this was still significant (see adapted table in Appendix E). Leave-one-out analysis (Viechtbauer, 2010) was also performed which showed no single study accounted for the high levels of heterogeneity.

Publication bias of Total LPFS estimates

Publication bias was assessed using funnel plots of Total LPFS IRR (see Appendix F). Funnel plots appeared symmetrical and trim and fill analysis did not highlight any missing studies. Visual inspection of forest plots (Figure 2) highlighted small sample studies accounted for large confidence intervals (Kampe et al., 2018; Preti et al., 2018; Morey, 2019; Zimmerman et al., 2014).

LPFS Domains

Eleven studies reported single-rater ICCs for the individual LPFS domains (again, one study reported three IRR tests using independent pools of raters). The main findings for each individual domain meta-analysis can be found in Table 3. This table provides information on the number of studies (*k*), pooled sample size (*N*), estimate of overall ICC (*ICC*), 95% confidence intervals, significance test of weighted effect size estimate (*z*) and amount of heterogeneity (*Q*). Significant heterogeneity was found in all domains with pooled ICCs ranging from *empathy* (*ICC*= .63, 95% CI [.43 – .77]), to *intimacy* (*ICC*= .73, 95% CI [.54 – .85]). All pooled domains were above the DSM-5 cut off for acceptability and fell within the moderate category of reliability according to ICC reporting guidelines (Kraemer et al., 2012; Koo & Li, 2016).

Sensitivity analysis of LPFS domain estimates

Sensitivity analysis was conducted by removing domain studies rated as having high risk of bias (i.e., Roche et al., 2018). This increased reliability (ICCs ranged from .73, 95% CI [.57-.84] to .82, 95% CI [.68-.90]), however heterogeneity remained significant (see adapted table in Appendix E). Leave-one-out analysis (Viechtbauer, 2010) was performed and showed no one study accounted for the high levels of heterogeneity.

Publication bias of LPFS Domain estimates

Publication bias was assessed using funnel plots of all domains (see Appendix. D) . Funnel plots appeared symmetrical, however trim and fill analysis highlighted one missing study for the *Identity* domain. However, pooled ICC for this domain remained at .70, 95% CI [0.54, 0.81] even when this study was accounted for. Visual inspection of forest plots (Figures 3-6) showed small sample studies accounted for large confidence intervals (Preti et al., 2018; Zimmerman et al., 2014).

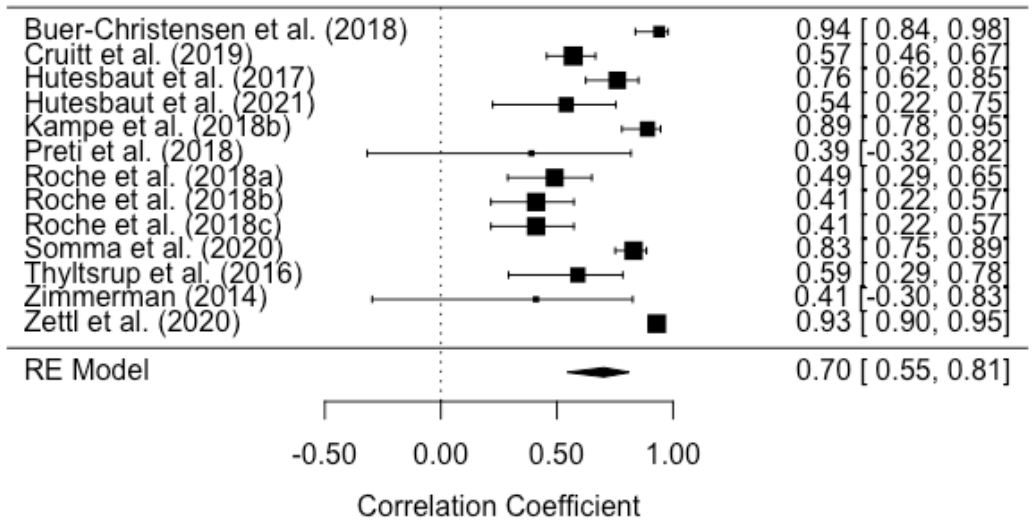

Figure 5. Forest plot for LPFS Domain – Identity IRR meta-analysis

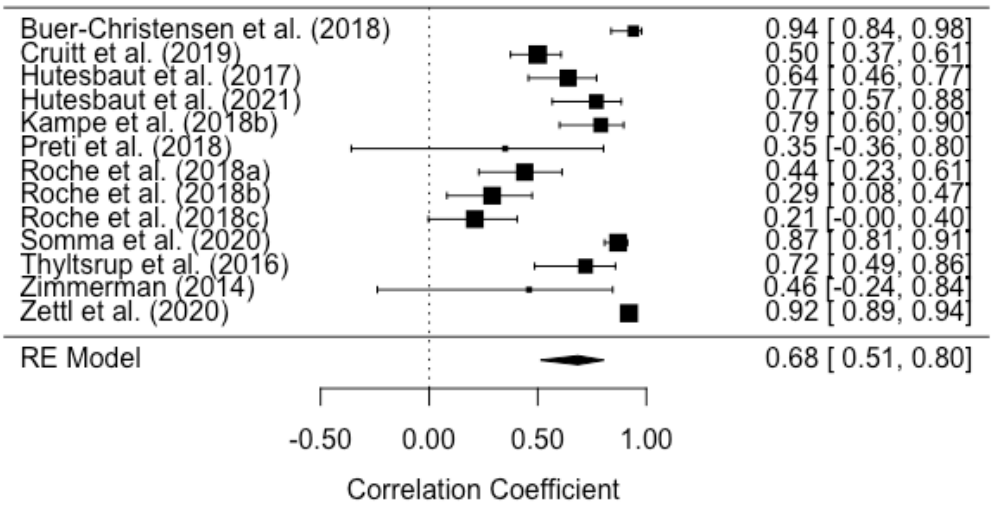

Figure 6. Forest plot for LPFS Domain – Self-Direction IRR meta-analysis

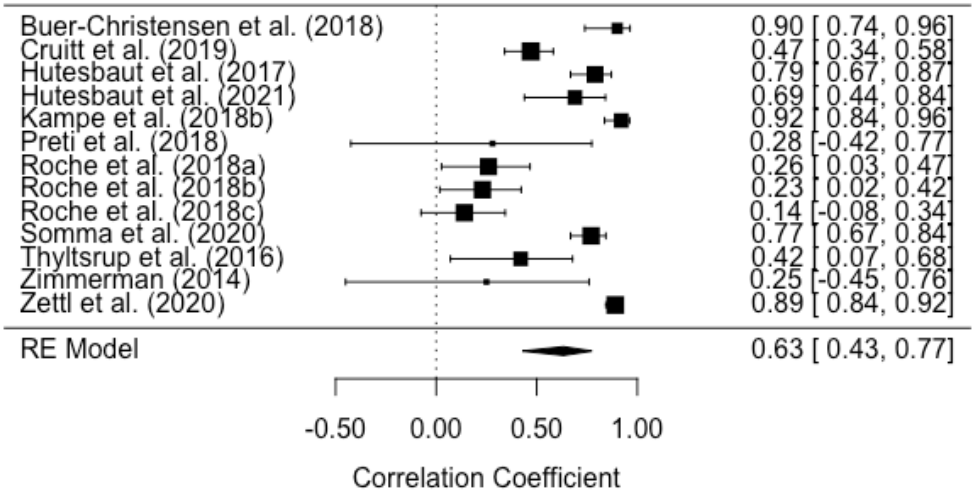

Figure 7. Forest plot for LPFS Domain – Empathy IRR meta-analysis

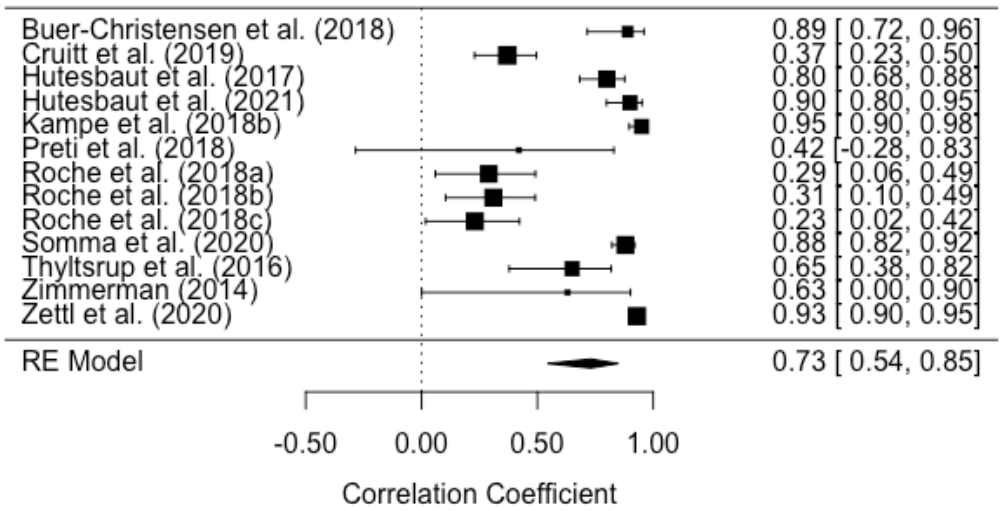

Figure 8. Forest plot for LPFS Domain – Intimacy IRR meta-analysis

Criterion B – Trait Domains

Three studies reported single-rater ICCs for Criterion B trait domains. Due to small study sample a meta-analysis was not considered appropriate. Heterogeneity is common within small sample sizes, however often not detected by statistical methods when few studies are included (Cuijpers, 2016). Variability within the studies is apparent with varying methods of assessment (interviews vs clinical vignettes) and measures used (SCID-AMPD vs PDLT-C vs DSM text). Summary IRR statistics can be seen in Table 4. Trait domains ranged from .92 (Disinhibition; Somma et al., 2020) to .43 (Detachment; Morey, 2019). No single domain fell below the recommended acceptability levels for the DSM-5 (Kraemer et al., 2012). However, several domains fell within the poor category of reliability (*ICC*<.50; Koo & Li, 2016). Somma's (2020) study using the SCID-5-AMPD (Bender et al., 2018) appears to have gained the highest IRR score with ICCs ranging from .79 to .92.

Table 4.
Interrater Reliability of Criteria B of the AMPD.

Study	N	Measure	IRR Stat	Negative Affectivity	Detachment	Antagonism	Disinhibition	Psychoticism
Garcia et al. (2018)	15	PDLT–C	ICC (2,1)	.50	.48	.45	.49	.59
Morey et al. (2019)	12	DSM text	ICC (1,1)	.58	.43	.51	.62	.67
Somma et al. (2020)	88	SCID-5-AMPD	ICC (1,1)	.80	.82	.79	.92	.85

Overall PD Diagnosis

Only three studies reported IRR for overall PD diagnosis. Statistical analysis varied, with two studies reporting kappa (Dereboy et al., 2018; Somma et al., 2019) and one ICC (Morey et al., 2019). Methods of assessment also varied across the three studies with two assessing only Borderline PD and one identifying PD vs no PD as the outcome. Studies were

thus considered too heterogenous to meta-analyse, particularly with a small sample (Cuijpers, 2016). Summary IRR summary statistics can be found in Table 5. There was considerable variance for IRR values across studies with kappa values ranging from moderate (0.68; Dereboy et al., 2018) for general PD to good (0.83; Somma et al., 2019). Both studies scored above the acceptability level for DSM-5 (Kraemer et al., 2012). However, dimensional reliability was poor (*ICC*=0.33; Morey, 2019) and below acceptable level for DSM-5 IRR thresholds (Koo & Li, 2016; Kraemer et al., 2012).

Table 5.
Interrater Reliability of overall PD Diagnosis of the AMPD

Study	N	Measures	PD	Statistic	IRR
Dereboy et al. (2018)	120	SCID-5-AMPD (Turkish version)	General PD	Kappa	.68
Somma et al. (2019)	84	SCID-5-AMPD	Borderline PD	Kappa	.83
Morey et al. (2019)	12	SCID-II with LPFS and PTRF*	Borderline PD	ICC (1,1)	.33

**PTRF= DSM-5 Clinicians' Personality Trait Rating Form*

Discussion

This review sought to summarise the literature concerning the IRR of the AMPD and its associated instruments. IRR is a fundamental component of reliability assessment within diagnostic systems and seeks to measure consistent application of diagnostic criteria (Kraemer et al., 2012). Previous issues concerning reliability of the categorical model of PD have been attributed to limited empirical evidence for categorical criteria, variations in the conceptualisation of personality and poor application of the model (Clark, 2007; Haslam et al., 2012;2020; Kotov et al., 2017). Therefore, it was hoped the provision of an empirically-based dimensional approach to PD diagnosis, would improve levels of IRR. Fifteen studies were included in the current review examining IRR of Criterion A, Criterion B, and overall diagnosis. ICCs in Criterion A and its individual domains were meta-analysed to provide a

pooled overall ICC for each of these. Sub-group and sensitivity analyses were carried out to explore high levels of heterogeneity and possible causes. Due to small study samples for Criterion B and overall diagnosis, IRR scores were reported in table format and discussed in relation to DSM-5 acceptability levels of IRR (Kraemer et al., 2012) and ICC reporting guidelines (Koo & Li, 2016).

LPFS Total Score

Meta-analysis found overall "good" IRR for the LPFS score across a total sample size of 902 (*ICC*=.75; 95% CI [.63 – .84]). This is above the acceptable threshold for inclusion in the DSM-5 (Kraemer et al., 2012) and higher than previous estimates of IRR of the categorical model (*k*=.40 to .52; Samuel, 2015). However, there was significant heterogeneity across these studies (I^2 =90.10%). Subgroup analysis highlighted considerable difference between assessment methods; assessment tools designed for the AMPD produced substantially higher IRR scores (*ICC=.87,* 95% CI [.79, .92]) than those that did not (*ICC*=.51, 95% CI [.42, .60]). Heterogeneity in the instrument group remained significant so results should be interpreted cautiously, however this tentatively supports previous research which suggests reliability is improved with the use of SCIs (Aboraya et al., 2006; Wood et al., 2002). This may also indicate inconsistency in either application or understanding of the model when using non-AMPD SCIs or relying solely on clinical judgement which has been found to be over-estimated by clinicians (Monahan, 1981; Kitamura & Kitamura, 2000). Research suggests these definitions vary considerably across culture (Caldwell-Harris & Ayçiçegi, 2006), however limited ethnicity reporting for either sample or raters makes this difficult to explore. Positively, the removal of high risk of bias studies during sensitivity analysis improved IRR (*ICC*=.80, 95% CI [70 – .87])

LPFS Individual Domains

All pooled ICCs of LPFS domains from a total sample of 775 fell in the "moderate" range for IRR and above the acceptable range of IRR for DSM-5 (>.40; Kraemer et al., 2012). However, heterogeneity remained significant across the domains and could not be fully explained through sensitivity analysis. Therefore, pooled ICCs should be interpreted with caution. Variability was found across the domains; *empathy* reported the lowest pooled agreement (*ICC*= .63, 95% CI [.43 – .77]) and *intimacy* the highest (*ICC*= .73, 95% CI [.54 – .85]). This suggests some domains may be more reliably agreed upon than others. Several factors could explain this finding. Firstly, there may be practical difficulties in assessing the domains. For example, intimacy dysfunction may be assessed through self or observer reports of relationships breakdowns. Whereas, assessment of empathy is more challenging, with additional skills in assessing reflective functioning required (Zimmerman et al., 2014). This may highlight a need for additional training and or the use of a specific AMPD SCI to guide this process. Secondly, the variation in IRR could highlight fundamental differences in the way the domains are understood by raters. This could be attributed to lack of familiarity with the concepts, difficulties in defining the concepts (particularly across cultures) or indeed flaws within the concepts themselves. This aligns with previous critiques of the model suggesting the domains are too complex to be universally understood (Pincus, 2011; Pilkonis et al., 2011).

Furthermore, trim and fill analysis highlighted a possible missing study within the *Identity* domain which could indicate publication bias (Duval &Tweedie, 2000; Viechtbauer, 2010). However, adding this study had no impact on the overall ICC score. No missing studies were found on any other domains. Removal of studies with high risk of bias improved domain scores (*Empathy ICC*= .73, 95% CI [.57-.84]; to *Intimacy ICC*=.82, 95% CI [.68-.90]), however variability between scores remained and heterogeneity remained significant.

Criterion B

Three studies reported IRR for Criterion B with a total sample size of 115, therefore meta-analysis was not deemed appropriate. This was unsurprising given the popularity of self-report measures for assessing personality traits (e.g., PID-5; Krueger et al., 2012). However, even with a robust self-report measure, self-report information cannot be relied upon. PD pathology may lead to inaccurate reporting due to lack of insight or deception (Carlson & Ottomanns, 2015). Therefore, further research into this area is important. However, of the studies that did report IRRs for personality trait domains, IRR scores varied both within and across studies. They ranged from excellent reliability (Disinhibition, *ICC*= .92 [Somma et al., 2020]) to poor reliability (Detachment, *ICC*= 0.43 [Morey, 2019]). Differences between raters could be attributed to the understanding and assessment of the criterion. Inhibition for example is a more familiar existing concept within mental health and also more easily observed and reported by others (e.g., impulsivity or risk taking). However, clinical expressions of detachment occur less frequently and can be more challenging to observe. Despite this, all scores are above the recommended threshold of acceptability within the DSM, which theoretically supports their inclusion in future editions of the manual (Kreamer et al., 2012). However, this also raises questions around the validity of acceptable thresholds of IRR (>.40) that would usually be classified as "poor" in other areas of scientific practice (Aboraya et al., 2006; Koo & Li, 2016). Promisingly, individual IRR statistics appear to indicate the use of a specific AMPD SCI increases overall reliability for personality trait domains as well as severity (e.g., Somma et al., 2020). However, it is difficult to make any conclusions around this based on such limited data.

Overall Diagnosis

Limited data was available for the IRR of overall PD diagnosis with a total sample of 216 reported from three studies. Thus, meta-analysis was not deemed appropriate. In the few

limited studies found by this review, the SCID-5-AMPD (Bender et al., 2018) continues to report higher IRR statistics in the summarised data, with both the English and Turkish version exceeding acceptable standards of IRR for the DSM-5 (Kraemer et al., 2012). However, there is a large amount of methodological variation between studies, therefore comparisons may be unwise. The study which fell below acceptability levels of the DSM (*ICC*=0.33; Morey, 2019) was based on written text using the DSM criteria only. Furthermore, two studies focus on Borderline PD, arguably the most familiar of the existing diagnoses, whereas the final study focused on the reliability of having or not having any PD. Although the information provided by these studies is useful, the overlap and disagreement within PD diagnosis is often between multiple traits or other diagnoses. Therefore, further studies examining more PD presentations would be helpful.

Limitations

To the best of the authors' knowledge, this is the first meta-analytic review of the IRR of the AMPD. However, this review does present with several limitations. Firstly, it is important to highlight the high levels of statistically significant heterogeneity within all meta-analyses which could not be conclusively explained. Although random effect models were used to increase generalizability, high heterogeneity does impact the confidence in which conclusions should be drawn from the results. Whilst it has not been possible to conclusively explain this heterogeneity through sensitivity analysis, it is likely caused by variation in methodological design within the studies including sample size and type (clinical vs non-clinical), rater number and experience in addition to methods of assessing and presenting cases (Higgins, 2008).

Secondly, methodological limitations are also important to consider. Searches were limited to English-language articles with translations not available despite contacting authors. Only published peer-review articles were included, which meant unpublished or grey

literature was potentially excluded. This may have led to the exclusion of potentially relevant studies. Furthermore, data extraction, coding and quality rating were mainly conducted by a single author, which could have led to reporting bias.

Thirdly, it is important to consider limitations around reporting of IRR statistics within studies. For the dimensional criteria, single-rater ICCs were the most frequent and appropriate reporting method. However, some studies were excluded for reporting different forms of ICC or other IRR statistics, with raw data not available. This was an important methodological decision to correctly pool data due to ICC of the "mean of *k* raters" always appearing larger than the corresponding single-rater type (Koo & Li, 2016). This was also more clinically appropriate as diagnoses in clinically practice are unlikely to be independently rated by a team of clinicians, so this inflated IRR is unlikely to be replicated in clinical practice.

Clinical Implications

The findings of this study are clinically important. Results indicate the pooled ICCs of the LPFS and individual domains all fall above the acceptable level for inclusion in DSM-5 (Kraemer et al., 2012). Furthermore, pooled ICCs suggested "good" reliability for LPFS and "moderate" reliability of the individual domains. It is hoped this will go some way to support the case for inclusion into future DSM editions. Furthermore, sub-group analysis shows that using a SCI designed for the AMPD (at least for Criterion A) improves reliability even further. This supports previous research and practice guidelines which recommend the use of SCIs in clinical practice (Samuel, 2015; Wood et al., 2002). It also advises caution around the use of unstructured clinical interviews. This mirrors previous research which suggests clinicians overestimate their ability in other areas of clinical decision making (Monahan, 1981; Kitamura & Kitamura, 2000). Despite this, the results show promising data for the severity domain of the AMPD and its learnability for clinicians, particularly considering

some studies opted for inexperienced or untrained raters. Similar patterns were also evidenced in Criterion B and overall diagnosis; however, conclusions were unable to be drawn due to limited data. However, it is important to highlight that even in research conditions, few "excellent" IRRs were reported across any domain. Taken together, this information may highlight the need for more consistent use of SCIs within clinical practice, additional training, or clarification around AMPD constructs and indicate the need for further research.

Future Research

Future research of IRR in Criterion B and overall diagnosis is warranted. Given tentative findings of this review, it would be beneficial to focus research using AMPD SCIs. For overall diagnosis, a particular focus on other presentations aside from Borderline PD would be beneficial. Furthermore, although the inclusion of inexperienced raters in research has been helpful to highlight learnability of the AMPD, further research focusing on implementation of the model in clinical practice would be advantageous to increase ecological findings. It will also be important for IRR studies to report both single-rater and "mean of *k* raters" ICCs for more accurate comparisons between studies to be made. Finally, it may be helpful to conduct this review again focussing on test-retest reliability which is considered another fundamental element of reliability alongside IRR (Kraemer et al., 2012).

Conclusions

Despite some methodological limitations, meta-analyses provided tentative support for the IRR of Criterion A of the AMPD. Pooled IRRs were reported for overall LPFS score and its specific domains, that all fell above the acceptable level for inclusion in the DSM and above previously report IRR scores for the categorical model of PD assessment. Sub-group analysis suggested IRR of LPFS score could be substantially improved using a SCI for assessment. Variation in ICC scores was observed across domains indicating some may be

more reliable than others. Tentative hypotheses were formed around the understanding and application of key constructs within the domains. However, heterogeneity was high in all analyses which could not be fully explained; therefore, results should be interpreted with caution. Less research has been conducted around Criterion B and overall PD diagnosis, with initial studies providing mixed results. Further research into these areas is recommended alongside a more specific focus on more ecologically valid methods utilising clinician administering SCIs designed specifically for the AMPD.

References

Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The Reliability of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis. *Psychiatry*, *3*(1), 41-50.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. (4th ed., revised). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. (5th ed.) Arlington (VA): American Psychiatric Association. https://doi.org/10.1176/APPI.BOOKS.9780890425596

Aviram, R. B., Brodsky, B. S., & Stanley, B. (2006). Borderline personality disorder, stigma, and treatment implications. *Harvard Review of Psychiatry*, *14*(5), 249–256. https://doi.org/10.1080/10673220600975121

Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are Informant Reports of Personality More Internally Consistent Than Self Reports of Personality? *Assessment*, *22*(4), 399. https://doi.org/10.1177/1073191114556100

Bender, D. S., Morey, L. C., & Skodol, A. E. (2011). Toward a Model for Assessing Level of Personality Functioning in DSM–5, Part I: A Review of Theory and Methods. *Journal of Personality Assessment*, *93*(4), 332–346. https://doi.org/10.1080/00223891.2011.583808

Bender, D. S., Skodol, A. E., First, M. B., & Oldham, J. M. (2018). Module I: Structured Clinical Interview for the Level of Personality Functioning Scale. In M. B. First, A. E. Skodol, D. S. Bender, & J. M. Oldham (Eds.), Structured Clinical Interview for the

DSM-5 Alternative Model for Personality Disorders (SCID-5-AMPD). *American Psychiatric Association.*

Boland, A., Cherry, G., & Dickson, R. (Eds.). (2017). *Doing a systematic review: A student's guide.* London: Sage.

Borenstein, M., Hedges, L. v., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. *Introduction to Meta-Analysis*, 1–421. https://doi.org/10.1002/9780470743386

Buer Christensen, T., Paap, M. C. S., Arnesen, M., Koritzinsky, K., Nysaeter, T.-E., Eikenaes, I., Germans Selvik, S., Walther, K., Torgersen, S., Bender, D. S., Skodol, A. E., Kvarstein, E., Pedersen, G., & Hummelen, B. (2018). Interrater reliability of the Structured Clinical Interview for the DSM–5 Alternative Model of Personality Disorders Module I: Level of Personality Functioning Scale. *Journal of Personality Assessment*, *100*(6), 630–641. https://doi.org/10.1080/00223891.2018.1483377

Caldwell-Harris, C. L., & Ayçiçegi, A. (2006). When Personality and Culture Clash: The Psychological Distress of Allocentrics in an Individualist Culture and Idiocentrics in a Collectivist Culture. *Collectivism (Particularly Vertical),* 43(3), 331–361. https://doi.org/10.1177/1363461506066982

Carlson, E. N., & Oltmanns, T. F. (2015). The Role of Metaperception in Personality Disorders: Do People with Personality Problems Know How Others Experience Their Personality? *Journal of Personality Disorders*, *29*(4), 449–467. https://doi.org/10.1521/PEDI.2015.29.4.449

Chinese Psychiatry Association. (2000). *Chinese classification of mental disorders. 3rd ed*. Shandong Science Press, Jinan.

Clark, L.A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual review of psychology, 58,* 227–257.https://doi.org/10.1146/annurev.psych.57.102904.190200

Clark, L. A., Vanderbleek, E. N., Shapiro, J. L., Nuzum, H., Allen, X., Daly, E., Kingsbury, T. J., Oiler, M., & Ro, E. (2015). The Brave New World of Personality Disorder-Trait Specified: Effects of Additional Definitions on Coverage, Prevalence, and Comorbidity: *Https://Doi.Org/10.5127/Pr.036314*, *a2*(1), 52–82. https://doi.org/10.5127/PR.036314

Clarkin, J. F., Caligor, E., Stern, B., & Kernberg, O. F. (2004). Structured interview of personality organization (STIPO). Personality Disorders Institute. *Weill Medical College of Cornell University.*

Clarkin, J. F., & Huprich, S. K. (2011). Do DSM-5 Personality Disorder Proposals Meet Criteria for Clinical Utility? *Journal of Personality Disorders. 25*(2), 192–205. https://doi.org/10.1521/PEDI.2011.25.2.192

Cochran, W. (1954). The Combination of Estimates from Different Experiments. Biometrics, 10(1), 101-129. https://doi:10.2307/3001666

Costa, P. T., Jr., & Widiger, T. A. (Eds.). (1994). *Personality disorders and the five-factor model of personality.* American Psychological Association. https://doi.org/10.1037/10140-000

Crawford, M. J., Koldobsky, N., Mulder, R., & Tyrer, P. (2011). Classifying Personality Disorder According to Severity. *Journal of Personality Disorders. 25*(3), 321–330. https://doi.org/10.1521/PEDI.2011.25.3.321

Cuchna, J. W., Hoch, M. C., & Hoch, J. M. (2016). The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. *Physical Therapy in Sport*, *19*, 57–65. https://doi.org/10.1016/J.PTSP.2015.12.002

Cuijpers, P. (2016). Meta-analyses in mental health research: A practical guide. Amsterdam, Netherlands, Arije universiteit Amsterdam.

Cruitt, P. J., Boudreaux, M. J., King, H. R., Oltmanns, J. R., & Oltmanns, T. F. (2019). Examining criterion a: DSM–5 level of personality functioning as assessed through life story interviews. *Personality Disorders: Theory, Research, and Treatment*, *10*(3), 224–234. https://doi.org/10.1037/per0000321

Dale, O., Sethi, F., Stanton, C., Evans, S., Barnicot, K., Sedgwick, R., Goldsack, S., Doran, M., Shoolbred, L., Samele, C., Urquia, N., Haigh, R., & Moran, P. (2017). Personality disorder services in England: findings from a national survey. *BJPsych Bulletin*, *41*(5), 247. https://doi.org/10.1192/PB.BP.116.055251

Dereboy, F., Dereboy, C., & Eskin, M. (2018). Validation of the DSM-5 Alternative Model Personality Disorder Diagnoses in Turkey, Part 1: LEAD Validity and Reliability of the Personality Functioning Ratings. *Journal of Personality Assessment*, *100*(6), 603–611. https://doi.org/https://dx.doi.org/10.1080/00223891.2018.1423989

Duval S, Tweedie R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 56(2):455-463.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult)*. London: Hodder and Stoughton.

First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., Davies, M., Borus, J., Howes, M. J., Kane, J., Pope, H. G., & Rounsaville, B. (2011). The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part II: Multi-Site Test-Retest Reliability Study. *Journal of Personality Disorders*, *9*(2), 92–104. https://doi.org/10.1521/PEDI.1995.9.2.92

Frances, A. (2012). *DSM-5 field trials discredit the American Psychiatric Association.* Huffington Post Science. The Blog. Posted 31 October 2012. Retrieved from: https://www.huffpost.com/entry/dsm-5-reliability-tests_b_1490857

Garcia, D. J., Skadberg, R. M., Schmidt, M., Bierma, S., Shorter, R. L., & Waugh, M. H. (2018). It's not that Difficult: An Interrater Reliability Study of the DSM-5 Section III Alternative Model for Personality Disorders. *Journal of Personality Assessment*, *100*(6), 612–620. https://doi.org/https://dx.doi.org/10.1080/00223891.2018.1428982

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine*, *42*(5), 903–920. https://doi.org/10.1017/S0033291711001966

Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: a meta-analysis of taxometric research. *Psychological Medicine*, *50*(9), 1418–1432. https://doi.org/10.1017/S003329172000183X

Hemmati, A., Rahmani, F., & Bach, B. (2021). The ICD-11 Personality Disorder Trait Model Fits the Kurdish Population Better Than the DSM-5 Trait Model. *Front. Psychiatry* 12:635813. https://doi:10.3389/fpsyt.2021.635813

Herkov, M. J., & Blashfield, R. K. (2010). Clinician Diagnoses of Personality Disorders: Evidence of a Hierarchical Structure. *Journal of Personality Assessment*, *65*(2), 313–321. https://doi.org/10.1207/S15327752JPA6502_7

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, 37(5), 1158–1160. https://doi.org/10.1093/ije/dyn204.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. https://doi.org/10.1002/sim.1186.

Hutsebaut, J., Berghuis, H., De Saeger, H., Kaasenbrood, A., & Ingenhoven, T. (2014). Semi-structured Interview for Personality Functioning DSM-5 (STiP-51). *The Podium DSM-5 Research Group of the Netherlands Centre of Expertise on Personality Disorders. Utrecht, the Netherlands: Trimbos Institute*.

Hutsebaut, J., Kamphuis, J. H., Feenstra, D. J., Weekers, L. C., & de Saeger, H. (2017). Assessing DSM5-oriented level of personality functioning: Development and psychometric evaluation of the semi-structured interview for personality functioning DSM5 (STiP-5.1). *Personality Disorders: Theory, Research, and Treatment*, *8*(1), 94–101. https://doi.org/10.1037/per0000197

Hutsebaut, J., Weekers, L. C., Tuin, N., Apeldoorn, J. S. P., & Bulten, E. (2021). Assessment of ICD-11 Personality Disorder Severity in Forensic Patients Using the Semi-structured Interview for Personality Functioning DSM-5 (STiP-5.1): Preliminary Findings. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.617702

Kampe, L., Zimmerman, J., Bender, D., Caligor, E., Borowski, A.-L., Ehrenthal, J. C., Benecke, C., & Hörz-Sagstetter, S. (2018). Comparison of the structured DSM–5 clinical interview for the Level of Personality Functioning Scale with the Structured Interview of Personality Organization. *Journal of Personality Assessment*, *100*(6), 642–649. https://doi.org/10.1080/00223891.2018.1489257

Kitamura, T., & Kitamura, F. (2000). Reliability of clinical judgment of patients' competency to give informed consent: A case vignette study. *Psychiatry and Clinical Neurosciences*, *54*(2), 245–247. https://doi.org/10.1046/J.1440-1819.2000.00665.X

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155. https://doi.org/10.1016/J.JCM.2016.02.012

Kotov, R., Jonas, K. G., Carpenter, W. T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Hobbs, K., Reininghaus, U., Slade, T., South, S. C., Sunderland, M., Waszczuk, M. A., Widiger, T. A., Wright, A., Zald, D. H., Krueger, R. F., & Watson, D. (2020). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. *World Psychiatry*, *19*(2), 151–172. https://doi.org/10.1002/wps.20730

Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. (2012). DSM-5: How Reliable Is Reliable Enough? *The American Journal of Psychiatry. 169*(1), 13–15. https://doi.org/10.1176/APPI.AJP.2011.11010050

Krueger, R. F., Skodol, A. E., Livesley, W. J., Shrout, P. E., & Huang, Y. (2007). Synthesizing dimensional and categorical approaches to personality disorders: refining the research agenda for DSM-V Axis II. *International Journal of Methods in Psychiatric Research*, *16*(1), S65. https://doi.org/10.1002/MPR.212

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial Construction of a Maladaptive Personality Trait Model and Inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879. https://doi.org/10.1017/S0033291711002674

Krueger, R. F., & Hobbs, K. A. (2020). An Overview of the DSM-5 Alternative Model of Personality Disorders. *Psychopathology*, *53*(3), 126–132. https://doi.org/10.1159/000508538

Lewis, G., & Appleby, L. (1988). Personality Disorder: the Patients Psychiatrists Dislike. *The British Journal of Psychiatry*, *153*(1), 44–49. https://doi.org/10.1192/BJP.153.1.44

Lorenzetti, D.L., Ghali, W.A. (2013) Reference management software for systematic reviews and meta-analyses: an exploration of usage and usability. *BMC Med Res Methodol* **13,** 141 https://doi.org/10.1186/1471-2288-13-141

Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of clinical epidemiology*, *63*(8), 854-861.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031/FULLARTICLE

Mezzich J. E. (2002). International surveys on the use of ICD-10 and related diagnostic systems. *Psychopathology*, *35*(2-3), 72–75. https://doi.org/10.1159/000065122

Milton, J. (2000). A postal survey of the assessment procedure for personality disorder in forensic settings. *Psychiatric Bulletin*, *24*(7), 254–257. https://doi.org/10.1192/PB.24.7.254

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, *339*(7716), 332–336. https://doi.org/10.1136/BMJ.B2535

Monahan, J. (1981). *Predicting Violent Behavior - An Assessment of Clinical Techniques.* Office of Justice Programs, Sage, Beverly Hills, CA.

Morey, L. C. (2019). Application of the DSM-5 Level of Personality Functioning Scale by Lay Raters. *Journal of Personality Disorders*, *32*(5), 709–720. https://doi.org/10.1521/pedi_2017_31_305

Mulder, R. T. (2021). ICD-11 Personality Disorders: Utility and Implications of the New Model. *Frontiers in Psychiatry*, *12*, 709. https://doi.org/10.3389/FPSYT.2021.655548/BIBTEX

Mulder, R. T., Newton-Howes, G., Crawford, M. J., & Tyrer, P. J. (2011). The Central Domains of Personality Pathology in Psychiatric Patients. *Journal of Personality Disorders*, *25*(3), 364–377. https://doi.org/10.1521/PEDI.2011.25.3.364

Mulder, R., & Tyrer, P. (2019). Diagnosis and classification of personality disorders: novel approaches. *Current Opinion in Psychiatry*, *32*(1), 27-31. https://doi:10.1097/YCO.0000000000000046

National Institute for Mental Health England. (2003). *Personality disorder: no longer a diagnosis of exclusion. Policy implementation guidance for the development of services for people with personality disorder.* London: Department of Health.

Newton-Howes, G., Clark, L. A., & Chanen, A. (2015). Personality disorder across the life course. *The Lancet*, *385*(9969), 727–734. https://doi.org/10.1016/S0140-6736(14)61283-6

Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *The American Journal of Psychiatry*, *149*(12), 1645–1653. https://doi.org/10.1176/AJP.149.12.1645

Pilkonis, P. A., Hallquist, M. N., Morse, J. Q., & Stepp, S. D. (2011). Striking the (Im)proper balance between scientific advances and clinical utility: Commentary on the DSM-5 proposal for personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *2*(1), 68–82. https://doi.org/10.1037/A0022226

Pincus, A. L. (2011). Some comments on nomology, diagnostic process, and narcissistic personality disorder in the DSM-5 proposal for personality and personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *2*(1), 41–53. https://doi.org/10.1037/A0021191

Roche, M. J., Jacobson, N. C., & Phillips, J. J. (2018). Expanding the Validity of the Level of Personality Functioning Scale Observer Report and Self-Report Versions Across Psychodynamic and Interpersonal Paradigms. https://doi.org/10.1080/00223891.2018.1475394

Preti, E., di Pierro, R., Costantini, G., Benzi, I. M. A., de Panfilis, C., & Madeddu, F. (2018). Using the Structured Interview of Personality Organization for DSM–5 Level of Personality Functioning Rating Performed by Inexperienced Raters. *Journal of Personality Assessment*, *100*(6), 621–629. https://doi.org/10.1080/00223891.2018.1448985

Samuel, D. B., Sanislow, C. A., Hopwood, C. J., Shea, M. T., Skodol, A. E., Morey, L. C., ... & Grilo, C. M. (2013). Convergent and incremental predictive validity of clinician, self-report, and structured interview diagnoses for personality disorders over 5 years. *Journal of consulting and clinical psychology*, *81*(4), 650.

Samuel, D. B. (2015). A review of the agreement between clinicians' personality disorder diagnoses and those from other methods and sources. *Clinical Psychology: Science and Practice,* 22(1), 1–19. https://doi.org/10.1111/cpsp.12088

Sansone, R. A., & Sansone, L. A. (2011). Gender Patterns in Borderline Personality Disorder. *Innovations in Clinical Neuroscience*, *8*(5), 16-20.

Schuster, C. (2004). A Note on the Interpretation of Weighted Kappa and its Relations to Other Rater Agreement Statistics for Metric Scales. *Educational and Psychological Measurement*, *64*(2), 243–253. https://doi.org/10.1177/0013164403260197

Skodol, A. E., Gunderson, J. G., McGlashan, T. H., Dyck, I. R., Stout, R. L., Bender, D. S., Grilo, C. M., Tracie Shea, M., Zanarini, M. C., Morey, L. C., Sanislow, C. A., & Oldham, J. M. (2002). Functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder. *American Journal of Psychiatry*, *159*(2), 276–283. https://doi.org/10.1176/APPI.AJP.159.2.276

Somma, A., Borroni, S., Carlotta, D., Giarolli, L. E., Barranca, M., Cerioli, C., Franzoni, C., Masci, E., Manini, R., Busso, S. L., Ruotolo, G., & Fossati, A. (2019). The inter-rater reliability and convergent validity of the Italian translation of the Structured Clinical

Interview for the DSM-5 Alternative Model of Personality Disorders Module III in a psychotherapy outpatient sample. *Journal of Psychopathology*, *25*(4), 195–204.

Somma, A., Borroni, S., Gialdi, G., Carlotta, D., Emanuela Giarolli, L., Barranca, M., Cerioli, C., Franzoni, C., Masci, E., Manini, R., Luca Busso, S., Ruotolo, G., Krueger, R. F., Markon, K. E., & Fossati, A. (2020). The Inter-Rater Reliability and Validity of the Italian Translation of the Structured Clinical Interview for DSM-5 Alternative Model for Personality Disorders Module I and Module II: A Preliminary Report on Consecutively Admitted Psychotherapy Outpatients. *Journal of Personality Disorders*, *34*(Supplement C), 95–123. https://doi.org/https://dx.doi.org/10.1521/pedi_2020_34_511

Thylstrup, B., Simonsen, S., Nemery, C., Simonsen, E., Noll, J. F., Myatt, M. W., & Hesse, M. (2016). Assessment of personality-related levels of functioning: A pilot study of clinical assessment of the DSM-5 level of personality functioning based on a semi-structured interview. *BMC Psychiatry*, *16*(1), 1–8. https://doi.org/10.1186/S12888-016-1011-6/TABLES/2

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tyrer, P. (2012). Diagnostic and Statistical Manual of Mental Disorders: A classification of personality disorders that has had its day [Editorial]. *Clinical Psychology & Psychotherapy*, 19(5), 372–374. https://doi.org/10.1002/cpp.1810

Tyrer, P. (2014). A comparison of DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment, 20*(4), 280-285. https://doi:10.1192/apt.bp.113.011296

Verheul, R., & Widiger, T. A. (2004). A meta-analysis of the prevalence and usage of the personality disorder not otherwise specified (PDNOS) diagnosis. *Journal of Personality Disorders*, *18*(4), 309–319. https://doi.org/10.1521/PEDI.18.4.309.40350

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. https://doi:10.18637/jss.v036.i03

Wallace, B. C., Dahabreh, I. J., Trikalinos, T. A., Lau, J., Trow, P., & Schmid, C. H. (2012). Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. Journal of Statistical Software, 49(5), 1–15.https://doi.org/10.18637/jss.v049.i05

Westen, D., & Shedler, J. (1999). Revising and assessing Axis II, Part 2: Towards an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry*, 156, 273-285.

Widiger, T. A., Livesley, W. J., & Clark, L. A. (2009). An Integrative Dimensional Classification of Personality Disorder. *Psychological Assessment*, *21*(3), 243–255. https://doi.org/10.1037/A0016606

Winsper, C., Bilgin, A., Thompson, A., Marwaha, S., Chanen, A. M., Singh, S. P., Wang, A., & Furtado, V. (2019). *The prevalence of personality disorders in the community: a global systematic review and meta-analysis*. https://doi.org/10.1192/bjp.2019.166

Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology, 53*(1), 519-543. https://doi.org/10.1146/annurev.psych.53.100901.135136

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva: World Health Organization.

Wright, A., & Zimmerman, J. (2015). At the nexus of science and practice: Answering basic clinical questions in personality disorder assessment and diagnosis with quantitative

modeling techniques. *Personality Disorders: Toward Theoretical and Empirical Integration in Diagnosis and Assessment.*, 109–144. https://doi.org/10.1037/14549-006

Zimmerman, J., Benecke, C., Bender, D. S., Skodol, A. E., Schauenburg, H., Cierpka, M., & Leising, D. (2014). Assessing DSM–5 Level of Personality Functioning From Videotaped Clinical Interviews: A Pilot Study With Untrained and Clinically Inexperienced Students. *96*(4), 397–409. https://doi.org/10.1080/00223891.2013.852563

Zettl, M., Volkert, J., Vogele, C., Herpertz, S. C., Kubera, K. M., & Taubner, S. (2020). Mentalization and criterion a of the alternative model for personality disorders: Results from a clinical and nonclinical sample. *Personality Disorders*, *11*(3), 191–201. https://doi.org/https://dx.doi.org/10.1037/per0000356

Chapter Three: Bridging Chapter

The meta-analysis presented in Chapter Two sought to summarise existing literature on the interrater reliability (IRR) of the Alternative Model of Personality Disorder (AMPD) currently included in Section III of the Diagnostic and Statistical Manual for Mental Disorders – Fifth Edition (DSM-5; APA, 2013). Despite some methodological limitations, meta-analysis provided tentative support for improved IRR of the AMPD when compared to the primary categorical model of Personality Disorder (PD) diagnosis. Pooled IRRs for Level of Personality Function Scale (LPFS; Bender et al., 2011) total score and individual domains fell above acceptability levels for DSM inclusion. Furthermore, sub-group analysis of Total LPFS score, suggests that IRR of the model could be further improved by using a specific AMPD structured clinical interview (SCI) as the primary assessment tool. Despite this, there was still some variation in IRR between the LPFS domains, with some domains (e.g., *intimacy*) reporting higher IRR scores than others (e.g., *empathy*). Less research has been conducted concerning the IRR of Criterion B and overall PD diagnosis, however similar patterns were found within the individual studies for the trait domains also. It was hypothesised that the domains demonstrating more disagreement between raters might be explained by differences in comprehension, interpretation, and application of the model. This is consistent with critiques of the model which suggest the AMPD may be too complex for clinical practice (Clarkin & Huprich, 2011; Tyrer, 2012; Pincus, 2011; Pilkonis et al., 2011), however this pattern was not seen across all domains.

Of course, it is acknowledged that despite improved IRR of PD diagnosis, excellent agreement was not achieved. Although it is possible to increase the reliability of these diagnoses with structured clinical assessment tools, these are rarely used within clinical settings (Milton, 2000; Perry, 1992). Furthermore, increased agreement between clinicians does not guarantee the criteria they are rating are meaningful or valid clusters of symptoms

that can predict the effectiveness of certain treatments. Therefore, one might reasonably question the appropriateness of providing a highly stigmatic diagnostic label on this basis.

This problem, however, is not exclusive to PD diagnoses. Many studies have highlighted the unreliability of diagnostic concepts (e.g., Aboraya, 2007; Bentall, 2003; Kinderman et al., 2013). The Chair of the DSM-5 task force himself highlighted the difficulties of diagnosis, explaining they had been waiting since the seventies to identify biological and genetic markers that lead to precise diagnoses (Kupfer, 2013). However, fifty years on, this still seems a somewhat futuristic concept. Furthermore, the head of the DSM-IV taskforce referred to DSM-5 Field Trials as a "disgrace to the field" (Frances, 2012). This was largely due to low levels of diagnostic agreement in the DSM-5 Field Trials, alongside attempts to minimise this by reducing reliability thresholds for diagnoses to be included in the manual.

Within Clinical Psychology, there has been a growing movement towards an individual formulation-based understanding of an individual's difficulties. One approach which has attempted to stimulate thinking in this area, particularly emphasising the role of the social context of mental health problems, is the Power, Threat, Meaning Framework (Johnstone & Boyle, 2018). However, there has been some resistance to this model (e.g., Salkovskis & Edge, 2018) and, for now, psychiatric diagnosis remains the dominant model within the Western world.

Detached from healthcare, psychiatric diagnosis also has a significant impact on other systems within society. Perhaps one of the most relevant of these systems is the Criminal Justice System in England and Wales. When an offender is thought to be suffering from a mental health disorder (within the meaning of s.1 of the Mental Health Act 1983 [MHA]), the court must obtain and consider a medical report before passing a custodial sentence (Sentencing Council, 2020). Significant weight is placed on this expert witness testimony,

which is usually provided by two medical doctors (usually psychiatrists), including one who is section 12 approved (MHA, 1983). Arguably, poor diagnostic IRR is often evident in the courtroom, where conflicting expert witness testimonies is commonplace (Peay, 2016). Regardless of the potential reason for these differences (and certainly the unreliability of diagnostic concepts may be one contributing factor), it is notable that the responsibility of weighing up this complex evidence falls to the judge, who has limited mental health training (Latham, 2017). Considering it is it is estimated that anywhere between 52% (HM Chief Inspector of Prisons, 2021) and 90% (Singleton et al., 1998) of the prison population has a "diagnosable mental health condition", diagnostic unreliability is likely to have strong negative consequences for many offenders. In fact, this real-life impact can be easily observed by studying Case Law. For example, in the case of *R. v Challen* (2019), a psychiatrist deemed the defendant to show no evidence of any mental illness and the defendant was found guilty of murder. However, a second psychiatrist then assessed Ms Challen and found her to have a diagnosis of Borderline PD and possibly Bi-Polar Disorder. Upon appeal, the conviction was quashed based on these new diagnoses. Furthermore, in the case of *R. v Fort* (2013), an 18-year-old was imprisoned for murdering his mother. At sentence, psychiatric evidence alluded to possible diagnosis of Autism, and he was sentenced to custody for life. Subsequently, whilst in custody he received a diagnosis of Schizoid Personality Disorder which eventually led to his sentence being changed to Section 45A of the Mental Health Act (1983), a hybrid hospital and penal order. Whilst being treated in hospital, he was later diagnosed with Paranoid Schizophrenia and, upon appeal, given a hospital treatment order.

The aforementioned examples from Case Law go some way to highlight the impact of psychiatric diagnosis on the sentencing and treatment of offenders with mental health difficulties. The reasons for this are multifactorial with diagnosis alone not being adequate to

avoid sentencing. However, if diagnostic reliability is poor and certain diagnoses are viewed as less treatable and more stigmatised than others (Lewis & Appleby, 1988), one can assume they have a considerable impact on courtroom decision. However, the diagnoses themselves, as mentioned, are only one part of the story and ultimately judges are responsible for the final sentencing decision.

The following chapter will therefore present an empirical paper which begins by outlining relevant research in relation to the sentencing of offenders with mental health problems, the unreliability of psychiatric diagnoses and associated stigmatic attitudes to these diagnoses. It will aim to explore whether sentencing outcomes can be predicted by psychiatric diagnosis, mental illness stigma or the extent to which sentencers agree with the guidance they are directed to follow. The empirical study aims, methodology, research findings and a discussion will follow before concluding with an overall critical review of the thesis portfolio.

Chapter Four: Empirical Project

An experimental clinical vignette study to explore the relationship between perceived mental illness stigma, psychiatric diagnosis and sentencing outcomes of offenders with mental health difficulties.

Samantha E. Young[a]*, George Baldwin[a], Ian Edwards[b] and Peter Beazley[a]

University of East Anglia

[a] Department of Clinical Psychology and Psychological Therapies, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, United Kingdom.

[b] School of Law, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, United Kingdom.

*Corresponding author: samantha.young@uea.ac.uk

Word Count (including references): 7,934 (including references).

This paper is being prepared for submission to *Psychiatry, Psychology and Law*.
See Appendix A for author guidelines. The word limit for this journal is 12000 inclusive of tables, references, figure captions.

This review did not receive a grant or funding and no conflicts of interest were reported. A data availability link is not provided as ethical permission was not requested to share data beyond the original study.

Abstract

The sentencing of offenders with mental health difficulties (OMHDs) is an under-researched area. Recent changes to sentencing guidelines now instruct judges to consider a hybrid treatment and custodial sentence (Section 45A of the Mental Health Act, 1983), before a hospital disposal. Limited guidance is available for the application of s.45A, and little is known about the impact of extra-legal factors on this process. This online study employed an experimental video-vignette methodology to explore the impact of psychiatric diagnosis, perceived mental illness stigma and *Vowles* Criteria on the sentencing of OMHDs. Participants acted as proxy judges and were allocated to one of three experimentally-manipulated diagnostic conditions. Sentencing outcome was significantly predicted by the extent to which a participant agreed the offender required punishment for the offence. However, no other significant predictors were found including mental illness stigma. Recommendations for future studies with a judicial population and implications for sentencing are discussed.

Keywords: Stigma, mental health, sentencing, criminal justice, extra-legal, personality disorder, schizophrenia

Introduction

The sentencing of offenders with mental health difficulties (OMHDs) is an under-researched area. In fact, very few studies have even looked at the prevalence rates of mental health problems in courts in the UK (Bradley Report, 2009). However, it is estimated that anywhere between 52% (HM Chief Inspector of Prisons, 2021) and 90% (Singleton et al., 1998) of the prison population has a "diagnosable mental health condition". Despite high prevalence rates in the criminal justice system, judges and legal professionals receive limited mental health training, with recent reports highlighting the need for this to be made mandatory (Latham, 2017). This can be problematic as judges make independent decisions on sentencing and can therefore overrule clinical opinion with no prior mental health knowledge (Peay, 2016). The consequences for OMHDs can be significant, with estimates suggesting over half of OMHDs do not have their mental health needs met whilst in prison (Jakobowitz et al., 2017). Since sentencing options can decide whether an individual is sent to prison or a specialist hospital, for instance, it is imperative that effective sentencing processes exist for OMHDs.

In England and Wales, when a jury has found a defendant guilty, it is the judge's responsibility to pass a sentence. If two medical doctors agree an offender has a mental illness (within the meaning of s.1 of the Mental Health Act 1983 [MHA]) a judge has several options at their disposal; they can pass a custodial sentence, a hospital order (s.37 MHA) with or without restrictions (s.41) or a Hospital and Limitation Direction (s.45A). S.37 is not considered a punishment and cannot include a penal element (s.142, Criminal Justice Act 2003). S.45A combines an initial period in hospital with a custodial sentence and was originally developed under the MHA (1983) for use with offenders diagnosed with "psychopathic disorder" during a period of increased homicides perpetrated by OMHDs (Royal College of Psychiatrists, 1996). The disposal was widely criticised by psychiatrists

(Chiswick, 1996; Eastman, 1996). However, when the MHA was amended in 2007, the scope of this section widened. Historically, there has been relatively low use of the order with 23 orders passed between 1997 and 2004 in England and Wales (Li & Foster, 2005). Low use has been attributed to a lack of awareness and ethical objections to the hybrid order (Whyte & Gupta, 2007). However, use of the order has been increasing steadily since 2007, and the Ministry of Justice (2018) reported a 39% increase in its use between 2013 and 2017. It is argued that this increase has been at the expense of OMHDs who would have previously been given a s.37 (Delmage et al., 2015). This increase is explained by the judgement in *R v Vowles (2015)* where the Court of Appeal ruled that, in all cases where a s.37 could be considered appropriate, a s.45A must always be considered first. The Court of Appeal also encouraged sentencers to consider four criteria before applying a s.45A could be applied; (1) the extent to which the offender's mental health requires treatment, (2) the extent to which offending is attributable to the mental health disorder, (3) the extent to which offending requires punishment and (4) the protection of the public when deciding release and regime of release. For present purposes these are referred to as the "*Vowles* Criteria". This guidance was upheld in the case of *R v Edwards* (2018) despite a previous appeal suggesting that supervision under s.37 offered enhanced public protection (*R v Ahmed*, 2016). This signified a change in sentencing for OMHDs as, prior to this, the law had allowed a court to deal with such an offender in the manner it considers to be "most appropriate" (s.166 of the Criminal Justice Act, 2003 [5]). This was further highlighted in the introduction of new sentencing guidelines that state before making a hospital order, sentencers must consider whether a s.45A would be more appropriate (Sentencing Council, 2020). The priority for OMHDs appears to have moved away from treatment as the priority and towards the priority of punishment.

Whilst the foregoing has emphasised the legal aspects of s.45A, the implications and clinical considerations – both in the decision making of sentencing, and then subsequently in the impact for clients – are significant. Research has highlighted clinical concerns over the use of s.45A including the "revolving door syndrome" of offenders (Peay, 2016). In practice, mental health needs of OMHD's often cannot be met in prison due to limited staffing and resources (O'Loughlin, 2021). Furthermore, there are clinical limitations of a s.45A during the hospital treatment phase, with individuals under this section unable to access important rehabilitative processes such as leave (MOJ, 2017). The regime for release also differs, with those on s.37 having a lower threshold for re-admission to hospital for deterioration in mental state and gaining eligibility to s.117 aftercare, which is not available on s.45A (Peay, 2015). Interestingly, research indicates lower rates of re-offending in forensic psychiatric populations compared to an equivalent prison population (Fazel et al., 2016). During the consultation process for the Wessely review of mental health legislation, no forensic psychiatrists came forward to commend the order (Taylor et al., 2021).

Sentencing decisions can only be made by a judge, however expert witness often make recommendations for sentencing (Beech et al., 2019). In a recent qualitative study, psychiatrists identified diagnosis as a key influencing factor for recommending a s.45A, with a diagnosis of personality disorder (PD) more likely to attract a s.45A, but a primary diagnosis of psychosis less likely, with a s.37 preferred instead (Beech et al., 2019). This is particularly problematic given the literature around the reliability and validity of diagnostic concepts; multiple studies have highlighted the unreliability of psychiatric diagnosis (Bentall, 2003; Kinderman et al., 2013; Aboraya, 2007). Furthermore, heterogeneity within individual diagnoses, and overlaps between diagnostic groups are high (Olbert et al., 2014; Allsopp et al., 2019). Yet, during trials, psychiatrists often present diagnoses in court as fact, even when opposing viewpoints are presented by other expert witnesses (Guthiel et al., 2006).

Regarding the process of decision making around sentencing, this is likely influenced by a number of factors that may be drawn from wider psychological literature. Faced with such uncertain decisions, judges will inevitably revert to heuristic techniques (Tversky, & Kahneman, 1974) which may increase the likelihood of cognitive and emotional biases (Ebbesen & Konecni, 1975; Sensibaugh & Allgeier, 1996). Yet, there is a distinct lack of research on this topic, with none considering the impact on sentencing in the context of mental health issues in the English and Welsh justice system. There is, however, good reason to think that such biases will be particularly relevant to OMHDs; nine out of ten people who experience mental health problems report experiencing stigma and discrimination (Corker et al., 2016). This is interesting from a Clinical Psychology perspective where structural stigma theory proposes that mental illness stigma can result in discrimination (Corrigan et al., 2004). Mental illness stigma creates increases inequality and wider societal discrimination including access to healthcare, employment, and education (Corrigan et al., 2006). The presence of a mental health label has been reported to increase predictions of recidivism (Nee & Witt, 2013). Furthermore, beliefs associated with particular diagnoses – whatever their underlying reliability – may also be particularly important. Although individuals diagnosed with schizophrenia are more likely to be seen as dangerous and unpredictable (Levey et al., 1995; Crisp et al., 2000; Varshney et al., 2016); studies also highlight that labelling offenders with psychotic disorders mitigates mock jurors' perceptions of blameworthiness and death penalty support (Edens et al., 2005). Although it is important to note, that individuals with schizophrenia are more likely to be a victim, rather than a perpetrator, of a violent crime (Brekke et al., 2002). In contrast, there is a lack of knowledge regarding PDs in the general public and, as a result, a risk that individuals with this diagnosis may be deemed to be "purposefully misbehaving" rather than experiencing an illness (Baker at al., 2021; Sheehan et al., 2016).

PD is a common mental health diagnosis with prevalence rates of 5% in the UK (Coid et al., 2006) and estimates of up to 70% in the prison population (MoJ & DoH, 2011). Research suggests that PD is more stigmatised than other mental illnesses including schizophrenia (Sheehan et al., 2016). Moreover, it has been reported that the public are less likely to view those with a PD as needing professional help, and more likely to view individuals as being able to cope on their own compared to other psychiatric diagnoses (Furnham et al., 2015). This conflation of responsibility and control around PDs arguably does not bode well in the court room, where culpability increases both the seriousness of the offence and the subsequent sentence (Sentencing Council, 2020). In fact, Blais and Forth (2014) found that an anti-social PD diagnosis increased the likelihood of a guilty verdict compared to those without any mental health diagnoses. Unfortunately, this perceived stigma towards PD has also been found to be present in healthcare professionals (Markham & Trower, 2003) and structurally within services making it difficult to access treatment (National Institute for Mental Health England, 2003). Research suggests that psychiatrists as a profession are more likely to view PDs as "untreatable" which may influence their recommendations as expert witnesses at sentencing (Lewis & Appleby, 1988). This may be due to effective treatments usually being more psychologically than psychiatrically led (NICE, 2015).

Goffman (1963, p.6) defines a stigma as when "an individual with an attribute which is deeply discredited by his/her society is rejected as a result of the attribute". One way of conceptualising this phenomenon may be through labelling theory (Scheff, 1966; Link et al., 1987). Labelling theory constructs mental illness as a "deviance" from societal norms (Rosenfield, 1997) and hypothesises that an individual's behaviour and self-identity are driven by the labels used to describe and classify them (Scheff, 1966,1974). The theory received some criticism for being too extreme (Gove, 1970) and a modified labelling

approach was developed highlighting the importance of stigma and stereotyping in mental illness outcomes (Link et al., 1987,1989). Link and Phelan (2001) emphasised the role of power within stigma. This may be important for OMHDs who are classified as deviating from societal norms not only for their experiences of mental health, but also as offenders (Rosenfield, 1997; Winnick, & Bodkin, 2008). Mental Illness Stigma literature broadly defines stigma as public (public attitudes which lead to discriminatory behaviour [Corrigan & Watson, 2002]) and self-stigma (when this public stigma is internalised and leads to negative consequences for the individual; [Corrigan & Rao, 2012]). In the context of sentencing, public stigma can trigger individual and structural discrimination, with the latter in the form of laws and sentencing guidelines which do not incorporate the needs of OMHDs (Link & Phelan, 2014). Thus, one may reasonably expect an interaction between stigmatic attitudes and sentencing, with increased stigmatising attitudes leading to more punitive decisions towards OMHDs.

The current study

The primary aim of the current study was to investigate whether the provision of psychiatric diagnosis, levels of stigmatic attitudes towards mental health problems in proxy judges and agreement with the four separate *Vowles* Criteria could predict the sentencing outcome in a clinical experimental vignette design.

Hypotheses

Hypothesis 1: Participants are more likely to recommend a custodial sentence when presented with an experimental vignette with PD compared to schizophrenia or complex mental health.

Hypothesis 2: Participants with highly stigmatised attitudes towards mental health problems are more likely to recommend a custodial sentence in an experimental vignette.

Hypothesis 3: The stronger the agreement with the "extent to which the offender's mental health requires treatment", the lower the likelihood of a custodial sentence.

Hypothesis 4: The stronger the agreement with, "the extent to which offending is attributable to the mental health disorder", the lower the likelihood of a custodial sentence.

Hypothesis 5: The stronger the agreement with, "the extent to which offending requires punishment", the higher the likelihood of a custodial sentence.

Hypothesis 6: The stronger the agreement with, "the protection of the public when deciding release and regime of release", the higher the likelihood of a custodial sentence.

Method

Design

An experimental design was used to test the above hypotheses using experimental clinical vignettes. The study took place online involving between-subjects randomisation to diagnosis (independent variable, three levels: schizophrenia, Borderline PD [BPD] and complex mental health problems) and within-subjects measurement of the two dependent variables – sentencing outcome (s.37, s.45A or prison), the *Vowles* criteria scales and the Perceived Devaluation and Discrimination Scale (PDD; Link et al., 1987). BPD was selected over other PD diagnoses for this study as it is the most researched PD (Bateman et al., 2015), the only distinct PD diagnostic pattern to be preserved in the updated ICD-11 (WHO, 2018), and one of the most well-known and prevalent PD diagnoses within the offending and inpatient psychiatric population (Bateman et al., 2015; Singleton et al., 1998). BPD was privileged over Anti-Social PD (the most prevalent PD within the prison population [Coid et al., 2006; Singleton et al., 1998]) due to the increased levels of stigma and recidivism associated with this diagnosis (Blais & Forth, 2004). Furthermore, PD pathways in England and Wales recommend treatment within the OPDPP, located within the Criminal Justice System, as

opposed to a hospital admission (NOMS, 2015). The "complex mental health" condition is taken from a previous study exploring jury decision making and intended to act as a control as per previous research (Baker at al., 2021).

Vignette based methodologies are frequently used in decision making research and bridge the high internal validity of experiments and the high external validity of survey research (Evans et al., 2015). They also seek to mediate the high social desirability bias that occurs in interviews (Spencer et al., 2015).

Measures

Perceived Devaluation and Discrimination Scale

Corrigan and Rao (2012) identified 22 different measures relating to stigma in mental illness. Due to the multiple measures and relative novelty of the phenomenon, Yang (2015) recommends measures are chosen in line with theoretical relevance. This study seeks to explore public stigma (as opposed to internalised *self-stigma*) and as such requires a perceived mental illness stigma measure which measures awareness of public stigma (Brohan et al., 2010). The PDD is the most used measure in this category (Brohan et al, 2010) and is a 12-item measure that looks at the extent to which a person believes that other people will devalue or discriminate against someone with a mental illness (Link et al., 1987). Items are measured on a six-point Likert scale of agreement. A scaled score is used by dividing the total score by 12 (number of items). Because the midpoint of the scale is 3.5 and is located directly between the strongly agree/strongly disagree poles, a mean above 3.5 indicates that the average person tends to endorse an item or a pool of items (as with the summative scale).

The PDD has been reported to have good psychometric properties with the internal consistency of the scale ranging from $\alpha = .75$ to $\alpha = .88$ (Ahn et al., 2015; Link et al., 2001) and a test-retest reliability score of .93 (Yang, 2015). In the current study, the scale demonstrated good reliability (Cronbach's alpha=.90). Furthermore, the measure frames

questions by asking what "most people" believe about people with a mental illness, which helps to reduce levels of socially desirable reporting (Link & Cullen, 1983).

Vowles Criteria

Each *Vowles* criteria were measured using a six-point Likert scale with responses ranging from strongly agree to strongly disagree (Appendix H). Each criterion was treated as an individual variable and not summed to make a total score.

Participants

A total of 285 participants completed the survey. Eighty-eight were excluded for failing the knowledge check questions (Appendix I). This left 197 participants included in the final data analysis. Females were overrepresented in the sample (77.2% female, 22.8% male). Participant age ranged between 18 and 69 years with a mean of 33.57 (SD=10.16) years. Participants came from a wide range of backgrounds, however only 7% of participants had experience in Criminal Law (see Table 1 for further demographics). "Non-legal" participants were non-professionals working in a legal setting e.g., managerial role.

Table 1

Participant demographics

Demographics	Percentage of sample % (n=197)
Age (years)	
M = 33.57, *SD* = 10.16, range =18-69	
Legal Experience -Years since qualification (N=89)	
M = 9.40 *SD* = 9.27	
Gender	
Female	77.20
Male	22.80
Ethnicity	
White English/Welsh/Scottish/Northern Irish/British	78.6
Other white background	6.60
Any other	14.79
Experience of mental health	
Self - None at all	8.60
A little	40.10
A moderate amount	28.90
A lot	14.20
A great deal	8.10
Other - None at all	1.00
A little	23.40
A moderate amount	27.90
A lot	30.50
A great deal	17.30
Occupation/Legal Experience	
Legal Professional	73.00
Other legal	*50.00*
Solicitor	*34.70*
Law student	*11.80*
Barrister	*3.50*
Non-legal professional	27.00
Area of Law (N=197)	
Criminal	7.10
Other (incl. property, commercial, corporate etc.)	56.90
N/A	36.00

Sample Size and Power

A prospective power analysis was carried out to estimate sample size for logistic regression. Green (1991) recommends N > 50 + 8m (where m is the number of IVs). In this case, the recommended sample would be >98.

Sampling Procedure

A three-staged recruitment strategy was implemented. Initially, a sample of judges currently practicing in England or Wales were sought through the Judicial Office (Appendix J). Secondly, recruitment of legal professionals took place through national law societies and on social media. Finally, the online survey participant platform, Prolific (www.prolific.co), was used when recruitment from previous methods proved challenging. Recruitment took place jointly with another Trainee (Appendix A) and advertising materials can be found in Appendix K. Baseline demographics (gender, age and ethnicity) were collected to assess group equivalences. Recruitment took place between April and October 2021. Participants recruited through Prolific were paid a small amount as a token for completing the study. Participants recruited through all other methods were entered into a prize draw with the chance of winning an e-voucher.

Procedure

The study was conducted online via a Qualtrics survey link. Participants were initially screened to ensure they met the inclusion criteria and were then directed to a lay summary (Appendix L) an online information sheet (Appendix M) and asked to provide informed consent (Appendix N). Participants completed a brief questionnaire gathering demographic information including age, gender, ethnicity, legal area, experience and mental health experience. Following completion, participants were randomly assigned to a diagnostic condition (schizophrenia, PD, or complex mental health). The “complex mental health” condition intended to act as a control as per previous research (Baker at al., 2021).

Participants were then asked to view a brief video vignette presenting pre-sentence information. Each video vignette described an offender who had been found guilty of Grievous Bodily Harm with Intent (GBH; Offences against the Person Act 1861, s.18), an offence punishable by a custodial sentence. The only difference between the three vignettes was the diagnosis; the vignettes were otherwise identical. The vignette script was created in conjunction with a legal expert to increase ecological validity (Appendix O). After participants viewed the video, they were provided with information listing and defining three sentencing options; prison, s.37 or s.45A (Appendix P).

Finally, due to the online nature of this study, a knowledge check was used to check the validity of this data collection method. Two multiple-choice questions were presented after the video and participants who failed either question, were not included in the study (Appendix I).

During the study, participants were also asked to complete the three outcome measures; the PDD, a quantified measure of the four *Vowles* criteria on a six-point Likert Scale (Appendix H) and The Mental Health Locus of Origin (MHLO) Scale (Bale and Hill, 1980) (given as part of another trainee's study – see Appendix A). The order of the above measures was randomised to control for ordering and fatigue effects.

Ethical Considerations

Ethical approval was gained from the UEA Faculty of Medicine and Health research ethics committee (Appendix Q). Additional ethical approval was also gained from the Judicial Office (Appendix J) however due to delays resulting from the COVID-19 pandemic, the approval was granted too late to start recruitment of judges given the timeline for this doctoral thesis. The study followed research guidelines and principles of BPS Code of Human Research Ethics and Conduct (2016) and Ethics Guidelines for Internet-mediated Research (BPS, 2017). Participants were recruited online and provided with an information

sheet and a debrief form (Appendix R), which included resources for further support if required. Due to the online nature of the study, participants could withdraw from the study at any time before submission. Given the nature of the vignettes, information was provided regarding how participants could seek additional support after the study if they were distressed by the material. Demographic information was collected, however no identifiable information was required.

Data analysis

Data analysis was conducted using IBM SPSS Statistics v28. Preliminary analyses were run to check if there were any differences in groups or if there was an impact of gender, age and vignette condition on PDD score. For the main analysis, a multinominal logistic regression was proposed due to the three-level categorical dependent variable of sentencing outcome. However, this was adapted to binary logistic regression as prison and s.45A were later merged to form one custodial category due to small numbers of participants choosing prison. The independent or predictor variables were psychiatric diagnosis with three levels (complex mental health problems, schizophrenia and PD), total PDD score (continuous variable) and the *Vowles* criteria Likert-scales (ordinal). Due to more than one hypothesis being tested, the Holm (1979) alpha correction was applied to adjust *p* values. The Holm alpha correction works by ordering the *p* values under .05 and adjusting relative to the number of hypotheses and rank number. To support interpretation, *p* values under .05 are reported at their original level with a notation of the Holm (1979) adjustment.

Results

Sentencing outcome

Participants were split evenly into three conditions; Table 2 shows the proportion of sentencing outcomes in each group. There were no observed differences between vignette conditions. The majority of participants (75.13%) selected s.45A. Due to the low number of

participants choosing the prison sentence (2.54%), prison and s.45A totals were combined to form a broad custodial category for the purpose of analysis.

Table 2

Sentencing Outcome Across the Three Experimentally Manipulated Diagnostic Conditions

Sentence	Vignette Condition			
	Schizophrenia	BPD	Complex Mental Health	Total
37/41	14	16	14	44
45A	55	47	46	148
Prison	0	5	0	5
Total	69	68	60	197

Total PDD Scores

Total stigma score was calculated from the PDD with higher scores indicating stronger perceptions of devaluation-discrimination. (M = 47.67, SD = 10.37, range = 21-71)**.** The mean scaled score lies above the 3.5 average (M = 3.97, SD =0.86, range = 1.75-5.91). The results of the Kolmogorov-Smirnov Test were not significant (D (197) = .052, p> .05) and therefore indicate normality within the data.

Preliminary analysis found moderately positive correlations between age and Total PDD score, however this was not significant (r[195]=.34, p=.64*).* Similarly, gender and Total PDD score were found to have a weak positive correlation, however this again was not significant (r[195]=.11, p=.10) therefore these factors are not included in the following analyses (see Appendix S for further details). There was also a weak positive correlation between stigma and diagnosis, however also not significant, suggesting no relationship between these variables (r[195]=.13, p=.75).

***Vowles* Criteria**

All four *Vowles* criteria were relatively strongly endorsed by participants across all three diagnostic conditions (see Table 3). Preliminary analysis found weak positive

correlations between Total PDD scores and each of the four *Vowles* criteria variables, all these correlations were non-significant (see Appendix S for further detail). This suggests no relationship between stigma (as measured by the PDD) and decision-making on the different *Vowles* criteria.

Table 3

Means And Standard Deviations of Vowles Criteria Likert Scales Ranging From 1-6

Vowles Criteria	Mean (SD) 1= Strong Disagree … 6 = Strongly Agree			
	Complex Mental Health (n = 60)	BPD (n = 68)	Schizophrenia (n = 69)	Total (N = 197)
(1) Extent to which Mental Health Requires Treatment	5.88 (.67)	5.97 (0.17)	5.59 (1.22)	5.18 (0.83)
(2) Extent to which Offence is Attributable to Mental Health	4.97 (.96)	4.81 (1.00)	4.99 (1.01)	4.92 (0.99)
(3) Extent to which Offence Requires Punishment	5.52 (.97)	5.49 (0.86)	5.41 (1.02)	5.46 (0.95)
(4) Extent to which Protection of the Public is Important	5.93 (.25)	5.82 (0.46)	5.62 (1.06)	5.79 (0.70)

Main Analysis

Binominal logistic regression was conducted to understand whether sentencing outcome was predicted by the different dependent variables (i.e. PDD total scores, video condition/diagnosis and *Vowles* Critera). Variables were entered into the model together to measure the impact of each factor whilst also controlling for the other. The continuous predictor variable (Total PDD score) was tested a priori to verify there were no assumptions violated using the Box-Tidwell (1962) test. Tolerance and Variance Inflation Factor (VIF) tests were .94 and 1.06 respectively, indicating no multicollinearity between factors (Tabachnick & Fidell, 2007).

Table 4

Logistic regression comparing likelihood of giving a custodial sentence to each variable

Variable	B	S.E.	Wald	*P*	Odds Ratio
Complex Mental Health			1.21	.55	1.00
BPD	.033	.45	.01	.94	1.03
Schizophrenia	.48	.49	.97	.32	1.62
Total PDD Score (Link, 1989)	-.01	.018	.19	.66	.99
Vowles 1-Treatment	.16	.21	.63	.43	1.18
Vowles 2-Attribution	-.12	.20	.35	.55	.89
Vowles 3-Punishment	.79	.19	16.77	<.01*	2.20
Vowles 4-Public Protection	.16	.24	.44	.51	1.18

**Denotes p values reaching statistical significance following Holm adjustment*

The logistic regression model was not statistically significant χ^2 (8, n = 197) = 3.12, p = .93. The model explained between 11% (Cox and Snell R Square) and 17% (Nagelkerke R^2) of the variance in sentencing and correctly classified 79.2% of cases. However, this level of accuracy was similar to the baseline model at 77.7%, which is likely a result of uneven sentencing groups. Neither diagnosis nor PDD Total Score were predictors of sentencing outcome. Out of the four *Vowles* criteria only the third criteria ("the extent to which the offence requires punishment") was a significant predictor of sentencing outcome. For every unit increase in this *Vowles* Scale, participants were 2.202 times more likely to recommend a custodial sentence. However, given the whole model was not significant, this finding should be interpreted with caution.

Discussion

This study sought to investigate factors which predicted sentencing outcome in an experimentally manipulated video vignette design. Following a review of current literature, it was hypothesised that psychiatric diagnosis would predict sentencing outcome, with a PD diagnosis more likely to attract a custodial sentence (Blais & Forth, 2014). It was also hypothesised that higher stigma, as measured by the PDD, would lead to more custodial sentences. Finally, it sought to examine the relevance of each of the *Vowles* criteria on judicial decision making. It was hypothesised that stronger agreement with the "extent to which the offender's mental health requires treatment" and "the extent to which offending is attributable to the mental health disorder" would decrease the likelihood of a custodial sentence. Conversely, it was also hypothesised that the stronger the agreement with, "the extent to which offending requires punishment" and "the protection of the public when deciding release and regime of release", the higher the likelihood of a custodial sentence. However, in this study, neither psychiatric diagnosis nor PDD score were significant predictors of sentencing outcome. Furthermore, only one *Vowles* criteria, "the extent to which the offence requires punishment", was a significant predictor of sentencing outcome. Preliminary analyses also found there were no statistical associations between PDD scores and age or sex, and no statistical differences in PDD scores between diagnostic groups.

Sentencing Outcome

Overall, most participants (75.13%) opted for a s.45A, regardless of experimental condition. Although not the primary aim of this study, it is nevertheless an interesting finding and indicates a high level of agreement between participants. Responses align with sentencing guidelines which direct judges to consider s.45A before a hospital disposal (Sentencing Council, 2020). However, participants may have been influenced by other heuristics not examined in this study. For example, it has been suggested that when faced

with uncertain decisions, individuals are likely to choose the middle option due to extremeness aversion (Tversky, & Kahneman, 1974; Neumann et al, 2016). However, whilst s.45A could appear to provide a middle-ground sentencing option by combining elements of treatment and punishment, the reality may be quite different. Hospital treatment under s.45A differs considerably from treatment on a s.37, with many arguing the positive effects of hospital treatment are reversed by the subsequent placement of OMHDs in inappropriate prison environments (Jakobowitz et al., 2017; O'Loughlin, 2021; Peay, 2016). Long-term treatment needs of OMHDs are largely unmet when released from prison as they are unable to access NHS aftercare (MOJ, 2017; O'Loughlin, 2021; Peay, 2015). Therefore, this perceived "middle option" can lead to declining mental health and thus increased risk of recidivism (Fazel et al., 2016; Peay, 2016).

Psychiatric Diagnosis

The hypothesis that sentencing outcome would be predicted by a specific psychiatric diagnosis was not supported by the results. This contradicts previous research around diagnostic suitability for s.45A (Beech et al., 2019), beliefs around control and treatability surrounding PD (Blaise &Forth, 2014; Furnhamet al., 2015), and schizophrenia and dangerousness (Levey et al., 1995; Crisp et al., 2000; Varshney et al., 2016). Several factors could explain this. On average, participants in the study reported PDD scores above the midpoint of 3.5 (M = 3.97, SD =0.86) indicating stigmatic views to mental illness. This data is similar to normative data gathered by Link et al (1989) on a community non-psychopathology sample (*M*=4.08, *SD*=0.80) and thus, representative of the general population. However, literature around PD-specific stigma is largely based on research with mental health professionals, who have greater exposure to this population and stigmatic attitudes (Sheehan et al., 2016; Markham & Trower, 2003). Therefore, it is possible that diagnostic specific stigmatic attitudes are less pervasive within a non-mental health

professional population and therefore all mental health labels are given the same weighting. This is supported by previous research which suggests any diagnostic label increases predictions of recidivism (Nee & Witt, 2013). This could suggest that the non-diagnostic label of "complex mental health", also attracts similar levels of stigma. This is important to consider given the drive to move to a non-diagnostic framework of mental health to reduce perceived stigma (e.g., the Power, Threat, Meaning Framework; Johnstone & Boyle, 2018).

Total PDD Scores

The hypothesis that higher total PDD scores would predict more custodial sentences, was not supported by the results. Total PDD scores did not significantly predict sentencing outcome. This contrasts with previous research which indicates high PDD leads to increased negative outcomes for OMHDs (Link et al., 1987;1989; Link & Phelan, 2014; Winnick, & Bodkin, 2008). There are various ways this finding could be understood. Firstly. several factors may have taken precedence over stigmatic attitudes towards mental health problems in the decision-making process, meaning there was little role for mental health stigma to operate. For instance, competing stigmatic attitudes towards the label of offending (Rosenfield, 1997; Winnick, & Bodkin, 2008), bias towards the middle-option (Tversky, & Kahneman, 1974; Neumann et al., 2016), or indeed specific elements of the vignette itself (including the violent nature of the crime). Finally, it is also important to consider the unequal distribution of sentencing options which makes prediction inherently harder (Field, 2015).

***Vowles* Criteria**

The third *Vowles* criteria, "the extent to which the offender requires punishment", was found to be a significant predictor of sentencing outcome within the model, with higher scores indicating a higher likelihood of specifying a sentence with a custodial element. This finding should be interpreted with caution given the overall model was not significant.

However, overall significance of the model examines the average contribution of a set of predictors and, given there was no multicollinearity between variables, it is likely that the model was not significant due to too many non-significant contributors. This could indicate that the need for punishment takes precedence over all other predictive variables (including the other *Vowles* criteria) within the model. Punishment remains an important function of sentencing (Sentencing Guidelines, 2020), despite contemporary penal discourse primarily focusing on the rehabilitation of offenders (MoJ, 2010,2013,2017). However, public opinion favours punishment and, in practice, custodial sentence length continues to rise annually in England and Wales to the highest it has been in over a decade (MoJ, 2021). Several factors included in the vignette have been identified as "aggravating factors" within the Sentencing Guidelines (2020) and may have contributed to stronger agreement on this criterion. These include the seriousness of the offence category (GBH with intent), the harm caused by the offender and use of a weapon (Sentencing Council, 2020).

The remaining three *Vowles* Criteria were not found to be significant predictors. The predictive value of these criteria may have been reduced by the need to merge prison and s.45A into one category for analysis. For example, strong agreement of the *Vowles* treatment criteria was found across the total sample (M=5.18; SD=0.83) with only a few individuals opting for a prison sentence without a treatment component (N=5). However, during analysis both the custodial and non-custodial categories contained an element of treatment because of the merge. Participants reported strong agreement with "the extent to which offending is attributable to the mental health disorder" criterion (M= 4.92; SD=0.99), however Sentencing Guidelines state a s.45A can still be given to an offender even if they meet criteria for a s.37, i.e., regardless of whether the offence is attributable to their mental health (Sentencing Council, 2020). Therefore, perhaps it makes sense that this criterion is less likely to predict the overall outcome. The final criterion concerning public protection when deciding release

and regime of release was also strongly endorsed (*M*=5.79; SD=0.70). Given most participants subsequently opted for a s.45A, it could suggest that the prison parole board and probation system are safer options than mental health services. However, mental health supervision in the community following a hospital disposal has been argued to be the most effective method for public protection where there is a strong link between offending and a deterioration in mental health (Peay, 2015; Fazel et al., 2016).

Strengths and Limitations

This study explores a novel and highly relevant research area relating to the sentencing of OMHDs. The video vignette methodology was designed to reflect real-world sentencing and thus, increase ecological validity (Evans et al., 2015). However, some limitations are present. Firstly, despite aiming to gather a legal sample, only 45.18% of the final sample consisted of qualified legal professionals, with the majority not specialising in criminal law and containing no judges. Judges ultimately have the responsibility of sentencing OMHDs and will arguably be more familiar with current sentencing guidelines and Case Law, despite limited mental health training (Latham, 2017). Furthermore, the study used convenience sampling which may increase bias within the sample (Clark-Carter, 2018). The study was clearly described during recruitment as being interested in the sentencing of OMHDs therefore it may have attracted those with higher interest in mental health research or primed participants to more socially desirable responses. Similar to other social science research, the sample was primarily White British (78.68%) and female (77.2%). Recent reports from the Judicial Office convey a similar lack of ethnic diversity with 92.6% of court judges in England and Wales identifying as White British (MOJ, 2020). However, only 32% of court judges identified as female (MOJ, 2020). Similarly, in this sample, the mean age of all participants was 33.57 years old. However, three-quarters of all judges in England and Wales are over 50 years old, with the senior judges who preside over serious criminal

offences often significantly older than this (MOJ, 2020). This implies the sample included in this study may not be representative of a judicial population who may demonstrate increased cohort effects for stigmatic attitudes towards mental health that were not observed in the current study. Positively, secondary analysis demonstrated that the population sample did not impact on the total PDD score with age, gender and mental health experience not significantly associated with Total PDD score.

With regards to the design of the study, a key limitation was the limited and forced three sentence option. Although this is ecologically valid and judges are forced to choose a sentence at the end of a trial, it provided limited information around the decision-making process and participant's confidence in this choice. This was further exacerbated by the need to combine the two custodial sentences due to small proportion of individuals choosing the prison sentence, thus decreasing the sensitivity of the test. Furthermore, due to limitations of video vignette methodology and a need to balance mental health elements and culpability elements of the vignette, this study could only explore a single crime (GBH) in a particular set of circumstances, with one individual with given specific demographic characteristics. Therefore, it is difficult to generalise the findings to a wider population. Moreover, a single vignette ordered in the same way presents challenges regarding the positioning of information. Initial information is deemed to be the most influential in decision making even if subsequent information is equally important (Tversky, & Kahneman, 1974).

It is also important to note, 88 participants were excluded from the main analysis for failing the knowledge check. This is a common difficulty in internet-mediated research where participants environments are not controlled, and they may be at higher risk of distraction (BPS Guidelines, 2017). This may also suggest participants included in the study might not have been attending to the details of the vignette, which could impact the overall quality of results.

Clinical Implications

The sentencing of OMHDs has real-life implications concerning therapeutic outcomes and the management of risk of harm to the public. Judges make complex decisions with limited mental health training and only vague sentencing guidelines alongside the *Vowles* Criteria to guide their practice. Although positively, neither stigmatic attitudes nor psychiatric diagnosis were found to predict sentencing outcome, this decision was predicted by the need for punishment and most people chose the middle option (s.45A). This is paralleled with the apparent increase in s.45A application in real-world sentencing. This has potentially harmful consequences for OMHDs who experience clinical limitations when on this section in hospital and are then subsequently transferred to prison where their mental health needs cannot be met (O'Loughlin, 2021). If the use of this hybrid order continues to increase, increased funding for mental health provision within the prison estate will be essential.

Recommendations

Ethical approval has been granted for this study to be carried out with the Judicial Office. It will be important to consider the limitations of this study and amend accordingly for the target population. The challenge of having three distinct categories could be alleviated with the addition of confidence scales. It will also be important to consider specific elements of the vignette including information presented and the order it is presented. It will be particularly important to explore the impact of the offence category and its impact on sentencing and the third *Vowles* Criteria around punishment. Furthermore, it may also be important to consider stigmatic attitudes of expert witnesses and how their testimony - particularly when conflicting - influences judicial decision making.

Conclusions

This paper outlines the impact of psychiatric diagnosis, perceived discrimination and devaluation and the *Vowles* Criteria on sentencing outcome. The primary finding of this paper

was the sole predictor of the third *Vowles* Criteria, "the extent to which punishment is required" and sentencing outcome. This study did not identify any additional significant predictors of sentencing outcome in the model including the other three *Vowles* Criteria, Psychiatric Diagnosis or PDD. The results suggest a narrative of punishment, which is also apparent politically as reflected in the updated Sentencing Guidelines for OMHDs (Sentencing Council, 2020; Taylor et al., 2021). This is particularly concerning given the poor outcomes for OMHDs in prison and the potential for harm due to lack mental health provision (O'Loughlin, 2021; Peay, 2016; Jakobowitz et al., 2017). Further research is necessary to identify whether these results are generalisable to a judicial population and explore levels of confidence in this decision-making process. Furthermore, it may also be helpful to explore the impact of offence category and expert witness testimony during sentencing, specifically focusing on stigmatic attitudes of mental health professionals.

References

Aboraya A. (2007). The Reliability of Psychiatric Diagnoses: Point-Our psychiatric Diagnoses are Still Unreliable. *Psychiatry, 4*(1), 22–25.

Ahn, J. H., Kim, W. H., Choi, H. J., Jeon, J. Y., Song, I. G., & Bae, J. N. (2015). Stigma of Mental Illnesses as Perceived by North Korean Defectors Living in South Korea. *Psychiatry Investigation*, *12*(1), 9. https://doi.org/10.4306/PI.2015.12.1.9

Allsopp, K., Read, J., Corcoran, R., Kinderman, P. (2019). Heterogeneity in psychiatric diagnostic classification, *Psychiatry Research, 279*, 15-22 https://doi.org/10.1016/j.psychres.2019.07.005

Baker, J., Edwards, I., & Beazley, P. (2021). Juror decision-making regarding a defendant diagnosed with borderline personality disorder. *Psychiatry, Psychology and Law*, *0*, 1-19. https://doi.org/10.1080/13218719.2021.1938273

Bateman, A. W., Gunderson, J., & Mulder, R. (2015). Treatment of personality disorder. *The Lancet*, *385*(9969), 735-743. https://doi.org/10.1016/S0140-6736(14)61394-5

Beech, V., Marshall, C., Exworthy, T., Peay J. & Blackwood, N. (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

Bentall, R. P. (2003). *Madness explained: Psychosis and human nature*. London: Allen Lane.

Blais, J., & Forth, A. E. (2014). Potential labelling effects: Influence of psychopathy diagnosis, defendant age, and defendant gender on mock jurors' decisions. *Psychology, Crime & Law, 20*(2), 116 134. https://doi.org/10.1080/1068316X.2012.749473

Bradley, K. J. C. B. (2009). *The Bradley Report: Lord Bradley's review of people with mental health problems or learning disabilities in the criminal justice system* (Vol. 7).

London: Department of Health. https://webarchive.nationalarchives.gov.uk/ukgwa/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_098698.pdf

Brekke, J. S., Prindle, C., Bae, S. W., & Long, J. D. (2001). Risks for individuals with schizophrenia who are living in the community. *Psychiatric services (Washington, D.C.)*, *52*(10), 1358–1366. https://doi.org/10.1176/appi.ps.52.10.1358

British Psychological Society (2016). *Code of Human Ethics.* Leicester: British Psychological Society.

British Psychological Society (2017) *Ethics guidelines for Internet-mediated research.* INF206/04.2017. Leicester: British Psychological Society.

Brohan, E., Gauci, D., Sartorius, N., Thornicroft, G., & GAMIAN-Europe Study Group (2011). Self-stigma, empowerment and perceived discrimination among people with bipolar disorder or depression in 13 European countries: the GAMIAN-Europe study. *Journal of affective disorders*, *129*(1-3), 56–63. https://doi.org/10.1016/j.jad.2010.09.001

Brohan, E., Slade, M., Clement, S., & Thornicroft, G. (2010). Experiences of mental illness stigma, prejudice and discrimination: a review of measures. *BMC health services research*, *10*, 80. https://doi.org/10.1186/1472-6963-10-80

Chiswick D. (1996). Sentencing mentally disordered offenders. *BMJ (Clinical research ed.)*, *313*(7071), 1497–1498. https://doi.org/10.1136/bmj.313.7071.1497

Clark-Carter, David. (2018) *Quantitative Psychological Research: The Complete Student's Companion: The Complete Student's Companion*. Psychology press.

Coid, J., Yang, M., Tyrer, P., Roberts, A., & Ullrich, S. (2006) Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, *188*(5), 423-431. https://doi.org/10.1192/bjp.188.5.423

Corrigan, P. W., & Rao, D. (2012). On the self-stigma of mental illness: Stages, disclosure, and strategies for change. *The Canadian Journal of Psychiatry, 57*(8), 464-469. https://doi.org/10.1177/070674371205700804

Corrigan, P. W., Larson, J. E., Watson, A. C., Boyle, M., & Barr, L. (2006). Solutions to discrimination in work and housing identified by people with mental illness. *The Journal of nervous and mental disease*, *194*(9), 716–718. https://doi.org/10.1097/01.nmd.0000235782.18977.de

Corrigan, P. W., Watson, A. C., Warpinski, A. C., & Gracia, G. (2004). Stigmatizing attitudes about mental illness and allocation of resources to mental health services. *Community Mental Health Journal*, *40*(4), 297–307. https://doi.org/10.1023/B:COMH.0000035226.19939.76

Crime (Sentences) Act 1997, Section 46, under the provisions for amending the Mental Health Act 1983 (mentally disordered offenders) and Commencement Order. Number 2, S.I. 1997, No. 2200.

Criminal Justice Act, 2003.

Crisp, A. H., Gelder, M. G., Rix, S., Meltzer, H. I., & Rowlands, O. J. (2000). Stigmatisation of people with mental illnesses. The British journal of psychiatry, *177*(1), 4-7. https://doi.org/10.1192/bjp.177.1.4

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

Department of Health (2008) *Code of Practice. Mental Health Act 1983.* London: The Stationery Office.

Eastman, N. (1996). Hybrid orders: an analysis of their likely effects on sentencing practice and on forensic psychiatric practice and services. *Journal of Forensic Psychiatry*, *7*(3), 481-494. https://doi.org/10.1080/09585189608415032

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., Stough, C. O., Canter, K. S., Robles, R., & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology, 15*(2), 160–170. https://doi.org/10.1016/j.ijchp.2014.12.001

Fazel, S., Fimińska, Z., Cocks, C., & Coid, J. (2016). Patient outcomes following discharge from secure psychiatric hospitals: systematic review and meta-analysis. The British Journal of Psychiatry, *208*(1), 17-25. https://doi.org/10.1192/bjp.bp.114.149997

Field, A. (2015*). Discovering Statistics Using IBM SPSS Statistics – Fifth Edition*. London: Sage.

Furnham, A., Lee, V., & Kolzeev, V. (2015). Mental health literacy and borderline personality disorder (BPD): what do the public "make" of those with BPD? *Social psychiatry and psychiatric epidemiology*, *50*(2), 317–324. https://doi.org/10.1007/s00127-014-0936-7

Goffman, E. (2009). *Stigma: Notes on the management of spoiled identity*. New York: Simon and Schuster.

Gove W. R. (1970). Societal reaction as an explanation of mental illness: an evaluation. *American Sociological Review*, *35*(5), 873–884. https://doi.org/10.2307/2093298

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*(3), 499-510. https://doi.org/10.1207/s15327906mbr2603_7

Gutheil, T. G., Schetky, D. H., & Simon, R. I. (2006). Pejorative testimony about opposing experts and colleagues: "fouling one's own nest". *The journal of the American Academy of Psychiatry and the Law*, *34*(1), 26–30.

HM Chief Inspector of Prisons (2021) *Annual Report 20-21*. London: HMSO. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1003082/hmip-annual-report-accounts-2020-21.pdf

Jakobowitz, S., Bebbington, P., McKenzie, N., Iveson, R., Duffield, G., Kerr, M., & Killaspy, H. (2017). Assessing needs for psychiatric treatment in prisoners: 2. Met and unmet need. *Social psychiatry and psychiatric epidemiology*, *52*(2), 231–240. https://doi.org/10.1007/s00127-016-1313-5

Johnstone, L., & Boyle, M. (2018). The Power Threat Meaning Framework: An Alternative Nondiagnostic Conceptual System. *Journal of Humanistic Psychology.* https://doi.org/10.1177/0022167818793289

Kinderman, P., Read, J., Moncrieff, J., & Bentall, R. P. (2013). Drop the language of disorder. *Evidence-Based Mental Health*, *16*(1), 2–3. http://dx.doi.org/10.1136/eb-2012-100987

Krajewski, C., Burazeri, G., & Brand, H. (2013). Self-stigma, perceived discrimination and empowerment among people with a mental illness in six countries: Pan European stigma study. *Psychiatry research*, *210*(3), 1136–1146. https://doi.org/10.1016/j.psychres.2013.08.013

Latham, D. (2017). *Mental health and fair trial: A Report by Justice*. London: Justice.

Levey, S., & Howells, K. (1995). Dangerousness, unpredictability and the fear of people with schizophrenia. *Journal of Forensic Psychiatry, 6*(1), 19–39. https://doi.org/10.1080/09585189508409874

Lewis, G., & Appleby, L. (1988). Personality disorder: the patients psychiatrists dislike. *British Journal of Psychiatry*, *153*(1), 44-49. https://doi.org/10.1192/bjp.153.1.44

Li, L. and Foster, S. (2005). *Statistics of Mentally Disordered Offenders 2004.* London: Home Office.

Link, B. G., & Cullen, F. T. (1983). Reconsidering the social rejection of ex-mental patients: Levels of attitudinal response. *American Journal of Community Psychology*, *11*(3), 261- 273. https://doi.org/10.1007/BF00893367

Link, B. G., & Phelan, J. (2014). Stigma power. *Social science & medicine*, *103*, 24-32. https://doi.org/10.1016/j.socscimed.2013.07.035

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual review of Sociology*, *27*(1), 363-385. https://doi.org/10.1146/annurev.soc.27.1.363

Link, B. G., Cullen, F. T., Frank, J., & Wozniak, J. F. (1987). The social rejection of former mental patients: Understanding why labels matter. *American Journal of Sociology, 92*(6), 1461–1500. https://doi.org/10.1086/228672

Link, B. G., Cullen, F. T., Struening, E. L., Shrout, P. E., & Dohrenwend, B. P. (1989). A modified labeling theory approach to mental disorders: An empirical assessment. American Sociological Review, *54*(3), 400–423. https://doi.org/10.2307/2095613

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2001). Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses. Psychiatric Services, *52*(12), 1621-1626 https://doi.org/10.1176/appi.ps.52.12.1621

Markham, D., & Trower, P. (2003). The effects of the psychiatric label 'borderline personality disorder' on nursing staff's perceptions and causal attributions for challenging behaviours. *British Journal of Clinical Psychology*, *42*(3), 243-256. https://doi.org/10.1348/01446650360703366

Ministry of Justice (2010) *Breaking the Cycle: Effective Punishment, Rehabilitation and Sentencing of Offenders*. London: HMSO. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/185936/breaking-the-cycle.pdf

Ministry of Justice (2013). *Transforming Rehabilitation: A Strategy for Reform*. London: HMSO.

Ministry of Justice (2017). *Mentally disordered offenders - the restricted patient system – Background Briefing.* London: HMSO.

Ministry of Justice and Department of Health (2011). *Working with personality disordered offenders: a practitioner's gui*de. London: National Offender Management Service and Department of Health.

Ministry of Justice. (2018). *Offender Management Statistics quarterly: October to December 2017.* https://www.gov.uk/government/statistics/offender-management-statistics-quarterly-october-to-december-2017

Ministry of Justice. (2020). *Diversity of the judiciary: Legal professions, new appointments and current post-holders.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/918529/diversity-of-the-judiciary-2020-statistics-web.pdf

Ministry of Justice. (2021). *Criminal Justice Statistics Quarterly June 2021* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1034457/criminal-justice-statistics-june-2021_Q2_revised.pdf

National Institute for Mental Health in England. (2003). *Personality Disorder: No Longer a Diagnosis of Exclusion.* London: Department of Health.

Nee, C., & Witt, C. (2013). Public perceptions of risk in criminality: The effects of mental illness and social disadvantage. *Psychiatry Research*, *209*(3), 675-683. https://doi.org/10.1016/j.psychres.2013.02.013

Neumann, N., Böckenholt, U., & Sinha, A. (2016). A meta-analysis of extremeness aversion. *Journal of Consumer Psychology*, *26*(2), 193-212. https://doi.org/10.1016/j.jcps.2015.05.005

O'Loughlin, A. (2021). Sentencing Mentally Disordered Offenders in England and Wales: Towards a Rights-Based Approach. *Criminal Law Review*, (2), 98-112.

Offences Against the Person Act, 1861.

Olbert, C. M., Gala, G. J., & Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *Journal of abnormal psychology*, *123*(2), 452–462. https://doi.org/10.1037/a0036068

Peay, J. (2016). Responsibility, culpability and the sentencing of mentally disordered offenders: objectives in conflict. *Criminal Law Review (3),* 152-164.

R v Edwards (2018) EWCA Crim 595

R v. Vowles (2015) EWCA Crim 45

Rosenfield, S. (1997). Labelling mental illness: The effects of received services and perceived stigma on life satisfaction. *American Sociological Review, 62*(4), 660–672. https://doi.org/10.2307/2657432

Royal College of Psychiatrists (1996). *Report of the confidential inquiry into homicides and suicides by mentally disordered offenders*. London: RCP.

Scheff, T. (1974). The Labelling Theory of Mental Illness. *American Sociological Review, 39*(3), 444-452. https://doi.org/10.2307/2094300

Scheff, T. J. (1966). *Being mentally ill III: A sociological theory*. Chicago: Aldine

Sheehan, L., Nieweglowski, K., & Corrigan, P. (2016). The Stigma of Personality Disorders. *Current psychiatry reports*, *18*(1), Article 11. https://doi.org/10.1007/s11920-015-0654-1

Singleton, N., Gatward, R., & Meltzer, H. (1998). *Psychiatric morbidity among prisoners in England and Wales.* London: Stationery Office.

Taylor, P., Eastman, N., Latham, R., & Holloway, J. (2021). Sentencing offenders with mental disorders, developmental disorders or neurological impairments: What does the new Sentencing Council Guideline mean for psychiatrists? *The British Journal of Psychiatry*, *218*(6), 299-301. https://doi.org/10.1192/bjp.2021.21

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Varshney, M., Mahapatra, A., Krishnan, V., Gupta, R., & Deb, K. S. (2016). Violence and mental illness: what is the true story?. *Journal of epidemiology and community health*, *70*(3), 223–225. https://doi.org/10.1136/jech-2015-205546

Whyte, S., & Gupta, J. (2007). Widening the scope of the hybrid order. *Medicine, science and the law*, *47*(1), 41-44.

Winnick, T. A., & Bodkin, M. (2008). Anticipated stigma and stigma management among those to be labelled "ex-con". *Deviant Behavior, 29*(4), 295-333. https://doi.org/10.1080/01639620701588081

Chapter Five: Discussion and Critical Evaluation

Overview of chapter

This final chapter summarises and integrates the main findings from both the meta-analysis outlined in Chapter Two, and the empirical research project reported in Chapter Four. Both pieces of work will be critically evaluated, and their strengths and limitations discussed. Considerations for future research will be reported, followed by an exploration of the overall clinical implications. Furthermore, the process of completing the thesis portfolio will also be reflected on. Finally, a conclusion of the whole portfolio is presented.

Main findings

Meta-analysis

The meta-analysis provided an overview of the interrater reliability (IRR) of the Alternative Model of Personality Disorder (AMPD) currently included in Section III of the Diagnostic and Statistical Manual for Mental Disorders – Fifth Edition (DSM-5; APA, 2013). The review was the first of its kind and offered a systematic and comprehensive review of relevant IRR literature of the model.

The review provided tentative support for the IRR of Criterion A of the AMPD, particularly when compared to the current categorical model of the DSM-5 (APA, 2013). Meta-analysis found an overall Intraclass Correlation Coefficient (ICC) of .75 (95% CI .63 – .84), for the total Level of Personality Functioning Scale (LPFS; Bender et al., 2011) score across a total sample size of 902. This is above the acceptability threshold for inclusion in DSM-5 (Kraemer et al., 2012) and higher than previous estimates of IRR of the categorical model (k=.40 to .52; Samuel, 2015). However, there was significant heterogeneity across these studies (I^2 =90.10%). Subgroup analysis highlighted considerable difference between assessment methods; studies which used structured clinical interviews designed for the AMPD produced substantially higher IRR scores (*ICC*=.87, 95% CI [.79, .92]) than those

that did not (*ICC*=.51, 95% CI [.42, .60]). These results support previous research which suggests reliability is improved with the use of semi-structured interviews (Aboraya et al., 2006; Wood et al., 2002). However, heterogeneity within the AMPD instrument group remained significant so results should be interpreted cautiously.

All pooled ICCs of LPFS domains from a total sample of 775 fell above the acceptable range of reliability for DSM-5 (>.40; Kraemer et al., 2012) and in the moderate range of reliability according to ICC reporting guidelines (Koo & Li, 2016). However, heterogeneity remained significant across the domains and again could not be fully explained through sensitivity analysis, therefore it is important to interpret pooled ICCs with caution. Pooled ICCs ranged from .63 (*Empathy*) to .73 (*Intimacy*), possibly highlighting variability in the understanding or application of the separate domains.

Fewer studies reported IRR for Criterion B with a total sample size of 115 from three studies, therefore a meta-analysis was not deemed appropriate. However, of the studies that did report IRRs for personality trait domains, IRR scores varied both within and across studies. They ranged from excellent reliability (Disinhibition, ICC= .92; Somma et al., 2020) to poor reliability (Detachment, ICC= 0.43; Morey, 2019) according to ICC reporting guidelines (Koo & Li, 2016).

Similarly, very few studies reported IRR scores for overall diagnosis within the AMPD, with a total sample of 216 reported from three studies. Therefore, a meta-analysis was not deemed appropriate. Within the studies found by this review, the structured clinical interviews designed for the AMPD continued to report higher IRR statistics in the summarised data (e.g., Structured Clinical Interview for DSM–5-AMPD; Bender et al., 2018), with both the English and Turkish version exceeding acceptable standards of IRR for DSM-5 (Kraemer et al., 2012). However, there is a large amount of methodological variation between studies, therefore comparisons may be unwise. Two studies focus on Borderline PD,

arguably the most familiar of the existing diagnoses, whereas the final study focused on the reliability of having, or not having, any PD. Although the information provided by these studies is useful, the overlap and disagreement within PD diagnosis is often between multiple traits or other diagnoses, therefore additional studies looking at these issues would be beneficial.

Overall, the results from the meta-analytic review provided tentative support for the IRR of Criterion A of the AMPD as a dimensional model of Personality Disorder (PD) diagnosis. Perhaps unsurprisingly, IRR appeared to improve using a structured clinical interview designed for use with the AMPD. There are likely two broad implications here; first, that the use of structured interviews in diagnosis may guard against idiosyncrasies in the practice of assessment, including in the process of making judgements about functional impairments. This is already outlined as good practice in various documents (Mental Health Act Code of Practice [21.6]). The second broad implication relates to the bigger question of why clinicians might choose not to use such measures, and it is therefore recommended that further research focuses on using these measures in ecologically valid settings to support clinical utility and ease of use for clinicians. Finally, additional research of clinician administered Criteria B instruments is warranted, alongside IRR of overall diagnostic category where very few existing studies were identified.

Empirical Project

Following on from the meta-analytic review, the empirical project sought to investigate the sentencing of offenders with mental health difficulties (OMHDs). This online study employed an experimental video-vignette methodology to explore the impact of psychiatric diagnosis, perceived mental illness stigma and the criteria laid out in *R v Vowles* which provide guidance on the sentencing of OMHDs. In Clinical Psychology, structural stigma theory proposes that mental illness stigma can result in discrimination (Corrigan et al.,

2004). Furthermore, different psychiatric diagnoses have been found to be more stigmatising than others (Baker at al., 2021; Sheehan et al., 2016). Participants acted as proxy judges and were allocated to one of three experimentally-manipulated diagnostic conditions (Schizophrenia, Borderline PD and Complex Mental Health). They were asked to take the role of sentencing a hypothetical offender who had been found guilty of Grievous Bodily Harm with Intent (GBH; Offences against the Person Act 1861, s.18), an offence punishable by a custodial sentence.

Overall, 75.13% of participants sentenced the offender to the hybrid Section 45A of the Mental Health Act (MHA; 1983), with the remainder being sentenced to a hospital disposal (s.37 MHA) or prison. This finding alone is interesting, since it potentially suggests that s.45A may be seen as a "middle option" between a custodial sentence and hospital order that appears to achieve the conflicting objectives of both sentencing options. However, this apparent simplicity may belie significant underlying complexity in the practical implications of the different options. Nonetheless, the primary aim of the study was to explore the hypotheses that sentencing outcome would be significantly predicted by psychiatric diagnosis, stigmatic attitudes (as measured by the Perceived Devaluation and Discrimination Scale; Link et al., 1987) and the separate *Vowles* Criteria. However, the overall binary logistic regression model was not significant.

In terms of contextualising this finding, it is noted that the sample was found to hold above average stigmatic attitudes towards mental illness as measured by the PDD ($M = 3.97$, $SD = 0.86$, range = 1.75-5.91), although positively these did not act as a predictor for sentencing. Furthermore, psychiatric diagnosis also did not significantly predict sentencing outcome and, in fact, only one *Vowles* Criteria was found to be a significant predictor within the model, "the extent to which a participant agreed the offender required punishment for the

offence". However, as the overall model was not significant [χ^2 (8, n = 197) = 3.12, p = .93], this finding should be interpreted tentatively.

Strengths and Limitations of the Thesis Portfolio

Taken together, the two papers inform one another; with one examining the reliability of psychiatric diagnoses (namely PD) and the other seeking to examine the real-world impact of these diagnoses on the under-researched area of the sentencing of OMHDs. The empirical project further adds to this research by examining the consistency of sentencing outcomes between participants alongside the examination of the psychological construct of mental illness stigma on these sentencing outcomes.

Meta-analysis

A key strength of this overall thesis portfolio is its novel contribution to two under-researched areas within mental health. PD in general is an under-researched area in comparison to other forms of psychopathology (Winship & Hardy, 2007). Criticisms of existing PD research include the lack of an agreed upon definition or assessment method (Tyrer et al., 2007). Therefore, reliable diagnosis to inform future research and evidence-based treatment is essential. Limited meta-analyses have been conducted for the IRR of psychiatric diagnoses and this is the first of its kind with regards to the novel dimensional framework, the AMPD. Meta-analyses of dimensional IRR statistics are even more rare within psychiatric diagnosis due to the dominance of categorical models. However, given the move towards a more dimensional framework for diagnoses within the DSM (e.g., Kotov et al., 2017), this will be an important methodology to summarise future reliability studies. This paper references existing meta-analyses from physical health settings to inform the methodological and analytical process which can hopefully act as a framework for future meta-analytic studies. The critical evaluation of studies was supported by the use of a quality assessment framework to assess risk of bias (Quality Appraisal of Reliability Studies; Lucas

et al., 2010). These tools are often not reported in published meta-analyses, however considered a vital part of the reviewing process (Cuijpers, 2016). This evaluation highlighted variability in IRR reporting across studies which unfortunately led to the exclusion of several studies due to appropriate statistics not being reported despite IRR reporting guidelines recommending doing so (Koo & Li, 2016). It was disappointing that these studies could not be included in the analysis, however recommendations for future studies were made because of the observed methodological weakness, with the aim of developing more robust studies in this area. Risk of bias assessment also enabled studies with high risk of bias to be removed from analyses through sensitivity analyses, which further highlighted improved IRR scores (e.g., Total LPFS pooled ICC of 0.80). Furthermore, subgroup analysis of assessment method for diagnosis was useful to inform barriers and facilitators to improved IRR and led to clinically useful information around the use of structured clinical interviews. However, high levels of heterogeneity were still present across this group. This is often the case in psychological research (Cuijpers, 2016) due to the variability of measures used and methodological differences, however this still presents a limitation of the generalizability of the presented findings.

Empirical Project

Similarly, the empirical project also provides an entirely novel contribution to the limited evidence base regarding the sentencing of OMHDs. This is an under-researched area with very few studies even reporting the prevalence of mental health difficulties in courts (Bradley, 2009). Generally, the few studies which have been conducted around OMHDs, focus on jury decision making as opposed to judicial sentencing (e.g., Baker et al., 2021). Furthermore, to the author's knowledge, the empirical project is the first of its kind to examine decision making involving the hybrid treatment and penal order, s.45A (MHA, 1983). The study responds to calls from legal commentaries (Peay, 2015) for research on the

hybrid order and builds on previous qualitative work completed by Beech et al. (2019) exploring the concerns of psychiatrists around recommending and working clinically with s.45A. The current project offers additional insights into this decision-making process by examining the impact of mental-illness stigma and psychiatric diagnoses, which have been hypothesised to evoke varying levels of stigma from individuals (Baker at al., 2021; Link & Phelan, 2001; Sheehan et al., 2016). Although the sample was found to hold stigmatic attitudes towards mental illness, this was not found to be a significant predictor of sentencing outcome within this non-judicial population. Several explanations of this are presented within the empirical project, however it is also important to consider the methodological limitations presented by three quarters of the sample choosing the same sentencing option. This inevitably made it harder to statistically determine which factors influenced the choice. However, as stigmatic attitudes were identified in this proxy sample, it is possible to hypothesise they would be present in other populations involved in the legal proceedings, including judges and clinical expert witnesses. This hypothesis is strengthened by research around diagnostic stigma which has largely been carried out using a mental health professional sample (Sheehan et al., 2016; Markham & Trower, 2003). As a result, it would be interesting to see how these potentially stigmatic attitudes influence evidence provided by clinicians when asked to comment on the culpability of OMHDs and the treatability of different diagnoses. Replication would be important to further explore this.

Sample

The decision to collaborate with another Trainee Clinical Psychologist for recruitment facilitated a larger overall sample for the Empirical Project. This was a key strength of this thesis and enabled an adequate sample to be gained for both projects. Recruitment was also improved by the online nature of the study, particularly given the COVID-19 pandemic and barriers to recruiting participants in person. The sentencing of OMHDs is the sole

responsibility of the Judicial Office, therefore this was the primary target population for the Empirical Project. Recruitment of Judges in England and Wales takes place through the official research request pathway, where ethical approval from the Judicial Office must first be granted. This process can be lengthy, and rejections are common. Furthermore, there were significant delays in gaining this approval during the COVID-19 pandemic. As a result, alternative recruitment strategies were also planned and implemented whilst awaiting this approval. Firstly, this involved recruiting legal professionals through law societies and social media. However, low recruitment rates from criminal law practitioners were expected due to well-publicised unmanageable workloads and limited free time (The Bar Council, 2022). Therefore, recruitment also took place using a paid online academic research participant platform who recruited participants from their existing pool of members if they had previously indicated they worked in the legal system. However, despite efforts to recruit a legal population, the Judicial Office ethical approval was received too late for recruitment given the time restrictions of a doctoral thesis and only 44.95% of the final sample consisted of qualified legal professionals. Out of this total, very few participants specialised in Criminal Law, and this did not include any members of the judiciary, who are solely responsible for sentencing. Although this therefore not a "gold standard" sample (as might have been obtained, for instance, through the "unused jury" methodology adopted by Cheryl Thomas [2010]), given the novel nature of this research it provides adequate feasibility for replication of the study with a judicial population.

Despite this, efforts were made to increase the ecological validity of the study. This included using video vignettes over written vignettes, as per recommendations of previous court research (Thomas, 2010). The vignette scripts were developed using elements of real cases involving violent offences and OMHDs. This information was obtained from the Court of Appeal records located on the Westlaw UK database. The vignettes (and additional study

materials) were developed jointly with another Trainee Clinical Psychologist with consultation from a member of the Law School (the secondary supervisor for this project). The aim of this method was to increase ecological validity without using a specific case example that some participants might recognise. It was hoped this would increase the generalisability of subsequent findings and provide additional support for the replication of these findings within the Judicial Office. However, despite attempts to increase ecological validity, there are of course limitations to the vignette methodology, with many arguing it can never replicate real-world phenomena (Evans et al., 2015). For example, the video only depicted a brief sentencing summary as opposed to a full trial. Although this was helpful to decrease the time taken to complete the study and increase study participation, it also simplifies a complex decision-making process. Previous studies focusing on the decision making of jurors have often been criticised for similarly omitting a trial procedure and jury discussion (e.g., Berryessa et al., 2015; Mossière & Maeder, 2016).

It is also important to note, 88 participants (just under half of the final sample) were excluded from the main analysis for failing the knowledge check. This is in line with previous research (e.g., Mossière & Maeder, 2015) and is a common difficulty in internet-mediated research where participants environments are not controlled, and they may be at higher risk of distraction (BPS Guidelines, 2017). Nevertheless, with such a high proportion of participants excluded, it is possible this may have impacted the overall results or excluded specific cohorts making the final sample less representative of the target population. Furthermore, it may also be important to consider the impact of attention and distraction on participants included in the final sample, which could also impact the overall quality of results.

Similar limitations were found within studies included in the meta-analysis. Not all studies assessed PD using a clinical sample and very few used qualified clinicians. In some

ways, the variation in sampling could be considered a strength, with the dimensional model by nature being applicable to both a clinical and non-clinical sample. However, generalisability of the findings may be impacted by the absence of this. For example, certain personality traits or settings (e.g., forensic) may act as barriers to the assessment process. Similarly, the application of a structured clinical interview, or indeed the absence of this tool, can vary depending on clinical skill and service pressures. Using lay raters may demonstrate the learnability of the AMPD, however arguably qualified clinicians may have a different understanding of concepts given their prior knowledge of the categorical model and training in various theoretical models. Taken together this highlights the benefits of proxy sample, however the need for this research to be replicated in more ecologically valid ways.

Diversity

One of the major limitations between both papers is the lack of diversity presented in the overall samples. Most studies included in the review element of this thesis did not even report ethnicities included in their sample and the sampling method excluded those studies not published in English. Furthermore, in the empirical project, the final sample consisted primarily of White British (78.68%) and female (77.2%) participants. The lack of diversity in research has been widely reported and psychological research has been criticised for reporting a skewed view of humanity by basing research largely on western culture (Brady et al., 2018). In terms of diagnosis, cultural norms will vary in what is considered the "norm" and therefore, what is the "disorder". Therefore, it is vital that research in these areas reflect these populations so Western views of these constructs are not forced upon cultures whom they do not apply to.

In the empirical project, the lack of ethnic diversity in the sample also reflects real-life problems around lack of ethnic diversity in the Judicial Office. This is somewhat problematic considering the population they preside over. Minoritized ethnic groups are over-represented

at every stage of the Criminal Justice System (MOJ, 2021). Furthermore, they are also more likely to be mental health service users in the NHS in England (NHS Digital, 2021). The interaction of these different characteristics is increasingly being considered in public health inequality research in the form of Intersectionality (Howard & Renfrow, 2014). This thesis only considered one aspect of this bias exploring stigmatic attitudes towards mental illness, however others are likely to be present in the courtroom. In some ways, this thesis missed an opportunity to empower these voices and sought consultation from legal professionals without the addition of service users and their families whom these decisions impact. Future research should be co-produced with OMHDs and those service users with experience of receiving a PD diagnosis.

Clinical Implications

Taken together, both papers contribute novel findings to areas of mental health research which both have considerable clinical implications. The AMPD was included in the "Emerging Measures and Models" as it was hoped this would encourage future research into the clinical utility and reliability of the model. The current meta-analysis aimed to summarise the existing literature around IRR of the model to better inform whether it should be included as the primary model of PD diagnosis in the future. Of course, it is acknowledged that other measures of reliability, validity and clinical utility will also need to be assessed before the model is included in future revisions of the DSM. However, this review tentatively supports its inclusion based on IRR improvements alone.

Of course, the review assumes a specific perspective on the ongoing existence of diagnostic concepts and approaches. It is acknowledged that some perspectives within Clinical Psychology, e.g., the Power, Threat, Meaning Framework (PTMF; Johnstone & Boyle, 2018), have argued that this approach should be abandoned entirely. However, it is also recognised that this view does not fully represent mainstream opinion, and whilst work

such as the PTMF has usefully highlighted the broader social context in which mental health and personality problems occur, diagnostic systems such as the DSM-5 and the ICD-11 continue to exert significant influence in psychiatric practice. If they are to continue to be used, efforts to consider how and under what situations they might be most reliably applied remain helpful. The AMPD seeks to provide more clinically useful descriptions of PD that will better inform treatment decisions and future research into PD, therefore remains a useful framework to review.

This review also provided useful information regarding the factors which impact reliability both in research and in clinical practice. For example, the subgroup analysis within this review tentatively highlighted that IRR of the AMPD could be improved with the use of a structured clinical interview designed specifically for the AMPD. This is a useful reminder for clinicians who have tendency to overestimate their abilities of assessment using clinical judgment alone (Monahan, 1981; Kitamura & Kitamura, 2000).

This message was echoed in the empirical project which examined the decision-making process for the sentencing of OMHDs. This process has real-life implications concerning therapeutic outcomes and management of risk for OMHDs. Judges do not have the benefit of using structured clinical assessments to guide their decision-making process and instead, have vague sentencing guidelines alongside the *Vowles* Criteria to guide their practice. Although positively, neither stigmatic attitudes nor psychiatric diagnosis were found to predict sentencing outcome, this decision was predicted by the need for punishment and most people chose the middle option (s.45A). This is paralleled with the apparent increase in s.45A application in real-world sentencing. As discussed in the empirical project, this has potentially harmful consequences for OMHDs who experience clinical limitations when on this section in hospital and are then subsequently transferred to prison where their mental health needs cannot be met (O'Loughlin, 2021). If the use of this hybrid order continues to

increase, increased funding for mental health provision within the prison estate will be essential.

Future Research

Future research examining the IRR of the AMPD should be focused on using structured clinical interviews designed specifically for the AMPD e.g., the SCID-5 (Bender et al., 2018). It would be beneficial for this research to focus on Criterion B of the framework and overall PD diagnosis which have very few existing studies. Research concerning overall PD should explore all available categories of PD and not just the familiar Borderline PD and also explore different cultural contexts. Furthermore, it may be helpful to repeat studies in more ecological valid settings, i.e., with clinicians in clinical practice. This may seek to address criticisms of inflated IRR scores in research settings (Aboraya, 2007) and demonstrate the learnability and usability of structured clinical interviews associated with the framework. Furthermore, future research should also explore similar research questions using the ICD-11 dimensional PD framework given this is the most commonly used diagnostic framework clinically in the United Kingdom and limited evidence is currently available (Mulder, 2021; Mulder & Tyrer, 2019).

As previously mentioned, more ecologically valid methods for the empirical project would also be beneficial. The project has already gained approval for replication with the Judiciary Office. Given the findings of the current project it may also be helpful to explore the impact of stigma in different ways. For example, examining the impact of mental illness stigma on expert witness testimonies with regards to culpability and treatability of certain diagnoses.

Reflections

The process of completing a doctoral thesis was a challenging but rewarding experience. It has not only broadened my knowledge of two clinically relevant subject areas,

but also enabled me to develop as both a facilitator and critical consumer of research. The meta-analysis provided me with an opportunity to critically evaluate the failings of the categorical model of PD diagnosis alongside the breadth of work that contributed to the development of the new dimensional model. Within Clinical Psychology, mental health difficulties are widely viewed as existing on a continua alongside "normal" functioning and understood in the context of an individual's experiences using formulations. However, many service users report positive experiences of being given a PD diagnosis, particularly commenting on feeling understood and being able to connect to others with the same diagnosis (Lester et al., 2020). Therefore, it is positive to see the move to a dimensional model of psychiatric diagnosis which seeks to bridge these two perspectives, particularly within PD. It is hoped this will go some way to reduce the harmful effects of the previous categorical system, including the issue of receiving multiple diagnoses. This is also in line with developments included in the current International Classification of Diseases and Related Health Problems (11th ed.; ICD-11; World Health Organisation, 2018), which has caused some to similarly request further research into structured assessments for use with this dimensional framework for PD diagnosis (e.g., Bach et al., 2022). As a trainee clinician about to embark on a career within the NHS, it was disheartening to learn that the main barriers to using these standardised assessments in clinical practice were underfunding, service pressures and lack of resources.

Interestingly, the possible move to a dimensional approach to diagnosis was contrasted quite starkly to the concrete categories of sentencing available within the context of the legal system and the application of the MHA (1983). It was shocking that, to this date, little research had been conducted into this area of mental health, despite it having such a large impact on an individual's life. Recruitment of legal professionals proved challenging despite adverts in multiple national legal networks. Unfortunately, this likely reflects the

limited resource and pressure around the legal system, which has undoubtedly been made worse by delays and backlogs of criminal cases caused by the COVID-19 pandemic. These delays also impacted the time it took to receive ethical approval from the Judiciary Office. This was a long, but useful process, that will be important to utilise for future projects. It will be important for future studies to be commissioned jointly with those within the system that hold power (e.g., the Judiciary Office) to ensure improved participation and ecological validity in such an important, yet under researched, area.

During the analysis, I was pleasantly surprised to find mental illness stigma did not predict sentencing outcome in the empirical project, however I was surprised by the emphasis on punishment. In some ways, this highlights the competing demands of sentencing held by judges for both punishment and rehabilitation. However, as a clinician with a duty of care to service users, this sits uncomfortably. Separate to causing distress, research has highlighted the ineffectiveness of prison on recidivism rates (Listwan et al., 2012). Moreover, the argument that prison acts as a deterrent has only been supported for white collar crime (Dölling, et al., 2009). This is particularly concerning when considering these consequences primarily impact some of the most disempowered and disadvantaged groups within society. This highlights the importance for Clinical Psychology as a profession to become more involved within Mental Health Law and the wider court system. This includes both contributing to Mental Health Act reform and acting as expert witnesses who can speak to the reliability and inexact science of mental health assessments. Furthermore, it similarly raises important question regarding the underfunding of mental health provision within the prison estate. One could argue, if the treatment of offenders within the prison estate improved with improved resources, facilities and care, the harmful effects of custodial sentences (hybrid or otherwise) would also reduce.

Overall Conclusion

The meta-analysis and empirical research project presented in this thesis portfolio provide novel contributions to the literature regarding psychiatric diagnosis, stigma, and the sentencing of OMHDs. The findings highlight improved IRR of Criterion A of the AMPD when compared with the current categorical model. However, the review identified a need for further research around clinician administered assessment of Criterion B and overall PD diagnosis. It would also be useful for current IRR research to be replicated in clinical settings using structured clinical interviews designed for the AMPD. Furthermore, the empirical project reported useful findings around the sentencing of OMHDs. Stigmatic attitudes and psychiatric diagnoses were not found to predict sentencing outcomes and only one of the *Vowles* Criteria, "the extent the offender requires punishment", was found to predict sentencing outcome. However, limitations of the study were identified and discussed. Replication with a judicial population and a further focus on expert witness testimony is advised. Finally, it is recommended that any future research has significant input from service users impacted by these sentencing and diagnostic processes.

Thesis Portfolio References

Aboraya A. (2007). The Reliability of Psychiatric Diagnoses: Point-Our psychiatric Diagnoses are Still Unreliable. *Psychiatry, 4*(1), 22–25.

Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The Reliability of Psychiatric Diagnosis Revisited: The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis. *Psychiatry (Edgmont)*, *3*(1), 41-50.

Ahn, J. H., Kim, W. H., Choi, H. J., Jeon, J. Y., Song, I. G., & Bae, J. N. (2015). Stigma of Mental Illnesses as Perceived by North Korean Defectors Living in South Korea. *Psychiatry Investigation*, *12*(1), 9. https://doi.org/10.4306/PI.2015.12.1.9

Allsopp, K., Read, J., Corcoran, R., Kinderman, P. (2019). Heterogeneity in psychiatric diagnostic classification, *Psychiatry Research, 279*, 15-22 https://doi.org/10.1016/j.psychres.2019.07.005

American Psychiatric Association. (1952). *Diagnostic and Statistical Manual: Mental Disorders.* Washington, DC.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. (4th ed., revised). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. (5th ed.) Arlington (VA): American Psychiatric Association. https://doi.org/10.1176/APPI.BOOKS.9780890425596

Aviram, R. B., Brodsky, B. S., & Stanley, B. (2006). Borderline personality disorder, stigma, and treatment implications. *Harvard Review of Psychiatry*, *14*(5), 249–256. https://doi.org/10.1080/10673220600975121

Bach, B., Kramer, U., Doering, S., di Giacomo, E., Hutsebaut, J., Kaera, A., de Panfilis, C., Schmahl, C., Swales, M., Taubner, S., & Renneberg, B. (2022). The ICD-11 classification of personality disorders: a European perspective on challenges and opportunities. *Borderline Personality Disorder and Emotion Dysregulation 2022 9:1*, *9*(1), 1–11. https://doi.org/10.1186/S40479-022-00182-0

Baker, J., Edwards, I., & Beazley, P. (2021). Juror decision-making regarding a defendant diagnosed with borderline personality disorder. *Psychiatry, Psychology and Law*, *0*, 1-19. https://doi.org/10.1080/13218719.2021.1938273

Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are Informant Reports of Personality More Internally Consistent Than Self Reports of Personality? *Assessment*, *22*(4), 399. https://doi.org/10.1177/1073191114556100

Beech, V., Marshall, C., Exworthy, T., Peay J. & Blackwood, N. (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

Bender, D. S., Morey, L. C., & Skodol, A. E. (2011). Toward a Model for Assessing Level of Personality Functioning in DSM–5, Part I: A Review of Theory and Methods. *Journal of Personality Assessment*, *93*(4), 332–346. https://doi.org/10.1080/00223891.2011.583808

Bender, D. S., Skodol, A. E., First, M. B., & Oldham, J. M. (2018). Module I: Structured Clinical Interview for the Level of Personality Functioning Scale. In M. B. First, A. E. Skodol, D. S. Bender, & J. M. Oldham (Eds.), Structured Clinical Interview for the DSM-5 Alternative Model for Personality Disorders (SCID-5-AMPD). *American Psychiatric Association.*

Bentall, R. P. (2003). *Madness explained: Psychosis and human nature*. London: Allen Lane.

Berryessa, C. M., Milner, L. C., Garrison, N. A., & Cho, M. K. (2015). Impact of psychiatric information on potential jurors in evaluating high-functioning autism spectrum disorder (hfASD). *Journal of Mental Health Research in Intellectual Disabilities*, *8*(3-4), 140-167. https://doi.org/10.1080/19315864.2015.1040176

Blais, J., & Forth, A. E. (2014). Potential labelling effects: Influence of psychopathy diagnosis, defendant age, and defendant gender on mock jurors' decisions. *Psychology, Crime & Law, 20*(2), 116 134. https://doi.org/10.1080/1068316X.2012.749473

Borenstein, M., Hedges, L. v., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. *Introduction to Meta-Analysis*, 1–421. https://doi.org/10.1002/9780470743386

Bradley, K. J. C. B. (2009). *The Bradley Report: Lord Bradley's review of people with mental health problems or learning disabilities in the criminal justice system* (Vol. 7). London: Department of Health. Retrieved from: https://webarchive.nationalarchives.gov.uk/ukgwa/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_098698.pdf

Brady, L. M., Fryberg, S. A., & Shoda, Y. (2018). Expanding the interpretive power of psychological science by attending to culture. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11406–11413. https://doi.org/10.1073/PNAS.1803526115/ASSET/1A789B81-6F28-4EC5-9DA3-54D63C34EA9E/ASSETS/GRAPHIC/PNAS.1803526115FIG01.JPEG

Brekke, J. S., Prindle, C., Bae, S. W., & Long, J. D. (2001). Risks for individuals with schizophrenia who are living in the community. *Psychiatric services (Washington, D.C.)*, *52*(10), 1358–1366. https://doi.org/10.1176/appi.ps.52.10.1358

Boland, A., Cherry, G., & Dickson, R. (Eds.). (2017). *Doing a systematic review: A student's guide.* London: Sage.

British Psychological Society (2016). *Code of Human Ethics.* Leicester: British Psychological Society.

British Psychological Society (2017) *Ethics guidelines for Internet-mediated research.* INF206/04.2017. Leicester: British Psychological Society.

British Psychological Society. (2013). *Classification of behaviour and experience in relation to functional psychiatric diagnoses: Time for a Paradigm Shift*. DCP Position Statement. Leicester. https://www.bps.org.uk/sites/www.bps.org.uk/files/Member%20Networks/Divisions/DCP/Classification%20of%20behaviour%20and%20experience%20in%20relation%20to%20functional%20psychiatric%20diagnoses.pdf

Brohan, E., Gauci, D., Sartorius, N., Thornicroft, G., & GAMIAN-Europe Study Group (2011). Self-stigma, empowerment and perceived discrimination among people with bipolar disorder or depression in 13 European countries: the GAMIAN-Europe study. *Journal of affective disorders*, *129*(1-3), 56–63. https://doi.org/10.1016/j.jad.2010.09.001

Brohan, E., Slade, M., Clement, S., & Thornicroft, G. (2010). Experiences of mental illness stigma, prejudice and discrimination: a review of measures. *BMC health services research*, *10*, 80. https://doi.org/10.1186/1472-6963-10-80

Buer Christensen, T., Paap, M. C. S., Arnesen, M., Koritzinsky, K., Nysaeter, T.-E., Eikenaes, I., Germans Selvik, S., Walther, K., Torgersen, S., Bender, D. S., Skodol, A. E., Kvarstein, E., Pedersen, G., & Hummelen, B. (2018). Interrater reliability of the Structured Clinical Interview for the DSM–5 Alternative Model of Personality

Disorders Module I: Level of Personality Functioning Scale. *Journal of Personality Assessment*, *100*(6), 630–641. https://doi.org/10.1080/00223891.2018.1483377

Caldwell-Harris, C. L., & Ayçiçegi, A. (2006). When Personality and Culture Clash: The Psychological Distress of Allocentrics in an Individualist Culture and Idiocentrics in a Collectivist Culture. *Collectivism (Particularly Vertical),* 43(3), 331–361. https://doi.org/10.1177/1363461506066982

Carlson, E. N., & Oltmanns, T. F. (2015). The Role of Metaperception in Personality Disorders: Do People with Personality Problems Know How Others Experience Their Personality? *Journal of Personality Disorders*, *29*(4), 449–467. https://doi.org/10.1521/PEDI.2015.29.4.449

Chinese Psychiatry Association. (2000). *Chinese classification of mental disorders. 3rd ed*. Shandong Science Press, Jinan.

Chiswick D. (1996). Sentencing mentally disordered offenders. *BMJ (Clinical research ed.)*, *313*(7071), 1497–1498. https://doi.org/10.1136/bmj.313.7071.1497

Clark-Carter, David. (2018) *Quantitative Psychological Research: The Complete Student's Companion: The Complete Student's Companion*. Psychology press.

Clark, L. A., Vanderbleek, E. N., Shapiro, J. L., Nuzum, H., Allen, X., Daly, E., Kingsbury, T. J., Oiler, M., & Ro, E. (2015). The Brave New World of Personality Disorder-Trait Specified: Effects of Additional Definitions on Coverage, Prevalence, and Comorbidity: *Https://Doi.Org/10.5127/Pr.036314*, *a2*(1), 52–82. https://doi.org/10.5127/PR.036314

Clark, L.A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual review of psychology, 58,* 227–257.https://doi.org/10.1146/annurev.psych.57.102904.190200

Clarkin, J. F., & Huprich, S. K. (2011). Do DSM-5 Personality Disorder Proposals Meet Criteria for Clinical Utility? Journal of Personality Disorders. 25(2), 192–205. https://doi.org/10.1521/PEDI.2011.25.2.192

Clarkin, J. F., Caligor, E., Stern, B., & Kernberg, O. F. (2004). Structured interview of personality organization (STIPO). Personality Disorders Institute. *Weill Medical College of Cornell University.*

Cochran, W. (1954). The Combination of Estimates from Different Experiments. Biometrics, 10(1), 101-129. https://doi:10.2307/3001666

Coid, J., Yang, M., Tyrer, P., Roberts, A., & Ullrich, S. (2006) Prevalence and correlates of personality disorder in Great Britain. *British Journal of Psychiatry*, *188*(5), 423-431. https://doi.org/10.1192/bjp.188.5.423

Corrigan, P. W., & Rao, D. (2012). On the self-stigma of mental illness: Stages, disclosure, and strategies for change. *The Canadian Journal of Psychiatry, 57*(8), 464-469. https://doi.org/10.1177/070674371205700804

Corrigan, P. W., Larson, J. E., Watson, A. C., Boyle, M., & Barr, L. (2006). Solutions to discrimination in work and housing identified by people with mental illness. *The Journal of nervous and mental disease*, *194*(9), 716–718. https://doi.org/10.1097/01.nmd.0000235782.18977.de

Corrigan, P. W., Watson, A. C., Warpinski, A. C., & Gracia, G. (2004). Stigmatizing attitudes about mental illness and allocation of resources to mental health services. *Community Mental Health Journal*, *40*(4), 297–307. https://doi.org/10.1023/B:COMH.0000035226.19939.76

Costa, P. T., Jr., & Widiger, T. A. (Eds.). (1994). *Personality disorders and the five-factor model of personality.* American Psychological Association. https://doi.org/10.1037/10140-000

Crawford, M. J., Koldobsky, N., Mulder, R., & Tyrer, P. (2011). Classifying Personality Disorder According to Severity. *Journal of Personality Disorders. 25*(3), 321–330. https://doi.org/10.1521/PEDI.2011.25.3.321

Crime (Sentences) Act 1997, Section 46, under the provisions for amending the Mental Health Act 1983 (mentally disordered offenders) and Commencement Order. Number 2, S.I. 1997, No. 2200.

Criminal Justice Act. (2003). (s.142). London: Stationery Office.

Crisp, A. H., Gelder, M. G., Rix, S., Meltzer, H. I., & Rowlands, O. J. (2000). Stigmatisation

Cruitt, P. J., Boudreaux, M. J., King, H. R., Oltmanns, J. R., & Oltmanns, T. F. (2019). Examining criterion a: DSM–5 level of personality functioning as assessed through life story interviews. *Personality Disorders: Theory, Research, and Treatment*, *10*(3), 224–234. https://doi.org/10.1037/per0000321

Cuchna, J. W., Hoch, M. C., & Hoch, J. M. (2016). The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. *Physical Therapy in Sport*, *19*, 57–65. https://doi.org/10.1016/J.PTSP.2015.12.002

Cuijpers, P. (2016). *Meta-analyses in mental health research: A practical guide*. Amsterdam, Netherlands, Arije universiteit Amsterdam.

Dale, O., Sethi, F., Stanton, C., Evans, S., Barnicot, K., Sedgwick, R., Goldsack, S., Doran, M., Shoolbred, L., Samele, C., Urquia, N., Haigh, R., & Moran, P. (2017). Personality disorder services in England: findings from a national survey. *BJPsych Bulletin*, *41*(5), 247. https://doi.org/10.1192/PB.BP.116.055251

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

Department of Health (2008) *Code of Practice. Mental Health Act 1983.* London: The Stationery Office.

Dereboy, F., Dereboy, C., & Eskin, M. (2018). Validation of the DSM-5 Alternative Model Personality Disorder Diagnoses in Turkey, Part 1: LEAD Validity and Reliability of the Personality Functioning Ratings. *Journal of Personality Assessment*, *100*(6), 603–611. https://doi.org/https://dx.doi.org/10.1080/00223891.2018.1423989

Dölling, D., Entorf, H., Hermann, D., & Rupp, T. (2009). Is deterrence effective? results of a meta-analysis of punishment. *European Journal on Criminal Policy and Research*, *15*(1–2), 201–224. https://doi.org/10.1007/S10610-008-9097-0/TABLES/12

Duval S, Tweedie R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 56(2):455-463.

Eastman, N. (1996). Hybrid orders: an analysis of their likely effects on sentencing practice and on forensic psychiatric practice and services. *Journal of Forensic Psychiatry*, *7*(3), 481-494. https://doi.org/10.1080/09585189608415032

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., Stough, C. O., Canter, K. S., Robles, R., & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology, 15*(2), 160–170. https://doi.org/10.1016/j.ijchp.2014.12.001

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult)*. London: Hodder and Stoughton.

Fazel, S., Fimińska, Z., Cocks, C., & Coid, J. (2016). Patient outcomes following discharge from secure psychiatric hospitals: systematic review and meta-analysis. The British Journal of Psychiatry, *208*(1), 17-25. https://doi.org/10.1192/bjp.bp.114.149997

Field, A. (2015*). Discovering Statistics Using IBM SPSS Statistics – Fifth Edition*. London: Sage.

First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., Davies, M., Borus, J., Howes, M. J., Kane, J., Pope, H. G., & Rounsaville, B. (2011). The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part II: Multi-Site Test-Retest Reliability Study. *Journal of Personality Disorders*, *9*(2), 92–104. https://doi.org/10.1521/PEDI.1995.9.2.92

Frances, A. (2012). *DSM-5 field trials discredit the American Psychiatric Association.* Huffington Post Science. The Blog. Posted 31 October 2012. Retrieved from: https://www.huffpost.com/entry/dsm-5-reliability-tests_b_1490857

Furnham, A., Lee, V., & Kolzeev, V. (2015). Mental health literacy and borderline personality disorder (BPD): what do the public "make" of those with BPD? *Social psychiatry and psychiatric epidemiology*, *50*(2), 317–324. https://doi.org/10.1007/s00127-014-0936-7

Garand, L., Lingler, J. H., Conner, K. O., & Dew, M. A. (2009). Diagnostic labels, stigma, and participation in research related to dementia and mild cognitive impairment. *Research in Gerontological Nursing*, *2*(2), 112–121. https://doi.org/10.3928/19404921-20090401-04

Garcia, D. J., Skadberg, R. M., Schmidt, M., Bierma, S., Shorter, R. L., & Waugh, M. H. (2018). It's not that Difficult: An Interrater Reliability Study of the DSM-5 Section III Alternative Model for Personality Disorders. *Journal of Personality Assessment*, *100*(6), 612–620. https://doi.org/https://dx.doi.org/10.1080/00223891.2018.1428982

Goffman, E. (2009). *Stigma: Notes on the management of spoiled identity*. New York: Simon and Schuster.

Gove W. R. (1970). Societal reaction as an explanation of mental illness: an evaluation. *American Sociological Review*, *35*(5), 873–884. https://doi.org/10.2307/2093298

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*(3), 499-510. https://doi.org/10.1207/s15327906mbr2603_7

Gutheil, T. G., Schetky, D. H., & Simon, R. I. (2006). Pejorative testimony about opposing experts and colleagues: "fouling one's own nest". *The journal of the American Academy of Psychiatry and the Law*, *34*(1), 26–30.

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine*, *42*(5), 903–920. https://doi.org/10.1017/S0033291711001966

Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: a meta-analysis of taxometric research. *Psychological Medicine*, *50*(9), 1418–1432. https://doi.org/10.1017/S003329172000183X

Hemmati A, Rahmani F and Bach B (2021) The ICD-11 Personality Disorder Trait Model Fits the Kurdish Population Better Than the DSM-5 Trait Model. *Front. Psychiatry* 12:635813. doi: 10.3389/fpsyt.2021.635813

Herkov, M. J., & Blashfield, R. K. (2010). Clinician Diagnoses of Personality Disorders: Evidence of a Hierarchical Structure. *Journal of Personality Assessment*, *65*(2), 313–321. https://doi.org/10.1207/S15327752JPA6502_7

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, 37(5), 1158–1160. https://doi.org/10.1093/ije/dyn204.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. https://doi.org/10.1002/sim.1186.

HM Chief Inspector of Prisons (2021) *Annual Report 20-21*. London: HMSO. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1003082/hmip-annual-report-accounts-2020-21.pdf

House of Commons Justice Committee (2021). *Mental Health in Prison. Fifth Report of the Session, 2021-2022.* House of Commons, London. Retrieved from: https://committees.parliament.uk/publications/7455/documents/78054/default/

Howard, J. A., & Renfrow, D. G. (2014). Intersectionality. In J. D. McLeod, E. J. Lawler, & M. Schwalbe (Eds.), *Handbook of the social psychology of inequality* (pp. 95–121). New York, NY: Springer.

Hutsebaut, J., Berghuis, H., De Saeger, H., Kaasenbrood, A., & Ingenhoven, T. (2014). Semi-structured Interview for Personality Functioning DSM-5 (STiP-51). *The Podium DSM-5 Research Group of the Netherlands Centre of Expertise on Personality Disorders. Utrecht, the Netherlands: Trimbos Institute*.

Hutsebaut, J., Kamphuis, J. H., Feenstra, D. J., Weekers, L. C., & de Saeger, H. (2017). Assessing DSM5-oriented level of personality functioning: Development and psychometric evaluation of the semi-structured interview for personality functioning DSM5 (STiP-5.1). *Personality Disorders: Theory, Research, and Treatment*, *8*(1), 94–101. https://doi.org/10.1037/per0000197

Hutsebaut, J., Weekers, L. C., Tuin, N., Apeldoorn, J. S. P., & Bulten, E. (2021). Assessment of ICD-11 Personality Disorder Severity in Forensic Patients Using the Semi-structured Interview for Personality Functioning DSM-5 (STiP-5.1): Preliminary Findings. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.617702

Jakobowitz, S., Bebbington, P., McKenzie, N., Iveson, R., Duffield, G., Kerr, M., & Killaspy, H. (2017). Assessing needs for psychiatric treatment in prisoners: 2. Met and unmet

need. *Social psychiatry and psychiatric epidemiology*, *52*(2), 231–240. https://doi.org/10.1007/s00127-016-1313-5

Johnstone, L., & Boyle, M. (2018). The Power Threat Meaning Framework: An Alternative Nondiagnostic Conceptual System. *Journal of Humanistic Psychology.* https://doi.org/10.1177/0022167818793289

Kampe, L., Zimmerman, J., Bender, D., Caligor, E., Borowski, A.-L., Ehrenthal, J. C., Benecke, C., & Hörz-Sagstetter, S. (2018). Comparison of the structured DSM–5 clinical interview for the Level of Personality Functioning Scale with the Structured Interview of Personality Organization. *Journal of Personality Assessment*, *100*(6), 642–649. https://doi.org/10.1080/00223891.2018.1489257

Kinderman, P., Read, J., Moncrieff, J., & Bentall, R. P. (2013). Drop the language of disorder. *Evidence-Based Mental Health*, *16*(1), 2–3. http://dx.doi.org/10.1136/eb-2012-100987

Kitamura, T., & Kitamura, F. (2000). Reliability of clinical judgment of patients' competency to give informed consent: A case vignette study. *Psychiatry and Clinical Neurosciences*, *54*(2), 245–247. https://doi.org/10.1046/J.1440-1819.2000.00665.X

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155. https://doi.org/10.1016/J.JCM.2016.02.012

Kotov, R., Jonas, K. G., Carpenter, W. T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Hobbs, K., Reininghaus, U., Slade, T., South, S. C., Sunderland, M., Waszczuk, M. A., Widiger, T. A., Wright, A., Zald, D. H., Krueger, R. F., & Watson, D. (2020). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. *World Psychiatry*, *19*(2), 151–172. https://doi.org/10.1002/wps.20730

Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. (2012). DSM-5: How Reliable Is Reliable Enough? *The American Journal of Psychiatry. 169*(1), 13–15. https://doi.org/10.1176/APPI.AJP.2011.11010050

Krajewski, C., Burazeri, G., & Brand, H. (2013). Self-stigma, perceived discrimination and empowerment among people with a mental illness in six countries: Pan European stigma study. *Psychiatry research*, *210*(3), 1136–1146. https://doi.org/10.1016/j.psychres.2013.08.013

Krueger, R. F., & Hobbs, K. A. (2020). An Overview of the DSM-5 Alternative Model of Personality Disorders. *Psychopathology*, *53*(3), 126–132. https://doi.org/10.1159/000508538

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial Construction of a Maladaptive Personality Trait Model and Inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879. https://doi.org/10.1017/S0033291711002674

Krueger, R. F., Skodol, A. E., Livesley, W. J., Shrout, P. E., & Huang, Y. (2007). Synthesizing dimensional and categorical approaches to personality disorders: refining the research agenda for DSM-V Axis II. *International Journal of Methods in Psychiatric Research*, *16*(1), S65. https://doi.org/10.1002/MPR.212

Kupfer, D. (2013, 3 May). Chair of DSM-5 task force discusses future of mental health research. American Psychiatric Association. Retrieved from https://www.madinamerica.com/wp-content/uploads/2013/05/Statementfrom-dsm-chair-david-kupfer-md.pdf

Lam, D. C. K., Salkovskis, P. M., & Hogg, L. I. (2016). 'Judging a book by its cover': An experimental study of the negative impact of a diagnosis of borderline personality disorder on clinicians' judgements of uncomplicated panic disorder. *British Journal of Clinical Psychology*, *55*(3), 253–268. https://doi.org/10.1111/BJC.12093

Latham, D. (2017). *Mental health and fair trial: A Report by Justice*. London: Justice.

Lester, R., Prescott, L., McCormack, M., & Sampson, M. (2020). Service users' experiences of receiving a diagnosis of borderline personality disorder: A systematic review. *Personality and Mental Health*, *14*(3), 263–283. https://doi.org/10.1002/PMH.1478

Levey, S., & Howells, K. (1995). Dangerousness, unpredictability and the fear of people with schizophrenia. *Journal of Forensic Psychiatry, 6*(1), 19–39. https://doi.org/10.1080/09585189508409874

Lewis, G., & Appleby, L. (1988). Personality Disorder: the Patients Psychiatrists Dislike. *The British Journal of Psychiatry*, *153*(1), 44–49. https://doi.org/10.1192/BJP.153.1.44

Link, B. G., & Cullen, F. T. (1983). Reconsidering the social rejection of ex-mental patients: Levels of attitudinal response. *American Journal of Community Psychology*, *11*(3), 261- 273. https://doi.org/10.1007/BF00893367

Link, B. G., & Phelan, J. (2014). Stigma power. *Social science & medicine*, *103*, 24-32. https://doi.org/10.1016/j.socscimed.2013.07.035

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual review of Sociology*, *27*(1), 363-385. https://doi.org/10.1146/annurev.soc.27.1.363

Link, B. G., Cullen, F. T., Frank, J., & Wozniak, J. F. (1987). The social rejection of former mental patients: Understanding why labels matter. *American Journal of Sociology, 92*(6), 1461–1500. https://doi.org/10.1086/228672

Link, B. G., Cullen, F. T., Struening, E. L., Shrout, P. E., & Dohrenwend, B. P. (1989). A modified labeling theory approach to mental disorders: An empirical assessment. American Sociological Review, *54*(3), 400–423. https://doi.org/10.2307/2095613

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2001). Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people

with mental illnesses. Psychiatric Services, *52*(12), 1621-1626 https://doi.org/10.1176/appi.ps.52.12.1621

Listwan, S. J., Sullivan, C. J., Agnew, R., Cullen, F. T., & Colvin, M. (2012). The Pains of Imprisonment Revisited: The Impact of Strain on Inmate Recidivism. *Http://Dx.Doi.Org/10.1080/07418825.2011.597772*, *30*(1), 144–168. https://doi.org/10.1080/07418825.2011.597772

Lorenzetti, D.L., Ghali, W.A. (2013) Reference management software for systematic reviews and meta-analyses: an exploration of usage and usability. *BMC Med Res Methodol* **13,** 141 https://doi.org/10.1186/1471-2288-13-141

Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of clinical epidemiology*, *63*(8), 854-861.

Markham, D., & Trower, P. (2003). The effects of the psychiatric label 'borderline personality disorder' on nursing staff's perceptions and causal attributions for challenging behaviours. *British Journal of Clinical Psychology*, *42*(3), 243-256. https://doi.org/10.1348/01446650360703366

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). *Paying the price. The cost of mental health care in England to 2026.* The Kings Fund, London.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031/FULLARTICLE

Mental Health Act, 1983. London: Stationery Office.

Mezzich J. E. (2002). International surveys on the use of ICD-10 and related diagnostic systems. *Psychopathology*, *35*(2-3), 72–75. https://doi.org/10.1159/000065122

Milton, J. (2000). A postal survey of the assessment procedure for personality disorder in forensic settings. *Psychiatric Bulletin*, *24*(7), 254–257. https://doi.org/10.1192/PB.24.7.254

Ministry of Justice (2010) *Breaking the Cycle: Effective Punishment, Rehabilitation and Sentencing of Offenders*. London: HMSO. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/185936/breaking-the-cycle.pdf

Ministry of Justice (2013). *Transforming Rehabilitation: A Strategy for Reform*. London: HMSO.

Ministry of Justice (2017). *Mentally disordered offenders - the restricted patient system – Background Briefing.* London: HMSO.

Ministry of Justice and Department of Health (2011). *Working with personality disordered offenders: a practitioner's gui*de. London: National Offender Management Service and Department of Health.

Ministry of Justice. (2018). *Offender Management Statistics quarterly: October to December 2017.* https://www.gov.uk/government/statistics/offender-management-statistics-quarterly-october-to-december-2017

Ministry of Justice. (2020). *Diversity of the judiciary: Legal professions, new appointments and current post-holders.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/918529/diversity-of-the-judiciary-2020-statistics-web.pdf

Ministry of Justice. (2021a). *Criminal Justice Statistics Quarterly June 2021* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1034457/criminal-justice-statistics-june-2021_Q2_revised.pdf

Ministry of Justice. (2021b). *Ethnicity and the Criminal Justice System - 2020.* Retrieved May 16, 2022, from https://www.gov.uk/government/statistics/ethnicity-and-the-criminal-justice-system-statistics-2020/ethnicity-and-the-criminal-justice-system-2020

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, *339*(7716), 332–336. https://doi.org/10.1136/BMJ.B2535

Monahan, J. (1981). *Predicting Violent Behavior - An Assessment of Clinical Techniques.* Office of Justice Programs, Sage, Beverly Hills, CA.

Morey, L. C. (2019). Application of the DSM-5 Level of Personality Functioning Scale by Lay Raters. *Journal of Personality Disorders*, *32*(5), 709–720. https://doi.org/10.1521/pedi_2017_31_305

Mossière, A., & Maeder, E. M. (2015). Defendant mental illness and juror decision-making: A comparison of sample types. International journal of law and psychiatry, 42-43, 58–66. https://doi.org/10.1016/j.ijlp.2015.08.008

Mossière, A., & Maeder, E. M. (2016). Juror decision making in not criminally responsible on account of mental disorder trials: Effects of defendant gender and mental illness type. *International Journal of Law and Psychiatry*, *49*, 47-54. https://doi.org/10.1016/j.ijlp.2016.05.008

Mulder, R. T., Newton-Howes, G., Crawford, M. J., & Tyrer, P. J. (2011). The Central Domains of Personality Pathology in Psychiatric Patients. *Journal of Personality Disorders*, *25*(3), 364–377. https://doi.org/10.1521/PEDI.2011.25.3.364

National Institute for Mental Health in England. (2003). *Personality Disorder: No Longer a Diagnosis of Exclusion.* London: Department of Health.

Nee, C., & Witt, C. (2013). Public perceptions of risk in criminality: The effects of mental illness and social disadvantage. *Psychiatry Research*, *209*(3), 675-683. https://doi.org/10.1016/j.psychres.2013.02.013

Neumann, N., Böckenholt, U., & Sinha, A. (2016). A meta-analysis of extremeness aversion. *Journal of Consumer Psychology*, *26*(2), 193-212. https://doi.org/10.1016/j.jcps.2015.05.005

Newton-Howes, G., Clark, L. A., & Chanen, A. (2015). Personality disorder across the life course. The Lancet, 385(9969), 727–734. https://doi.org/10.1016/S0140-6736(14)61283-6

NHS Benchmarking Network. (2018). Waiters for mental health hospital transfer and remission: Analysis of NHS England Specialised Commissioning and Health & Justice, and Her Majesty's Prison and Probation Services audits. Retrieved from: https://s3.eu-west-2.amazonaws.com/nhsbn-static/Other/2018/Data%20Transfers%20and%20Remissions%20Census%20Report.pdf

NHS Digital. (2021). *Use of NHS mental health, learning disability and autism services - GOV.UK Ethnicity facts and figures*. Retrieved May 16, 2022, from https://www.ethnicity-facts-figures.service.gov.uk/health/mental-health/adults-using-nhs-funded-mental-health-and-learning-disability-services/latest

NICE. (2009). *Borderline personality disorder: recognition and management. Clinical guideline* (Guideline CG78). Retrieved from: https://www.nice.org.uk/guidance/cg78

NICE. (2017). *Mental health of adults in contact with the criminal justice system.* (Guideline NH66). Retrieved from: https://www.nice.org.uk/guidance/ng66

O'Loughlin, A. (2021). Sentencing Mentally Disordered Offenders in England and Wales: Towards a Rights-Based Approach. *Criminal Law Review*, (2), 98-112.

O'Loughlin, A. (2014). The Offender Personality Disorder Pathway: Expansion in the Face of Failure? *The Howard Journal of Criminal Justice*, *53*(2), 173–192. https://doi.org/10.1111/HOJO.12058

Offences Against the Person Act, 1861.

Olbert, C. M., Gala, G. J., & Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *Journal of abnormal psychology*, *123*(2), 452–462. https://doi.org/10.1037/a0036068

Peay, J. (2016). Responsibility, culpability and the sentencing of mentally disordered offenders: objectives in conflict. *Criminal Law Review (3),* 152-164.

Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *The American Journal of Psychiatry*, *149*(12), 1645–1653. https://doi.org/10.1176/AJP.149.12.1645

Pilkonis, P. A., Hallquist, M. N., Morse, J. Q., & Stepp, S. D. (2011). Striking the (Im)proper balance between scientific advances and clinical utility: Commentary on the DSM-5 proposal for personality disorders. *Personality Disorders: Theory, Research, and Treatment*, *2*(1), 68–82. https://doi.org/10.1037/A0022226

Pincus, A. L. (2011). Some comments on nomology, diagnostic process, and narcissistic personality disorder in the DSM-5 proposal for personality and personality disorders. Personality Disorders: Theory, Research, and Treatment, 2(1), 41–53. https://doi.org/10.1037/A0021191

Preti, E., di Pierro, R., Costantini, G., Benzi, I. M. A., de Panfilis, C., & Madeddu, F. (2018). Using the Structured Interview of Personality Organization for DSM–5 Level of Personality Functioning Rating Performed by Inexperienced Raters. *Journal of*

Personality Assessment, *100*(6), 621–629. https://doi.org/10.1080/00223891.2018.1448985

R v Edwards (2018) EWCA Crim 595

R v Fort. EWCA Crim 2332. (2013).

R v. Vowles (2015) EWCA Crim 45

R. v Challen. EWCA Crim 916. (2019).

Roche, M. J., Jacobson, N. C., & Phillips, J. J. (2018). Expanding the Validity of the Level of Personality Functioning Scale Observer Report and Self-Report Versions Across Psychodynamic and Interpersonal Paradigms. https://doi.org/10.1080/00223891.2018.1475394

Rosenfield, S. (1997). Labelling mental illness: The effects of received services and perceived stigma on life satisfaction. *American Sociological Review, 62*(4), 660–672. https://doi.org/10.2307/2657432

Royal College of Psychiatrists (1996). *Report of the confidential inquiry into homicides and suicides by mentally disordered offenders*. London: RCP.

Samuel, D. B. (2015). A review of the agreement between clinicians' personality disorder diagnoses and those from other methods and sources. *Clinical Psychology: Science and Practice,* 22(1), 1–19. https://doi.org/10.1111/cpsp.12088

Samuel, D. B., Sanislow, C. A., Hopwood, C. J., Shea, M. T., Skodol, A. E., Morey, L. C., ... & Grilo, C. M. (2013). Convergent and incremental predictive validity of clinician, self-report, and structured interview diagnoses for personality disorders over 5 years. *Journal of consulting and clinical psychology*, *81*(4), 650.

Sansone, R. A., & Sansone, L. A. (2011). Gender Patterns in Borderline Personality Disorder. *Innovations in Clinical Neuroscience*, *8*(5), 16-20.

Scheff, T. (1974). The Labelling Theory of Mental Illness. *American Sociological Review, 39*(3), 444-452. https://doi.org/10.2307/2094300

Scheff, T. J. (1966). *Being mentally ill III: A sociological theory*. Chicago: Aldine

Schuster, C. (2004). A Note on the Interpretation of Weighted Kappa and its Relations to Other Rater Agreement Statistics for Metric Scales. *Educational and Psychological Measurement*, *64*(2), 243–253. https://doi.org/10.1177/0013164403260197

Sentencing Council. (2020). *Sentencing offenders with mental disorders, developmental disorders, or neurological impairments.* Retrieved from: https://www.sentencingcouncil.org.uk/overarching-guides/magistrates-court/item/sentencing-offenders-with-mental-disorders-developmental-disorders-or-neurological-impairments/

Sheehan, L., Nieweglowski, K., & Corrigan, P. (2016). The Stigma of Personality Disorders. *Current psychiatry reports*, *18*(1), Article 11. https://doi.org/10.1007/s11920-015-0654-1

Singleton, N., Gatward, R., & Meltzer, H. (1998). *Psychiatric morbidity among prisoners in England and Wales.* London: Stationery Office.

Skodol, A. E., Gunderson, J. G., McGlashan, T. H., Dyck, I. R., Stout, R. L., Bender, D. S., Grilo, C. M., Tracie Shea, M., Zanarini, M. C., Morey, L. C., Sanislow, C. A., & Oldham, J. M. (2002). Functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder. *American Journal of Psychiatry*, *159*(2), 276–283. https://doi.org/10.1176/APPI.AJP.159.2.276

Somma, A., Borroni, S., Carlotta, D., Giarolli, L. E., Barranca, M., Cerioli, C., Franzoni, C., Masci, E., Manini, R., Busso, S. L., Ruotolo, G., & Fossati, A. (2019). The inter-rater reliability and convergent validity of the Italian translation of the Structured Clinical

Interview for the DSM-5 Alternative Model of Personality Disorders Module III in a psychotherapy outpatient sample. *Journal of Psychopathology*, *25*(4), 195–204.

Somma, A., Borroni, S., Gialdi, G., Carlotta, D., Emanuela Giarolli, L., Barranca, M., Cerioli, C., Franzoni, C., Masci, E., Manini, R., Luca Busso, S., Ruotolo, G., Krueger, R. F., Markon, K. E., & Fossati, A. (2020). The Inter-Rater Reliability and Validity of the Italian Translation of the Structured Clinical Interview for DSM-5 Alternative Model for Personality Disorders Module I and Module II: A Preliminary Report on Consecutively Admitted Psychotherapy Outpatients. *Journal of Personality Disorders*, *34*(Supplement C), 95–123. https://doi.org/https://dx.doi.org/10.1521/pedi_2020_34_511

Taylor, P., Eastman, N., Latham, R., & Holloway, J. (2021). Sentencing offenders with mental disorders, developmental disorders or neurological impairments: What does the new Sentencing Council Guideline mean for psychiatrists? *The British Journal of Psychiatry*, *218*(6), 299-301. https://doi.org/10.1192/bjp.2021.21

The Bar Council (2022). The Impact of the Covid-19 Pandemic on the Criminal Bar. Retrieved from: https://www.barcouncil.org.uk/resource/bar-council-data-analysis-criminal-bar-april-2022.html

Thomas, C. (2010). Are juries fair? (Ministry of Justice Research Series 1/10) London: Ministry of Justice. Retrieved from https://www.justice.gov.uk/downloads/publications/research-andanalysis/moj-research/are-juries-fair-research.pdf

Thylstrup, B., Simonsen, S., Nemery, C., Simonsen, E., Noll, J. F., Myatt, M. W., & Hesse, M. (2016). Assessment of personality-related levels of functioning: A pilot study of clinical assessment of the DSM-5 level of personality functioning based on a semi-

structured interview. *BMC Psychiatry*, *16*(1), 1–8. https://doi.org/10.1186/S12888-016-1011-6/TABLES/2

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tyrer, P. (2012). Diagnostic and Statistical Manual of Mental Disorders: A classification of personality disorders that has had its day [Editorial]. *Clinical Psychology & Psychotherapy*, 19(5), 372–374. https://doi.org/10.1002/cpp.1810

Tyrer, P. (2014). A comparison of DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment, 20*(4), 280-285. https://doi:10.1192/apt.bp.113.011296

Tyrer, P., Coombs, N., Ibrahimi, F., Mathilakath, A., Bajaj, P., Ranger, M., Rao, B., & Din, R. (2007). Critical developments in the assessment of personality disorder. *The British Journal of Psychiatry*, *190*(S49), s51–s59. https://doi.org/10.1192/BJP.190.5.S51

Varshney, M., Mahapatra, A., Krishnan, V., Gupta, R., & Deb, K. S. (2016). Violence and mental illness: what is the true story?. *Journal of epidemiology and community health*, *70*(3), 223–225. https://doi.org/10.1136/jech-2015-205546

Verheul, R., & Widiger, T. A. (2004). A meta-analysis of the prevalence and usage of the personality disorder not otherwise specified (PDNOS) diagnosis. *Journal of Personality Disorders*, *18*(4), 309–319. https://doi.org/10.1521/PEDI.18.4.309.40350

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. https://doi:10.18637/jss.v036.i03

Wallace, B. C., Dahabreh, I. J., Trikalinos, T. A., Lau, J., Trow, P., & Schmid, C. H. (2012). Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. Journal of Statistical Software, 49(5), 1–15.https://doi.org/10.18637/jss.v049.i05

Westen, D., & Shedler, J. (1999). Revising and assessing Axis II, Part 2: Towards an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry*, 156, 273-285.

Widiger, T. A., Livesley, W. J., & Clark, L. A. (2009). An Integrative Dimensional Classification of Personality Disorder. *Psychological Assessment*, *21*(3), 243–255. https://doi.org/10.1037/A0016606

Winnick, T. A., & Bodkin, M. (2008). Anticipated stigma and stigma management among those to be labelled "ex-con". *Deviant Behavior, 29*(4), 295-333. https://doi.org/10.1080/01639620701588081

Winship, G., & Hardy, S. (2007). Perspectives on the prevalence and treatment of personality disorder. *Journal of Psychiatric and Mental Health Nursing*, *14*(2), 148–154. https://doi.org/10.1111/J.1365-2850.2007.01057.X

Winsper, C., Bilgin, A., Thompson, A., Marwaha, S., Chanen, A. M., Singh, S. P., Wang, A., & Furtado, V. (2019). The prevalence of personality disorders in the community: a global systematic review and meta-analysis. https://doi.org/10.1192/bjp.2019.166

Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology, 53*(1), 519-543. https://doi.org/10.1146/annurev.psych.53.100901.135136

World Health Organisation (2018). International Statistical Classification of Diseases and Related Health Problems (11th ed.). https://icd.who.int/

World Health Organisation (2019). International Statistical Classification of Diseases and Related Health Problems (11th ed.). https://icd.who.int/

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva: World Health Organization.

Wright, A., & Zimmerman, J. (2015). At the nexus of science and practice: Answering basic clinical questions in personality disorder assessment and diagnosis with quantitative modeling techniques. *Personality Disorders: Toward Theoretical and Empirical Integration in Diagnosis and Assessment.*, 109–144. https://doi.org/10.1037/14549-006

Zettl, M., Volkert, J., Vogele, C., Herpertz, S. C., Kubera, K. M., & Taubner, S. (2020). Mentalization and criterion a of the alternative model for personality disorders: Results from a clinical and nonclinical sample. *Personality Disorders*, *11*(3), 191–201. https://doi.org/https://dx.doi.org/10.1037/per0000356

Zimmerman, J., Benecke, C., Bender, D. S., Skodol, A. E., Schauenburg, H., Cierpka, M., & Leising, D. (2014). Assessing DSM–5 Level of Personality Functioning From Videotaped Clinical Interviews: A Pilot Study With Untrained and Clinically Inexperienced Students. *96*(4), 397–409. https://doi.org/10.1080/00223891.2013.852563

Appendix A: Summary of Linked Project

Research Study Summary

The Empirical Project involves elements of joint work completed by myself and anther Trainee under the supervision of Dr Peter Beazley. This required a joint ethics application to the University of East Anglia. Together, we devised the script for the video vignette, which was used within our joint empirical project. Recruitment and data collection also took place jointly with the same Trainee. However, they measured a separate psychological construct (origins of mental health) and had separate research questions. These are stated below:

1. How consistently is *Vowles* applied?
2. Are individual *Vowles* criteria associated with sentencing outcome?
3. Are differences in beliefs about the origin of mental health associated with different application of *Vowles*?
4. Does diagnosis predict sentencing outcome?
5. Are differences in beliefs about the origin of mental health associated with different sentencing outcomes?

Participants' beliefs about the origin of mental health difficulties was measured using The Mental Health Locus of Origin (MHLO) Scale (Bale and Hill, 1980). The MHLO scale consists of twenty items on a six-point Likert scale, measuring a single dimension with two polarised types of belief: "endogenous" (biogenetic) and "interactional" (environmental). All participants completed all measures, however this measure was solely for the purpose of the other trainee project and data was not analysed by me. This study used the Perceived Devaluation-Discrimination Scale (Link, 1987) which was used solely for the purposes of this project and not analysed by the other Trainee. The order of these two measures was

always counterbalanced to alleviate order effects. All statistical analysis was conducted independently, to test different hypotheses based on largely different bodies of literature.

Appendix B: Manuscript Guidelines

Journal of Personality Assessment

Preparing Your Paper

Structure

Your manuscript should be compiled in the following order: title page; abstract; keywords; main text introduction, methods, results, discussion; acknowledgments; declaration of interest statement; references; table(s) with caption(s) (on individual pages); figures; figure captions (as a list).

Word Limits

Please include a word count for your manuscript.

A typical manuscript for this journal should be no more than 40 pages, inclusive of the abstract, tables, references, figure captions, footnotes, etc. However, longer manuscripts can be considered if absolutely necessary; if so, please seek permission from the Editor-in-Chief prior to submission (and such permission should be clearly stated in a cover letter).

Style Guidelines

Please refer to these quick style guidelines when preparing your manuscript, rather than any published articles or a sample copy. Your manuscript should be carefully prepared in accordance with APA style guidelines. Please refer to The Publication Manual of the American Psychological Association, Seventh Edition, when preparing your manuscript.

Please use American spelling style consistently throughout your manuscript.

Manuscripts should be submitted in MS Word format.

Formatting and Templates

Papers may be submitted in Word format. Figures should be saved separately from the text. To assist you in preparing your manuscript, we provide formatting template(s).

Word templates are available for this journal. Please save the template to your hard drive, ready for use.

If you are not able to use the template via the links (or if you have any other template queries) please contact us here.

References

Please use this reference guide when preparing your manuscript.

Appendix C: Identification and Selection of Studies Interrater Reliability

Table 1.
Title and Abstract Screening Interrater Reliability

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.831	.092	5.060	<.001
N of Valid Cases		37			

Note: a = Not assuming the null hypothesis.
b = Using the asymptotic standard error assuming the null hypothesis

Table 2.
Full Article Screening Interrater Reliability

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.851	.142	3.220	<.001
N of Valid Cases		14			

Appendix D: Methodological quality and risk of bias of included studies assessed with the Quality Appraisal of Reliability Studies*

Study	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	% yes
Buer Christensen et al. (2018)	Y	Y	Y	N/A	Y	Y	Y	N/A	N/A	Y	Y	72.7
Cruitt et al. (2019)	Y	Y	Y	N/A	Y	Y	Y	U	N/A	Y	Y	72.7
Dereboy et al. (2018)	Y	Y	Y	N/A	Y	Y	U	N	N/A	Y	Y	63.6
Garcia et al. (2018)	N	Y	Y	N/A	Y	Y	Y	N/A	N/A	Y	Y	63.6
Hutsebaut et al. (2017)	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	90.9
Hutsebaut et al. (2021)	Y	Y	Y	N/A	Y	Y	Y	N	Y	Y	Y	81.8
Kampe et al. (2018)	Y	Y	Y	N/A	N/A	U	U	Y	Y	Y	Y	63.6
Morey (2019)	N	Y	Y	N/A	Y	Y	Y	Y	N/A	Y	Y	72.7
Preti et al. (2018)	Y	Y	Y	N/A	Y	Y	Y	N/A	N/A	Y	Y	72.7
Roche et al. (2018)	N	Y	Y	N/A	N/A	Y	Y	N/A	N/A	Y	Y	54.5
Somma et al. (2019)	Y	Y	Y	N/A	N/A	Y	Y	Y	N/A	Y	Y	72.7
Somma et al. (2020)	Y	Y	Y	N/A	Y	Y	Y	Y	Y	Y	Y	90.9
Thylstrup et al. (2016)	Y	Y	Y	N/A	N/A	Y	Y	N/A	N/A	Y	Y	63.6
Zettl et al (2020)	Y	Y	Y	N/A	Y	Y	U	N	N/A	Y	Y	63.6
Zimmerman et al. (2014)	Y	Y	Y	N/A	Y	Y	Y	U	N/A	Y	Y	72.7

**Assesses study quality based on 11 items. Items: 1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? 2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? 3. Were raters blinded to the findings of other raters during the study? 4. Were raters blinded to their own prior findings of the test under evaluation? 5. Were raters blinded to the results of the accepted reference standard or the disease status for the target disorder (or variable) being evaluated? 6. Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? 7. Were raters blinded to additional cues that were not part of the test? 8. Was the order of examination varied? 9. Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time-interval among repeated measures? 10. Was the test applied correctly and interpreted appropriately? 11. Were appropriate statistical measures of agreement used? N, no; NA, not applicable; U, unclear; Y, yes (marked in b*

Appendix E-Sensitivity Analysis

Table 1.

Meta-analysis outcomes for sensitivity analysis of Criterion A LPFS and Criteria B Trait Domains Interrater Reliability, removal of studies with high risk of bias (i.e.Roche et al., 2018)

Variable		*k*	ICC	95% Confidence LL	95% Confidence UL	SE	*p*	T^2	*z*	*Q*	*df*	*p*	I^2
TOTAL LPFS		14	0.80	0.70	0.87	0.1207	<.0001	0.1497	9.2039	100.6601	13	<.0001	85.90
Domains	Identity	10	0.77	0.62	0.86	0.1502	<.0001	0.1829	6.7355	92.1048	9	<.0001	89.49
	Self-Direction	10	0.77	0.63	0.85	0.1390	<.0001	0.1516	7.3127	94.5120	9	<.0001	87.59
	Empathy	10	0.73	0.57	0.84	0.1483	<.0001	0.1774	6.2784	84.0936	9	<.0001	89.20
	Intimacy	10	0.82	0.68	0.90	0.1620	<.0001	0.2188	7.0857	146.4512	9	<.0001	91.06

Appendix F - Funnel Plot assessing publication bias for Total LPFS

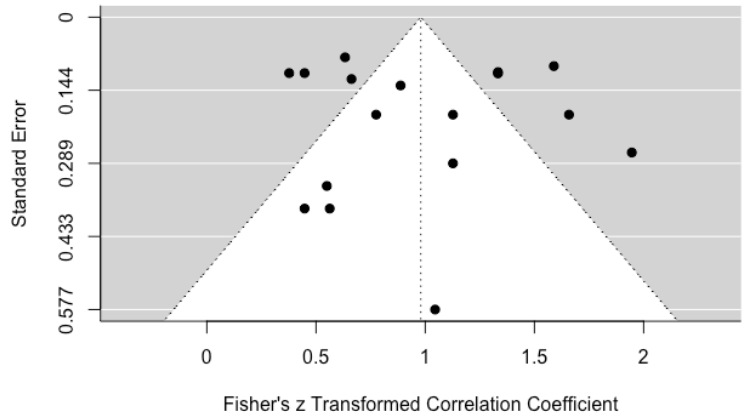

Figure 1. Funnell plot assessing bias for Total LPFS Score meta-analysis

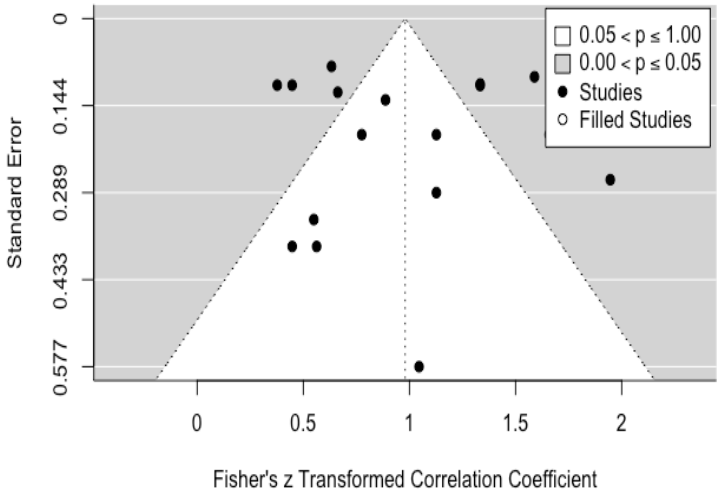

Figure 2. Funnell plot assessing bias for Total LPFS Score meta-analysis – Trim and Fill

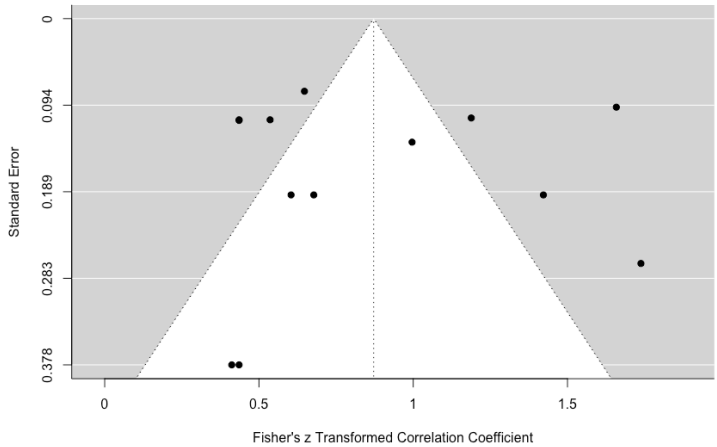

Figure 3. Funnell Plot for assessing publication bias of LPFS Domain – Identity

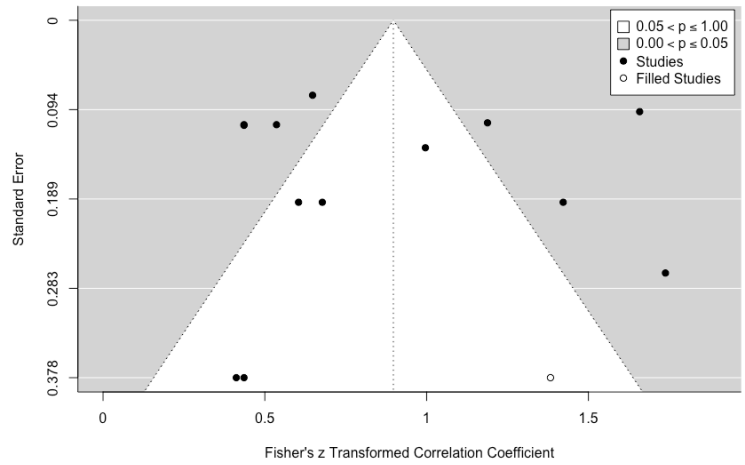

Figure 4. Funnell Plot for assessing publication bias of LPFS Domain Identity – Trim and Fill analysis

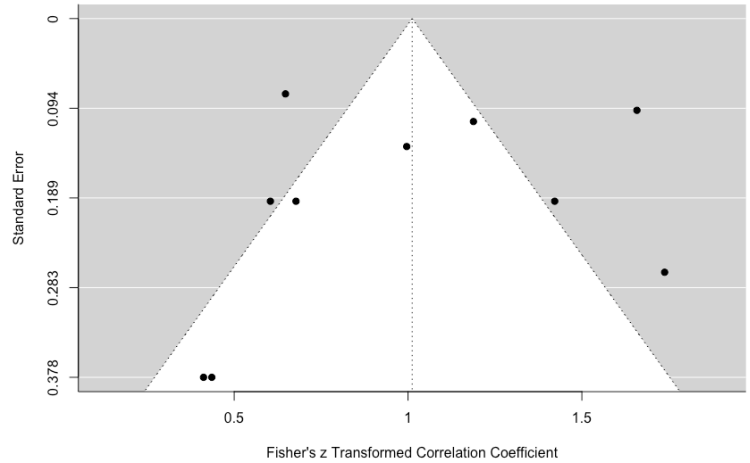

Figure 5. Funnell Plot for assessing publication bias of LPFS Domain – Identity – Sensitivity Analysis

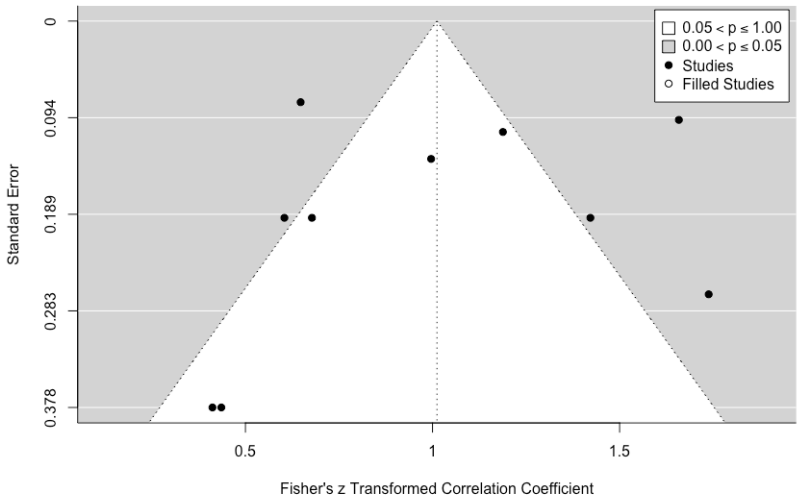

Figure 6. Funnell Plot for assessing publication bias of LPFS Domain – Identity – Sensitivity Analysis- Trim and Fill analysis

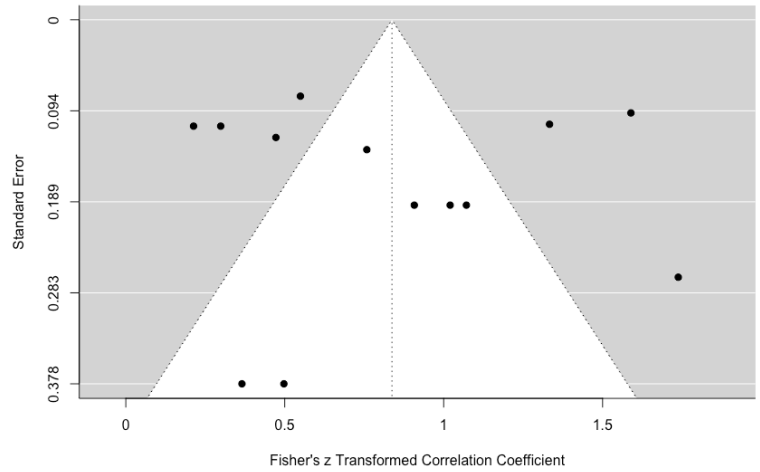

Figure 7. Funnell Plot for assessing publication bias of LPFS Domain - Self-Direction

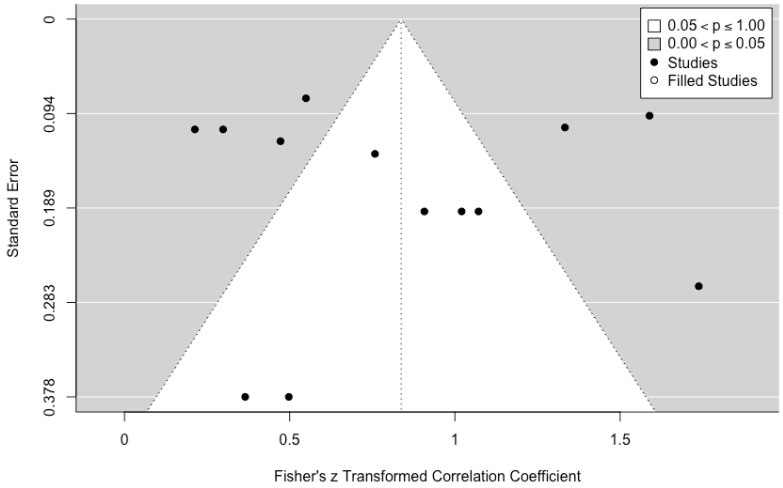

Figure 8. Funnell Plot for assessing publication bias of LPFS Domain - Self-Direction – Trim and Fill analysis

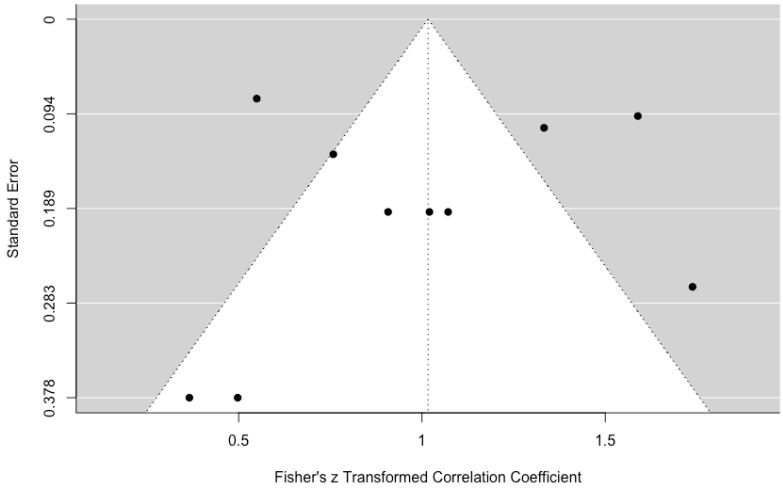

Figure 9. Funnell Plot for assessing publication bias of LPFS Domain - Self-Direction – Sensitivity Analysis

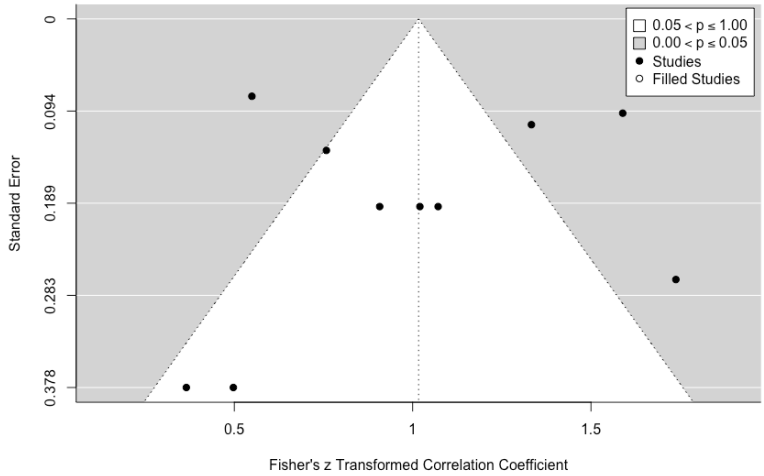

Figure 10. Funnell Plot for assessing publication bias of LPFS Domain - Self-Direction – Sensitivity Analysis – Trim and Fill analysis

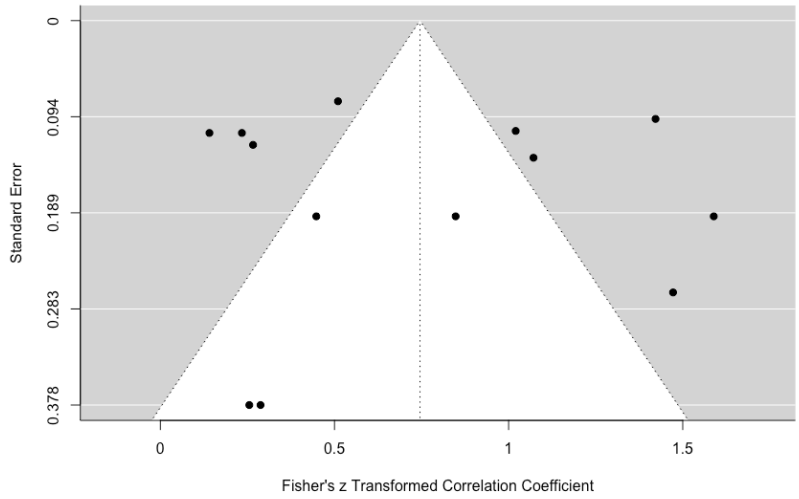

Figure 11. Funnell Plot for assessing publication bias of LPFS Domain - Empathy

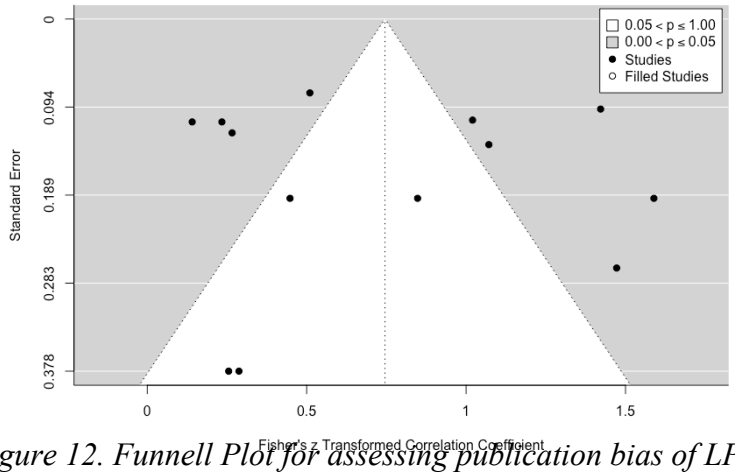

Figure 12. Funnell Plot for assessing publication bias of LPFS Domain - Empathy – Trim and Fill analysis

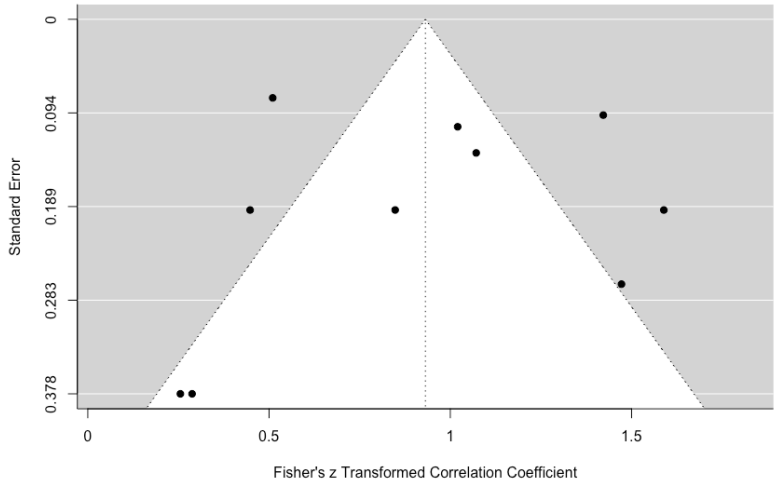

Figure 13. Funnell Plot for assessing publication bias of LPFS Domain - Empathy - Sensitivity Analysis.

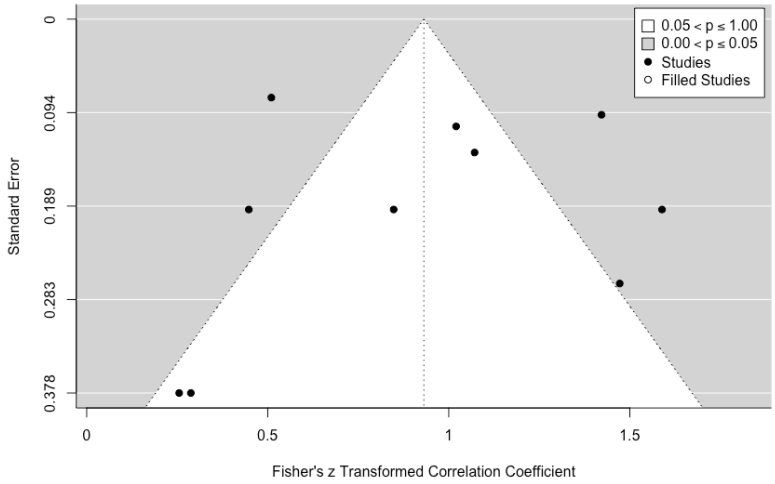

Figure 14. Funnell Plot for assessing publication bias of LPFS Domain - Empathy - Sensitivity Analysis - Trim and Fill analysis

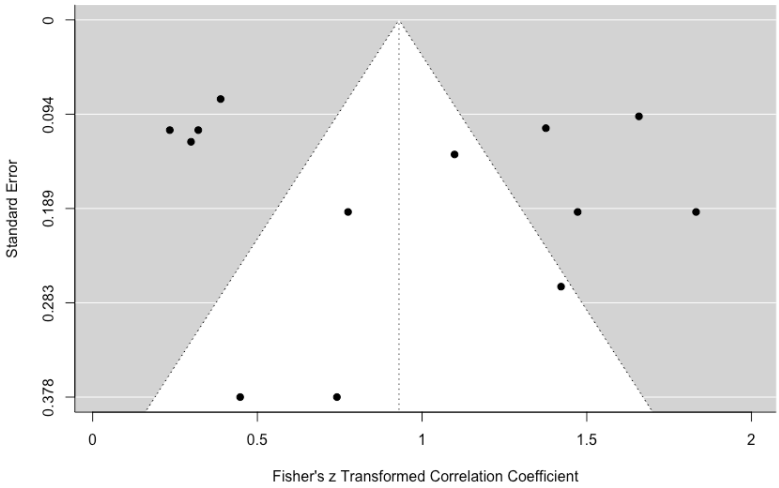

Figure 15. Funnell Plot for assessing publication bias of LPFS Domain – Intimacy

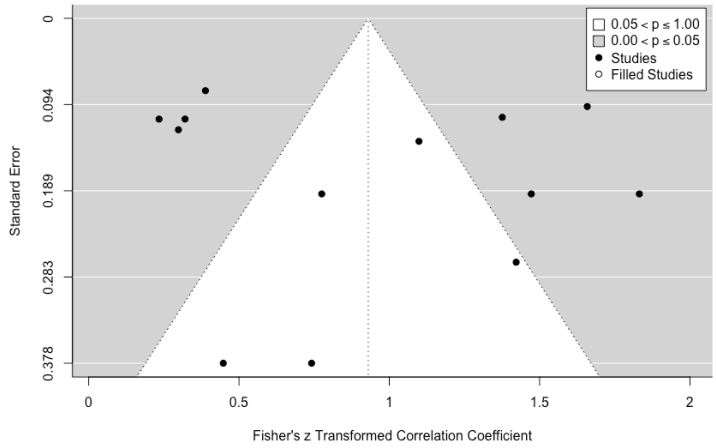

Figure 16. Funnell Plot for assessing publication bias of LPFS Domain – Intimacy - Trim and Fill analysis

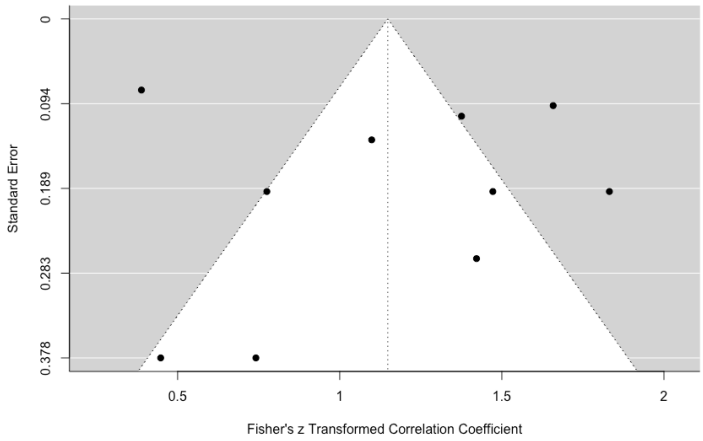

Figure 17. Funnell Plot for assessing publication bias of LPFS Domain – Intimacy – Sensitivity Analysis

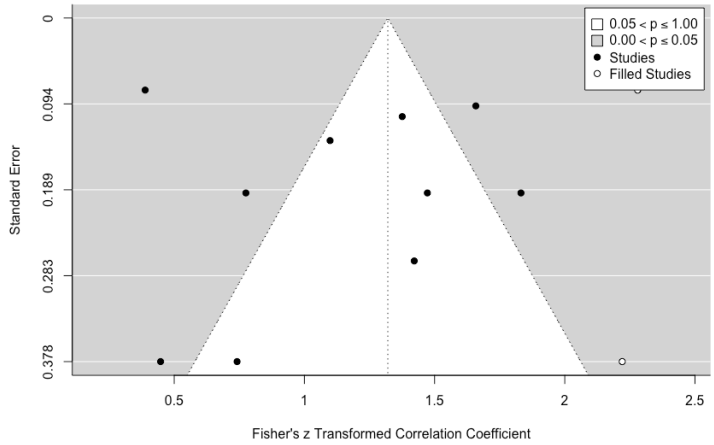

Figure 18. Funnell Plot for assessing publication bias of LPFS Domain – Intimacy – Sensitivity Analysis - Trim and Fill analysis

Appendix G. Manuscript Guidelines

Psychiatry, Psychology and Law

Preparing Your Paper
All authors submitting to medicine, biomedicine, health sciences, allied and public health journals should conform to the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, prepared by the International Committee of Medical Journal Editors (ICMJE).

Structure
1) Main document with author details: Your paper should be compiled in the following order: title page; abstract; keywords; main text (introduction, materials and methods, results, discussion); acknowledgments; disclosure and ethical standards statement; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figures; figure captions (as a list). Please label this file 'Main document – with full author details'. A separate title page may also be uploaded if desired, labelled 'Title page (not for review)'.

2) Anonymised manuscript: Please also upload an anonymised version of your manuscript with a title page but with no identifying author information in the title page or body of the manuscript. Please label this file 'Main document – Anonymous'.

3) Tables and figures: Please add any tables or figures as separate documents. Please label these file as 'Tables' and/or 'Figures' as appropriate.

Word Limits
Please include a word count for your paper.
A typical paper for this journal should be no more than 12000 words, inclusive of tables, references, figure captions.

Style Guidelines
Please refer to these quick style guidelines when preparing your paper, rather than any published articles or a sample copy.
Please use British (-ise) spelling style consistently throughout your manuscript.
Any form of consistent quotation style is acceptable. Please note that long quotations should be indented without quotation marks.

Manuscripts should be prepared depending on whether they are psychological, psychiatric, or legal in nature:
Psychological manuscripts should be prepared in accordance with the format and style specified in the 'Publication Manual of the American Psychological Association', fifth edition. Pages should be numbered consecutively. References should be cited in the text as specified in the Publication Manual of the American Psychological Association, seventh or current edition. A concise description of APA referencing style can be found at https://www.tandf.co.uk//journals/authors/style/reference/tf_APA.pdf . Personal communications should be cited as such in the text and should not be included in the reference list. For an overview of APA style (including referencing) visit https://apastyle.apa.org/

Formatting and Templates

Papers may be submitted in Word format. Figures should be saved separately from the text.
To assist you in preparing your paper, we provide formatting template(s).
Word templates are available for this journal. Please save the template to your hard drive, ready for use.
If you are not able to use the template via the links (or if you have any other template queries) please contact us here.

References

Please use this reference guide when preparing your paper.

Taylor & Francis Editing Services

To help you improve your manuscript and prepare it for submission, Taylor & Francis provides a range of editing services. Choose from options such as English Language Editing, which will ensure that your article is free of spelling and grammar errors, Translation, and Artwork Preparation. For more information, including pricing, visit this website.

Checklist: What to Include

1. Author details. Please ensure everyone meeting the International Committee of Medical Journal Editors (ICMJE) requirements for authorship is included as an author of your paper. Please ensure all listed authors meet the Taylor & Francis authorship criteria. All authors of a manuscript should include their full name and affiliation on the cover page of the manuscript. Where available, please also include ORCiDs and social media handles (Facebook, Twitter or LinkedIn). One author will need to be identified as the corresponding author, with their email address normally displayed in the article PDF (depending on the journal) and the online article. Authors' affiliations are the affiliations where the research was conducted. If any of the named co-authors moves affiliation during the peer-review process, the new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after your paper is accepted. Read more on authorship.
2. Should contain an unstructured abstract of 150 words.
3. Graphical abstract (optional). This is an image to give readers a clear idea of the content of your article. It should be a maximum width of 525 pixels. If your image is narrower than 525 pixels, please place it on a white background 525 pixels wide to ensure the dimensions are maintained. Save the graphical abstract as a .jpg, .png, or .tiff. Please do not embed it in the manuscript file but save it as a separate file, labelled GraphicalAbstract1.
4. You can opt to include a video abstract with your article. Find out how these can help your work reach a wider audience, and what to think about when filming.
5. At least 10 keywords. Read making your article more discoverable, including information on choosing a title and search engine optimization.
6. Funding details. Please supply all details required by your funding and grant-awarding bodies as follows:

For single agency grants…

This work was supported by the [Funding Agency] under Grant [number xxxx].

For multiple agency grants…

This work was supported by the [Funding Agency #1] under Grant [number xxxx]; [Funding Agency #2] under Grant [number xxxx]; and [Funding Agency #3] under Grant [number xxxx].

7. Disclosure statement. This is to acknowledge any financial or non-financial interest that has arisen from the direct applications of your research. If there are no relevant competing interests to declare please state this within the article, for example: The authors report there are no competing interests to declare. Further guidance on what is a conflict of interest and how to disclose it.
8. Supplemental online material. Supplemental material can be a video, dataset, fileset, sound file or anything which supports (and is pertinent to) your paper. We publish supplemental material online via Figshare. Find out more about supplemental material and how to submit it with your article.
9. Figures. Figures should be high quality (1200 dpi for line art, 600 dpi for grayscale and 300 dpi for colour, at the correct size). Figures should be supplied in one of our preferred file formats: EPS, PS, JPEG, TIFF, or Microsoft Word (DOC or DOCX) files are acceptable for figures that have been drawn in Word. For information relating to other file types, please consult our Submission of electronic artwork document.
10. Tables. Tables should present new information rather than duplicating what is in the text. Readers should be able to interpret the table without reference to the text. Please supply editable files.
11. Equations. If you are submitting your manuscript as a Word document, please ensure that equations are editable. More information about mathematical symbols and equations.
12. Units. Please use SI units (non-italicized).

Using Third-Party Material in your Paper

You must obtain the necessary permission to reuse third-party material in your article. The use of short extracts of text and some other types of material is usually permitted, on a limited basis, for the purposes of criticism and review without securing formal permission. If you wish to include any material in your paper for which you do not hold copyright, and which is not covered by this informal agreement, you will need to obtain written permission from the copyright owner prior to submission. More information on requesting permission to reproduce work(s) under copyright.

Disclosure and Ethical Standards Statement

The journal now has a requirement for all manuscripts that are submitted to it to include a statement to confirm compliance with ethical standards and ethical approval. This requirement has been introduced to ensure that the journal complies with the minimal requirements for Medline indexing. All manuscripts submitted to the journal must include (in the 'main document with author details') a statement in the manuscript as outlined below in the exact form of either Option 1 or Option 2 (below), as well as including relevant information about ethics and informed consent in the Methods section (see 'Complying with Ethics of Experimentation') below. Manuscripts that do not include this statement, will not be considered for publication in the journal. Please include the appropriate statement (choosing option 1 or 2) in your 'main document_with author details'.

Disclosure and Ethical Standards Statement Option 1: Studies with no human participants

Declaration of conflicts of interest

Author A [add name of author here] has declared no conflicts of interest
Author B [add name of author here] has declared no conflicts of interest
Author C [add name of author here] has declared no conflicts of interest

Ethical approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Disclosure and Ethical Standards Statement Option 2: Studies with human participants

Declaration of conflicts of interest
Author A [add name of author here] has declared no conflicts of interest
Author B [add name of author here] has declared no conflicts of interest
Author C [add name of author here] has declared no conflicts of interest

Ethical approval
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee [insert as appropriate, including name of approving committee and any approval numbers] and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent
Informed consent was obtained from all individual participants included in the study

Disclosure of Benefit or Interest Statement
Authors are required to disclose and acknowledge any financial benefit or interest that has arisen from the direct applications of your research. If you have benefits or interests to declare, this must be included in the disclosure and ethical standards statement. If you have no interests to declare, please state this using the wording in the disclosure and ethical standards statement. For all NIH/Wellcome-funded papers, the grant number(s) must be included in the declaration of interest statement. Read more on declaring conflicts of interest.

Clinical Trials Registry
In order to be published in a Taylor & Francis journal, all clinical trials must have been registered in a public repository at the beginning of the research process (prior to patient enrolment). Trial registration numbers should be included in the abstract, with full details in the methods section. The registry should be publicly accessible (at no charge), open to all prospective registrants, and managed by a not-for-profit organization. For a list of registries that meet these requirements, please visit the WHO International Clinical Trials Registry Platform (ICTRP). The registration of all clinical trials facilitates the sharing of information among clinicians, researchers, and patients, enhances public confidence in research, and is in accordance with the ICMJE guidelines.

Complying With Ethics of Experimentation
Please ensure that all research reported in submitted papers has been conducted in an ethical and responsible manner, and is in full compliance with all relevant codes of experimentation and legislation. All papers which report in vivo experiments or clinical trials on humans or animals must include a written statement in the Methods section. This should explain that all work was conducted with the formal approval of the local human subject or animal care committees (institutional and national), and that clinical trials have been registered as legislation requires. Authors who do not have formal ethics review committees should include a statement that their study follows the principles of the Declaration of Helsinki.

Consent

All authors are required to follow the ICMJE requirements on privacy and informed consent from patients and study participants. Please confirm that any patient, service user, or participant (or that person's parent or legal guardian) in any research, experiment, or clinical trial described in your paper has given written consent to the inclusion of material pertaining to themselves, that they acknowledge that they cannot be identified via the paper; and that you have fully anonymized them. Where someone is deceased, please ensure you have written consent from the family or estate. Authors may use this Patient Consent Form, which should be completed, saved, and sent to the journal if requested.

Health and Safety

Please confirm that all mandatory laboratory health and safety procedures have been complied with in the course of conducting any experimental work reported in your paper.
Please ensure your paper contains all appropriate warnings on any hazards that may be involved in carrying out the experiments or procedures you have described, or that may be involved in instructions, materials, or formulae.
Please include all relevant safety precautions; and cite any accepted standard or code of practice. Authors working in animal science may find it useful to consult the International Association of Veterinary Editors' Consensus Author Guidelines on Animal Ethics and Welfare and Guidelines for the Treatment of Animals in Behavioural Research and Teaching.
When a product has not yet been approved by an appropriate regulatory body for the use described in your paper, please specify this, or that the product is still investigational.

Submitting Your Paper

This journal uses ScholarOne Manuscripts to manage the peer-review process. If you haven't submitted a paper to this journal before, you will need to create an account in ScholarOne. Please read the guidelines above and then submit your paper in the relevant Author Centre, where you will find user guides and a helpdesk.
Please note that Psychiatry, Psychology and Law uses Crossref™ to screen papers for unoriginal material. By submitting your paper to Psychiatry, Psychology and Law you are agreeing to originality checks during the peer-review and production processes.
On acceptance, we recommend that you keep a copy of your Accepted Manuscript. Find out more about sharing your work.

Publication Charges

There are no submission fees, publication fees or page charges for this journal.
Colour figures will be reproduced in colour in your online article free of charge. If it is necessary for the figures to be reproduced in colour in the print version, a charge will apply.

Copyright Options

Copyright allows you to protect your original material, and stop others from using your work without your permission. Taylor & Francis offers a number of different license and reuse options, including Creative Commons licenses when publishing open access. Read more on publishing agreements.

Complying with Funding Agencies

We will deposit all National Institutes of Health or Wellcome Trust-funded papers into PubMedCentral on behalf of authors, meeting the requirements of their respective open access policies. If this applies to you, please tell our production team when you receive your article proofs, so we can do this for you. Check funders' open access policy mandates here. Find out more about sharing your work.

My Authored Works
On publication, you will be able to view, download and check your article's metrics (downloads, citations and Altmetric data) via My Authored Works on Taylor & Francis Online. This is where you can access every article you have published with us, as well as your free eprints link, so you can quickly and easily share your work with friends and colleagues.
We are committed to promoting and increasing the visibility of your article. Here are some tips and ideas on how you can work with us to promote your research.

Queries
Should you have any queries, please visit our Author Services website.

Appendix H: Six-point *Vowles* Likert Scale Measuring Agreement with Each Criteria.

Participants were asked to rate each of the four criteria on a six-point Likert scale of 1-6 (strongly disagree - 1, moderately disagree - 2, somewhat disagree - 3, somewhat agree - 4, moderately agree - 5, strongly agree - 6)

1) The extent to which the offender's mental health requires treatment
2) The extent to which offending is attributable to the mental health disorder
3) The extent to which offending requires punishment
4) The protection of the public when deciding release and regime of release

Appendix I: Knowledge Check

1. Please confirm the defendant's name.
    a. John Smith
    b. David Hughes
    c. Brian Barnes
2. What offence has the defendant been found guilty of?
    a. Grievous Bodily Harm
    b. Fraud
    c. Burglary

Appendix J: Ethical Approval from Judicial Office

From: Judicial Research Requests <researchrequest@judiciary.uk>
Sent: 24 November 2021 15:10
To: Samantha Young (MED - Postgraduate Researcher) <Samantha.Young@uea.ac.uk>
Cc: George Baldwin (MED - Postgraduate Researcher) <G.Baldwin@uea.ac.uk>; Judicial Research Requests <researchrequest@judiciary.uk>
Subject: RE: Sponsorship request for research within the HMCTS

Dear Samantha and George,

Thank you for your patience. The senior judiciary has now reviewed your application and has given permission for you to conduct your research involving the judiciary as set out in your application.

I understand you would like our support in recruiting participants. Can I ask how many and exactly what type of judges you need assistance with recruiting?

I would take this opportunity to remind you that judges should not be asked to comment on matters of government policy or on individual cases.

We wish you the best with your research and would be very grateful for a copy of your findings once they are finalised.

Best wishes,
Aliza

Aliza Catlin
Deputy Head of Jurisdictional Strategy

Judicial Office
11th Floor, Thomas More Building, Royal Courts of Justice
www.judiciary.uk

Appendix K: Advertisement

Appendix L: Lay Summary

This research study is looking at factors that impact Crown Court sentences for offenders with mental health conditions. Previous research has shown almost a quarter of prisoners arrive having had previous contact with mental health services. So we want to understand how reliably the court decides if a guilty offender goes to hospital or prison. This study will help us understand the process more. We will use a scale to measure your beliefs about the causes of mental health to see how certain beliefs may influence the sentences given. We also want to see what role mental health diagnosis has in deciding what sentence is given so you will be randomly allocated a diagnosis. We hope to see if there is any difference in how sentencing criteria is rated as well as the final sentencing verdict.

The study will present you with a mock court case video and then your mental health beliefs will be measured. We will also measure your ratings for each part of the sentencing guidance and then ask for your decision on the final sentence. This will help us to understand factors affecting this decision making process.

We estimate around 25,000 people with previous mental health problems receive prison sentences every year. We want to make sure the decision between going to hospital, prison or a combination of the two is reliable. This study could help to inform us how reliable these decisions are currently.

Appendix M: Participant Information Sheet

The participant information sheet was attached to the online survey as a PDF document. All participants had to confirm they had read this document and wished to participate.

PARTICIPANT INFORMATION STATEMENT

(1) What is this study about?
You are invited to take part in this study looking into factors influencing sentencing for offenders with mental health problems. We are recruiting participants from the law population, to understand the decision-making process involved in sentencing during Crown Court cases.

This Information Statement outlines the study to help you decide whether you would like to take part, please read it carefully and raise any questions you may have. Your participation is voluntary and you retain the right to withdraw at any point.

By giving consent to take part in this study you are telling us that you:
- Understand what you have read.
- Agree to take part in the research study as outlined below.
- Agree to the use of your personal information as described.
- You have received a copy of this Participant Information Statement to keep.

(2) Who is running the study?
This study is being conducted by: George Baldwin and Samantha Young, ClinPsyD Researchers, Norwich Medical School, University of East Anglia.

(3) What will the study involve for me?
Your participation requires completion of an online survey, which has multiple sections and will take approximately 30 minutes. You will be provided information about sentencing options, followed by case material. You will then be asked to complete two questionnaires measuring your beliefs about the causes of mental health and then asked to give a sentencing verdict.

(4) How much of my time will the study take?
The survey will take approximately 30 minutes to complete.

(5) Do I have to be in the study? Can I withdraw from the study once I've started?
Participation is voluntary, your decision whether to participate will not affect current or future relationships with anyone associated with the University of East Anglia. You can withdraw from the study prior to completion. Once you have started the survey, you will need to contact us to request that your data is not be saved.

(6) Are there any risks or costs associated with being in the study?

This study is not expected to cause any distress, however you are advised to stop completing the survey if at any time you feel uncomfortable. If you complete the survey and then experience distress, please contact us by email (g.baldwin@uea.ac.uk / samantha.young@uea.ac.uk) to discuss issues of concern and signpost you to further support if needed. You can also contact your GP for mental health support. Samaritans offer a 24/7 listening service via 116 123.

(7) Are there any benefits associated with being in the study?
This study will hopefully provide insight into factors influencing sentencing of offenders with mental health difficulties, to inform real life processes and safeguard from unreliable and/or biased real-life sentencing.

(8) What will happen to information about me that is collected during the study?
By consenting to participate, you are agreeing to the personal information shared to be collected and used for the purpose of this research study. Any information provided will only be used for the purposes outlined in this Participant Information Statement unless you consent otherwise. The 2018 General Data Protection Regulation Act and the University of East Anglia Research Data Management Policy (2019) will be adhered to at all times. Your information will be stored securely using UEA cloud storage and your identity/information will only be disclosed with your permission, except as required by law. Findings from this study may be included in publication, but you will not be identifiable. Data will be stored until analysis and publication are completed and then retained for ten years.

(9) What if I would like further information about the study?
When you have read this information, we will be available to discuss it with you further and answer any questions you may have. You can contact us via g.baldwin@uea.ac.uk / samantha.young@uea.ac.uk).

(10) Will I be told the results of the study?
You have a right to receive feedback about the overall results of this study. You can request this by contacting us via g.baldwin@uea.ac.uk / samantha.young@uea.ac.uk). Overall results will be provided in the form of a one page lay summary which you will receive after the study is finished.

(11) What if I have a complaint or any concerns about the study?
The ethical aspects of this study have been approved under the regulations of the University of East Anglia's Faculty of Medicine and Health Sciences Ethics Committee.

If there is a problem please let us know. You can contact us via the University at the following address:
George Baldwin & Samantha Young
Norwich Medical School
Faculty of Medicine and Health Sciences
University of East Anglia
NORWICH NR4 7TJ

(g.baldwin@uea.ac.uk / samantha.young@uea.ac.uk)

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the administration team who will direct your concerns to a senior faculty member: med.reception@uea.ac.uk

(12) OK, I want to take part – what do I do next?

You need to return to the online survey and click to confirm you have read this form and wish to participate.

Appendix N: Consent Form

The consent form below was included as an item within the online survey, following the participant information sheet. Participants could not continue the survey without confirming consent.

By acknowledging that I have read this consent form and clicking to proceed with the online survey, I agree to take part in this research study.

In giving my consent I state that:

- I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.
- I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.
- The researchers have answered any questions that I had about the study and I am happy with the answers.
- I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of East Anglia now or in the future.
- I understand that I can withdraw from the study at any time.
- I understand that I may stop the interview at any time if I do not wish to continue, and that unless I indicate otherwise any recordings will then be erased and the information provided will not be included in the study. I also understand that I may refuse to answer any questions I don't wish to answer.
- I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.
- I understand that the results of this study may be published, but these publications will not contain my name or any identifiable information about me.

Appendix O: Video Transcript

The same script was used for all three conditions with only the diagnosis changing (schizophrenia/BPD/complex mental health). The interchangeable diagnostic labels are underlined and in bold.

Your honour, the defendant, Mr James Smith, DOB: 4/10/99, has pleaded guilty to committing the offence of unlawfully and maliciously causing grievous bodily harm with intent to cause grievous bodily harm, contrary to section 18 of the Offences Against the Person Act 1861. He attacked the victim, Robert Peterson, with a weapon causing grievous bodily harm.

The facts of the case are as follows. On the 13th December 2020, the victim and defendant were seen arguing on the corner of London Road. The victim was Mr Smith's site manager at Lions Construction, where the defendant had worked as a labourer. Whilst working at the construction site Mr Smith had been given numerous warnings for repeatedly turning up to work late, failing to follow instructions and frequently getting into arguments with other site workers. The victim had approached Mr Smith after he arrived an hour late for work, to inform him he was no longer required and instructed him to leave the premises. The victim testified that Mr Smith was angry and aggressive, swearing at him and storming off. The following day, the victim reported encountering Mr Smith near the construction site on the corner of London Road. Mr Smith waited for the victim to finish work, where he was on his own, then entered the site, blocking the victim's exit. The victim reported Mr Smith to be loud and aggressive and difficult to follow, talking quickly and incoherently about his job. When the victim asked Mr Smith to leave, Mr Smith grabbed a steel scaffold pole from the floor and immediately struck the victim five times, including once to the head, causing permanent facial disfigurement and resulting in the victim being unable to work for 3 months.

Mr Smith was arrested at the scene after a witness from the adjacent construction site alerted the police to the incident. The victim's personal statement states "The actions of Mr Smith have completely changed my life. I spent over three months in hospital and despite numerous surgeries, I still see the damage caused by Mr Smith's attack every time I look in the mirror. Since the attack, I have been unable to return to work which has also meant that I am struggling financially. I am no longer the confident and care-free man I was."

Your Honour, as the judge presiding over this case, it is your job to determine Mr Smith sentence.

For the purposes of sentencing, I present as evidence the report of Dr Robert Taylor, a psychiatrist instructed to interview the defendant and report on the defendant's mental health condition in relation to the offence. His expert opinion has been corroborated by a second opinion from psychiatrist, Dr Amanda Bell. As this report confirms, Dr Taylor states that the defendant suffers from [**Borderline Personality Disorder which is a recognised medical condition or Schizophrenia which is a recognised medical condition or a complex mental health problem]**. Dr Taylor notes as part of this condition, unstable emotions (rapidly changing from being calm to angry), paranoid thoughts (expecting others to harm him), auditory hallucinations (hearing voices) and impulsive behaviours are present. Dr Taylor notes in the report that it is not uncommon for these symptoms to be worsened by stressful life events, such as a job loss. Indeed, during his childhood, Mr Smith attended a number of different schools. He described initially moving schools due to experiencing bullying from an early age as he would often turn up to school with worn and dirty clothing. However, later, Mr Smith begun to present with challenging behaviours which resulted in him being suspended and expelled from a number of schools, eventually leading him to be placed in a pupil referral unit for his challenging behaviour. Despite this history, Mr Smith does not have any previous convictions. He states remorse for the incident for which he has

plead guilty, but has also insisted that the victim firing him was provocation. Mr Smith has found it difficult to find stable employment, he has had approximately seven jobs in the last year, with many roles ending due to disputes or poor attendance. In 2019, Mr Smith's Employment Advisor at the Job Centre attempted to refer him to mental health services due to some odd behaviour, rushed speech and paranoid beliefs (including believing that previous colleagues had plotted against him) that were shared in an appointment. Mr Smith was diagnosed with [**Borderline Personality Disorder or Schizophrenia or having a complex mental health problem]** during his mental health assessment, however he subsequently disengaged with treatment offered and was discharged from the service. Aside from this, however, Mr Smith has had no other contact with mental health services.

Although Dr Taylor was certain that Mr Smith's diagnosed mental health condition would have played a part in the offence, it is difficult to know whether his mental health condition can fully explain his behaviour on the day in question. Certainly, his history suggests that his dismissal may well have led Mr Smith to experience extreme emotions of anger. Further, both experts have suggested that preoccupation with mental health symptoms earlier in the day, may have led him to be late for work and linked to his fears – or paranoia – that some of the workers at the site wanted to 'do him in'. It is perhaps even possible that these fears influenced his reaction to his boss dismissing him. However, both experts found it difficult to extract more detailed information from Mr Smith on this point and there is significant uncertainty. However, as stated, both experts have agreed that Mr Smith's presentation is consistent with **[a diagnosis of Borderline Personality Disorder or Schizophrenia or having a complex mental health problem]**. Both experts agree that Mr Smith could benefit from a period of treatment within a hospital environment. Therefore, as the honourable judge presiding over this case, it is down to you to determine Mr Smith's sentencing.

Appendix P: Sentencing Options with Descriptive Text

Participants were asked to select one sentence from the options below, with the descriptive text presented alongside each option:

1. Section 37/41 Hospital Order: the offender would go to a secure hospital and receive mental health treatment until deemed well enough to be discharged by a Mental Health Tribunal. They would then be monitored in the community by forensic mental health services.
2. Prison: the offender would go to prison until either the expiry of their sentence or they become eligible for parole. After release, they would be monitored in the community by probation services.
3. Section 45A Hospital Order: the offender would go to a secure hospital and receive mental health treatment, however there would be a minimum sentence attached, meaning if the Mental Health Tribunal felt the offender no longer needed to be in hospital, they could be transferred to prison for the remainder of their sentence until eligible for parole. If released from hospital, they would be monitored by forensic mental health services, if released from prison, they would be monitored by probation services.

Appendix Q: Confirmation Letter from the UEA FMH Research Ethics Committee

Faculty of Medicine and Health Sciences Research Ethics Committee

University of East Anglia

NORWICH MEDICAL SCHOOL
Bob Champion Research & Educational Building
Rosalind Franklin Road
University of East Anglia
Norwich Research Park
Norwich NR4 7UQ

Email: fmh.ethics@uea.ac.uk
www.med.uea.ac.uk

Samantha Young and George Baldwin
Norwich Medical School
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

23rd April 2021

Dear Samantha and George

Project Title: The sentencing of Mentally Disordered Offenders: an experimental clinical vignette study to explore the reliability of judgements made using Vowles and the relationships between psychiatric diagnosis, mental health locus of control, mental illness stigma and final sentencing outcome.

Reference: 2020/21-067

Thank you for your email of 22nd April 2021 notifying us of the amendments you would like to make to your above proposal. These have been considered and I can confirm that your amendments have been approved.

Please can you ensure that any further amendments to either the protocol or documents submitted are notified to us in advance, and that any adverse events which occur during your project are reported to the Committee.

Approval by the FMH Research Ethics Committee should not be taken as evidence that your study is compliant with GDPR and the Data Protection Act 2018. If you need guidance on how to make your study GDPR compliant, please contact your institution's Data Protection Officer.

Please can you arrange to send us a report once your project is completed.

Yours sincerely

Dr Jackie Buck
Chair
FMH Research Ethics Committee

Appendix R: Debrief

The debrief sheet was attached to the end of the online survey as a PDF document. All participants had to confirm they had read this document before submitting the online survey.

DEBRIEF

Thank you for taking part in this study looking into factors influencing sentencing for offenders with mental health problems. If you wish for your data to be removed or you are experiencing any distress following the survey, please contact us by email (g.baldwin@uea.ac.uk / samantha.young@uea.ac.uk) or telephone (01603 592308) to discuss issues of concern and signpost you to further support if needed.

You can also contact us to request a lay summary of our findings via the University at the following address:

George Baldwin & Samantha Young
Norwich Medical School
Faculty of Medicine and Health Sciences
University of East Anglia
NORWICH NR4 7TJ

(g.baldwin@uea.ac.uk / samantha.young@uea.ac.uk)

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the administration team who will direct your concerns to a senior faculty member: med.reception@uea.ac.uk

Kind regards,

George Baldwin & Samantha Young

Appendix S: Preliminary Analyses

Table 1.

Descriptive data Total PDD Score and Sex

					95% Confidence Interval of the Mean			
	N	Mean	SD	SE	Lower	Upper	Min	Max
Male	45	45.44	10.55	1.57	42.27	48.61	27.00	71.00
Female	152	48.33	10.26	0.83	46.68	49.94	21.00	71.00
Total	197	47. 67	10.37	0.74	46.21	49.13	21.00	71.00

Table 2.

Independent Samples T-Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. 2-tailed	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.01	.91	-1.65	195	.101	-2.88	1.75	-6.34	0.57
Equal variances not assumed			-1.62	18.96	.110	-2.88	1.78	-6.43	0.66

Table 3

Pearson Correlations – PDD Scores

	N	Pearson Correlation	Sig. (2-tailed)
Age*Total PDD Score	197	.34	.64
Gender*Total PDD Score	197	.11	.10
Diagnostic condition*Total PDD Score	197	.13	.75
Vowles Criteria 1*Total PDD Score	197	.11	.13
Vowles Criteria 2*Total PDD Score	197	.06	.40
Vowles Criteria 3*Total PDD Score	197	.01	.88
Vowles Criteria 4*Total PDD Score	197	.10	.18

Table 4.

Descriptive data Total PDD Score and Diagnosis

					95% Confidence Interval of the Mean		
	N	Mean	SD	SE	Lower	Upper	Min	Max
Complex Mental Health	60	46.35	10.39	1.34	43.67	49.03	27.00	71.00
BPD	68	49.44	10.16	1.23	46.98	51.90	21.00	70.00
Schizophrenia	69	47.07	10.47	1.26	44.56	49.59	26.00	71.00
Total	197	47. 67	10.37	0.74	46.21	49.13	21.00	71.00

Table 5

One-way ANOVA examining whether psychiatric diagnosis is associated with difference in Total PDD Score (N = 197)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	342.50	2	171.25	1.60	.20
Within Groups	20751.05	194	106.96		
Total	21093.55	196			