EXPLORING CLINICIANS' EXPERIENCES OF SHARED DECISION-MAKING IN THE TREATMENT OF MULTIPLE SCLEROSIS

by

Rebecca Gordon

A THESIS SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM FOR THE DEGREE OF DOCTOR OF CLINICAL PSYCHOLOGY

VOLUME ONE

Department of Clinical Psychology

School of Psychology

The University of Birmingham

June 2020

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Thesis Overview

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Clinical Psychology at the University of Birmingham. The thesis comprises of two volumes. All identifying information has been anonymised for confidentiality.

Volume One

This volume comprises three chapters. The first chapter is a meta-analysis of the diagnostic accuracy of the Word Memory Test in detecting dissimulation within analogue studies. The second chapter is a qualitative study that investigates the experience of clinicians participating in shared decision-making with patients with Multiple Sclerosis. The third chapter is a press release document, providing an overview of the meta-analysis and qualitative study.

Volume Two

This volume comprises five clinical practice reports (CPRs). The first report presents the case of a 14-year-old female presenting with anxiety and intense threat beliefs formulated from a cognitive behavioural and psychodynamic perspective. The second report presents a service evaluation exploring staff perspectives of client access to psychological therapies within mental health services for children and young people. The third report presents a case study of a 53-year-old female with Obsessive-Compulsive Disorder, using cognitive behavioural and systemic models. The fourth report is a single-case experimental design, evaluating the effectiveness of a behavioural intervention with a 70-year-old male with a Learning Disability. The final report presents an abstract of an oral presentation of a cognitive-behavioural intervention with a 72-year-old female presenting with anxiety, depression and suicidal ideation.

Acknowledgements

I would like to thank all the participants who gave up their time to take part in this research and share their experiences with me. To all of the clients that I have worked with both before and throughout my clinical training, every one of you has taught me something different. It has been a privilege to be even a small part of your lives.

Thank you to Dr Chris Jones and Dr Ruth Howard for supporting me throughout development (and many drafts) of this thesis. Along with my appraisal tutor, Dr Michelle Fisher your support, patience and understanding has been invaluable in reaching this point. Thank you also to Dr David Hacker and Dr John Woolmore (NHS supervisors) for your substantial contributions.

I would also like to thank my clinical supervisors: Dr Sarah Flanagan, Dr Erik Evenson, Dr Yajnah Bheenick, Dr Sue Adams and Dr Amy Naylor – for sharing your knowledge and time and supporting me to develop both personally and professionally. I have learnt so much from each of you and will take this with me in my future career.

To the friends I have made on this course – Becca, Caroline, Emily, Hollie, Jack, Joe, Nat and Leila – thank you for the laughter, the hugs (and the cake). The world is gaining eight incredible Clinical Psychologists and I will always be grateful to call you my friends!

Finally, and maybe most importantly, to my family and friends. Thank you for understanding when I needed to go “off grid” (and still being there when I returned). To Mum and Dad, I would never have made it this far without your love, support and encouragement – I hope I have made you proud. To Dan, thank you for your unwavering love and support from day one, for looking after me when I forget to take care of myself and for always being there. Here’s to the next adventure…

Table of Contents

a) Volume One: Research Component

i) Literature Review: A Meta-Analysis of Analogue Studies of Diagnostic Accuracy of the Word Memory Test 1

Abstract 2

Introduction 3

Method 11

Meta-Analytic Syntheses 32

Discussion 51

References 59

ii) Empirical Paper: Exploring clinician experiences of Shared Decision-Making in Multiple Sclerosis: An Interpretative Phenomenological Analysis 66

Abstract 67

Introduction 68

Method 76

Results 85

Discussion 105

References 112

iii) Literature Review Press Release 121

Spotting those we ask to exaggerate 122

iv) Empirical Paper Press Release 124

“Just get on with it” – experiences of doctors as they navigate Shared Decision-Making 125

b) Appendices for Volume One

Appendix 1: Search Terms by Database 128

Appendix 2: Data Extracted from Primary Studies 130

Appendix 3: Criteria defining the level of sophistication/coaching of simulating participants 131

Appendix 4: Univariate Analyses for Immediate Recall, Delayed Recall and Consistency 132

Appendix 5: sROC Curves for Immediate Recall, Delayed Recall and Consistency 135

Appendix 6: University of Birmingham Sponsorship Letter 137

Appendix 7: Health Research Authority Approval Letter 138

Appendix 8: Research and Development No Objection Letter 141

Appendix 9: Participant Information Sheet 143

Appendix 10: Participant Consent Form 145

Appendix 11: Interview Schedule 146

Appendix 12: Step 2 of analysis (Initial coding) 147

Appendix 13: Step 3 of analysis (Emergent themes) 148

Appendix 14: Step 5 of analysis (Connections between participants) 149

c) Volume Two: Clinical Practice Reports Component

i) Psychological Models Clinical Practice Report: The assessment and formulation of Zainab* from Cognitive-Behavioural and Psychodynamic Perspectives. 1

Abstract 2

Case Background and Assessment 4

Formulation 14

Reflections 28

References 31

ii) Service Evaluation Clinical Practice Report: Exploring staff perceptions of client access to psychological therapies within mental health services for children and young people. 35

Abstract 36

Introduction 37

Aims 44

Method 45

Results 49

Discussion 64

References 71

iii) Integrated Case Study Clinical Practice Report: An integrated case-study report of Sue* 77

Abstract 78

Case Background 80

Assessment 82

Formulation 93

Intervention 104

Evaluation 113

Reflections 115

References 117

iv) Single Case Experimental Design Clinical Practice Report: A Single-Case Experimental Design with Tom* 124

Abstract 125

Case Background and Assessment 127

Formulation 137

Intervention 144
Method 148
Results 150
Discussion 155
References 159
v) Case Study Clinical Practice Report: A case study with Sheila*. 167
Abstract 168

d) Appendices for Volume Two

Appendix 1: Service Evaluation NRES Guidance Table 171
Appendix 2: Service Evaluation Staff Questionnaire 172
Appendix 3: Service Evaluation Team Manager Interview Guide 175
Appendix 4: Sue's diary of Intrusive Thoughts and Compulsive Behaviours 176
Appendix 5: Sue's Behavioural Experiment Record Sheet 178
Appendix 6: Tom's Mood Diary 179
Appendix 7: Example ABC Diary for Staff to complete 180

List of Illustrations

a) Volume One: Research Component

i) Literature Review:

Figure 1 PRISMA chart summarising systematic search results and application of inclusion criteria 14
Figure 2 Forest plot of WMT sensitivity for Overall index by cut-off 37
Figure 3 Forest plot of WMT specificity for Overall index by cut-off 38
Figure 4 Forest plot of log-transformed Positive Likelihood Ratios 40
Figure 5 Forest plot of log-transformed Negative Likelihood Ratios 41
Figure 6 Forest plot of log-transformed Diagnostic Odds Ratio 43
Figure 7 Forest plot of WMT specificity for Overall index by control population 44
Figure 8 Subgroup analysis of DOR and confidence intervals for Naïve and Coached simulators from studies reporting failure at standard cut-offs 45
Figure 9 Subgroup analysis of DOR and confidence intervals for Incentivised and Non-incentivised simulators from studies reporting failure at standard cut-offs 46
Figure 10 sROC plot of sensitivity and false positive rates for studies reporting Overall index failure at standard cut-off 48
Figure 11 Subgroup analysis of DOR and confidence intervals for WMT sub-tests 50

b) Volume Two: Clinical Practice Reports Component

i) Psychological Models Clinical Practice Report

Figure 1 Zainab's relationship map based on assessment information 11
Figure 2 Longitudinal Cognitive-Behavioural formulation based upon Beck (1976) 18
Figure 3 Five areas maintenance formulation based upon Padesky & Mooney (1990) 20
Figure 4 Five areas maintenance formulation of Zainab's concentration difficulties 21
Figure 5 Triangle of Conflict based upon Malan (1995) 25
Figure 6 Triangle of Person based upon Malan (1995) 27

ii) Service Evaluation Clinical Practice Report

Figure 7 The CAPA model of service delivery (York & Kingsbury, 2009) 44

Figure 8 Percentage of staff referring to therapeutic interventions by type 50
Figure 9 Staff agreement of awareness of available therapeutic interventions 51
Figure 10 Staff agreement relating to policies for access to therapeutic interventions 52
Figure 11 How staff determine type of therapeutic intervention to refer for 54
Figure 12 Barriers to accessing therapeutic intervention 57

iii) Integrated Case Study Clinical Practice Report

Figure 13 Timeline representation of Sue's pathway within mental health services 80
Figure 14 Genogram of Sue's family 85
Figure 15 Timeline of Sue's significant life experiences 86
Figure 16 Longitudinal formulation of Sue's OCD based upon Salkovskis et al. (1998) 96
Figure 17 Sue's example maintenance cycle 98
Figure 18 Example feedback loop 100
Figure 19 Example of Sue's hierarchy of (avoided) activities and SUDS 108
Figure 20 Graphical representation of Sue's scores on Y-BOCS questionnaire at start and end of therapy 110
Figure 21 Graphical representation of Sue's scores on CORE-OM questionnaire at start and end of therapy 111

iv) Single Case Experimental Design Clinical Practice Report

Figure 22 Timeline of Tom's significant life experiences and referrals 128
Figure 23 Newcastle Formulation compiled with staff based upon James (2011) 137
Figure 24 Tom's example vicious cycle of worry 141
Figure 25 Visual representation of baseline and intervention frequency data 149
Figure 26 Median line graph of baseline data 150
Figure 27 Median line graph of baseline and intervention data 151

List of Tables

a) Volume One: Research Component

i) Literature Review

Table 1 Table of additional WMT sub-tests (not part of WMT effort indices) 6
Table 2 Search terms for meta-analysis 11
Table 3 Criteria for the exclusion of studies 12
Table 4 Overview of included studies 21
Table 5 Criteria for rating risk of bias (adapted from Whiting et al., 2011) 24
Table 6 Risk of bias in primary studies 27
Table 7 Univariate test performance indicators 33
Table 8 Overview of validity data reported by studies using WMT (Overall Index) 35
Table 9 Global measures of test performance 42
Table 10 Bivariate random effects model for Positive and Negative Likelihood Ratios 49
Table 11 Bivariate random effects model for individual WMT sub-tests 49

ii) Empirical Paper

Table 12 Complementary expertise of patient and healthcare professionals in the Shared Decision Making process 71
Table 13 Behaviours identified as essential in the Shared Decision Making process 72
Table 14 Participant demographic information 79
Table 15 Outline and description of steps of IPA 80
Table 16 Principles of validity in qualitative research 82
Table 17 Overview of superordinate and subordinate themes identified through analysis 85

b) Volume Two: Clinical Practice Reports Component

i) Psychological Models Clinical Practice Report

Table 1 RCADS and SDQ scores as completed during Choice assessment……9

ii) Service Evaluation Clinical Practice Report

Table 2 NICE guidelines for children and adults…….. 40
Table 3 Questionnaire items, response format and corresponding aims…….. 47
Table 4 Clinical presentations for which therapeutic interventions are deemed valuable….. 53
Table 5 Clinician responses relating to the process of referring to therapies from Choice assessment. 55
Table 6 Clinician responses relating to the process of referring to therapies from Partnership. 56
Table 7 Frequency of clinician endorsement for additional methods to support decision-making…….. 58
Table 8 Frequency of clinician endorsement for delivery of additional support……... 58
Table 9 Participant views on how additional support may benefit clients ……..... 59
Table 10 Participant views on how additional support may benefit clinicians…… 60
Table 11 Internal and external influences on the organisational delivery of psychological therapies……..... .61

iii) Integrated Case Study Clinical Practice Report

Table 12 Overview and description of formal outcome and assessment measures used at the start of assessment and during final intervention session..... 83
Table 13 Sue's CORE-OM scores at start of intervention..... 89
Table 14 Sue's Y-BOCS scores and Symptom Checklist responses at start of intervention..... 89
Table 15 Structure and overview of assessment and intervention sessions..... 102
Table 16 Techniques used to support cognitive change..... 105
Table 17 Integration of CBT and Systemic models within therapy phases 112

iv) Single Case Experimental Design Clinical Practice Report

Table 18 Overview and description of formal outcome and assessment measures used at start of assessment and final intervention session 132
Table 19 Summary of Motivational Assessment Scale scores for refusal of personal care..... 134
Table 20 Key factors to be considered in the assessment and formulation of behaviour that challenges (NICE, 2019)..... 135
Table 21 Outline of session content in baseline and intervention phases 142
Table 22 Outline of content of Tom's PBS plan 144
Table 23 Scores on the CORE-LD and the Mini-MANS 152

Chapter I

Literature Review

A Meta-Analysis of Analogue Studies of Diagnostic Accuracy of the Word Memory Test

Abstract

Background

The accuracy of interpreting cognitive assessment data in the context of suspected neurological disorder or impairment is dependent on the active engagement of the client. Performance Validity Tests (PVT's) like the Word Memory Test (WMT) are advocated for routine use within neuropsychological evaluations. Due to the nature and challenges in the assessment of effort and motivation, evaluating the diagnostic accuracy of PVTs is dependent upon either 'known-group' or experimental methodologies. This study therefore sought to investigate the diagnostic accuracy of the WMT amongst studies using analogue simulators.

Method

A systematic search of the literature was undertaken using PsycINFO, Medline, EMBASE and Web of Science databases. A total 57 datasets were extracted for quantitative synthesis from 18 primary studies. Meta-analytic synthesis of univariate and bivariate indicators of test accuracy was performed.

Results

Results demonstrated that the WMT holds strong diagnostic accuracy in the detection of individuals instructed to simulate cognitive impairment, with a DOR 146.986 (95% CI 61.138 – 353.378). Bivariate meta-analysis produced an area under the curve of 0.97 and the associated sROC curve also shows strong diagnostic accuracy.

Conclusions

The strength of the conclusions able to be drawn were limited by the available literature. Administration of the WMT according to test publisher instructions, and more accurate reporting (including Genuine Memory Impairment Profile) would enable further synthesis of WMT classification accuracy to be undertaken.

Introduction

Effort Testing / Performance Validity Tests

Neuropsychological assessment uses several sources of information when exploring the cognitive, neurological, emotional and behavioural consequences of neural insult. Alongside interviews, questionnaires (self/other report) and observations, formalised neurocognitive testing is a mainstay of the profession. Neurocognitive assessment requires active, effortful participation from patients. The validity of applying neuropsychological interpretation to such data is predicated on the assumption of this effortful engagement (Helibronner et al., 2009; McMillan et al., 2009). Many psychological and motivational formulations for suboptimal effort have been cited, including emphasising the legitimacy and severity of neurocognitive symptoms, conversion and/or functional neurologic disorders, disinterest or disengagement with the assessment process or active dissimulation of neurocognitive deficits (malingering or factitious disorder). The misidentification of psychologically or motivationally mediated test failure as evidence of neurocognitive deficit may have profound and enduring consequences for the client. They may receive inappropriate services, receive inappropriate expectations for recovery and fail to receive potentially beneficial treatments targeted at the psychological or motivational basis for the test performance.

Multiple tests have been developed in an attempt provide a quantifiable assessment of the validity of individual performance during testing. Although sometimes described as ‘Symptom Validity Tests’ (SVTs), the more accurate terminology for this group of measures

would be 'Performance Validity Tests' (PVT's)[1]. Whilst some neuropsychological tests (e.g., Wechsler Memory Scale (WMS-IV; Weschler, 2009) have indices embedded to assess performance validity, specific PVT's are completed in addition to other neuropsychological tests. Standalone PVT's typically take the form of a forced-choice (FC) recognition memory task. Designed to appear as difficult cognitive tasks in fact, they rely largely or wholly on cognitive processes that are robust to cognitive insult (e.g., Goodrich-Hunsaker & Hopkins, 2009).

The majority of PVT's use a performance cut-off to identify respondents at risk of negative response bias – performance which falls below an established cut-point is deemed as 'invalid' and likely indicative of suboptimal effort (Bigler, 2014). This permits determination of the presence of suboptimal effort with considerable certainty and in some circumstances (for example, below chance responders) one may be able to establish intentionality. It is not possible for a PVT alone to determine the motivational (or other) origins of suboptimal performance. For example, whilst individuals with a dementia diagnosis may be likely to 'fail' a PVT they will demonstrate a characteristic pattern of performance (Green & Merten, 2013). There will also be evidence from additional clinical information and observation of a presentation consistent with severe cognitive impairment – i.e., requiring significant support with daily functioning. Where test failure is due to genuine and severe impairment and not suboptimal effort this would be deemed a false positive (FP) result. Accordingly, it is incorrect for PVT's to be referred to as tests of malingering, as PVT performance alone cannot establish the motivational basis for the behaviour.

[1] SVTs are predominantly used to assess the validity of self-reported symptomatology (Lippa, 2018) whereas PVT's are concerned with the identification of suboptimal effort or 'noncredible effort' on neurocognitive ability measures.

PVT's should be a key component of all neuropsychological assessments. Multiple position papers published in the United States advocate for the use of both stand-alone and embedded PVT's as the most valid and essential approach to determining the validity of test performance in neuropsychological evaluation (Bush et al., 2005; Heilbronner et al., 2009). A good practice framework for the routine use of PVT's in clinical practice across contexts - i.e., outside of forensic and litigation situations, has also been published by the British Psychological Society (McMillan et al., 2009).

PVT's form a significant part of the multidimensional framework for classification of malingering for many years (Slick et al., 1999). In a recent update (2020) revised criterion lists somatoform, factitious and function neurological disorders as conditions where a classification of malingering can be attributed in the presence of non-credible PVT performance (see Sherman et al., 2020 for further discussion).

Word Memory Test

The Word Memory Test (WMT; Green et al., 1996) is a computerised PVT. It is an assessment of verbal learning and recognition memory, which involves a list of twenty semantically related word pairs presented twice to the respondent. Following two learning trials, the Immediate Recognition (IR) sub-test is undertaken. This is a forced-choice recognition task inviting the patient to choose the word from the original list of 40 new word pairs. After a thirty-minute delay, a new group of 40-word pairs are presented as part of the Delayed Recognition (DR) trial. The Consistency index (CNS) is calculated according to the consistency of responses between the IR and DR sub-tests (percentage of words correctly identified on both trials). An 82.5% (Standard cut-off) score or below on any of these three indices is suggested to be indicative of suboptimal effort (Green, 2003). This cut-score is derived as three standard deviations below those with severe traumatic brain injury (Green et

al., 2002). The WMT also includes four measures of memory ability (containing secondary validity indicators) which are administered after the IR and DR. These are listed in Table 1:

Table 1

Additional WMT subtests (not part of primary effort indices)

WMT Subtest	Administration
Multiple Choice (MC)	Respondents are presented with the first word from the pair and instructed to select its pair from eight presented options.
Paired Association (PA)	Individuals are presented with the first word and invited to tell the tester the second.
Delayed Free Recall (DFR) & Long Delayed Free Recall (LDFR)	Individuals recall as many words as possible in any order, and after a further 20-minute delay.

Patient performance on the WMT may be compared to multiple comparison groups. These groups include patients with; Moderate-Severe Traumatic Brain Injury (TBI), neurological patients, impaired verbal memory, 'normal' controls, dementia, learning disability, persons with differing motivations to perform optimally (e.g., those tested in forensic settings or in relation to compensation claims) and people asked to simulate the effects of brain injury. To address the potential for FP's for suboptimal effort in the severely impaired, the Genuine Memory Impairment Profile (GMIP or Severe Impairment Profile, SIP) was developed by authors of the WMT. The GMIP seeks to differentiate between below cut-

off performance due to cognitive impairment and that resulting from either low engagement or deliberate underperformance. The GMIP is calculated based on mean effort sub-test scores as thirty points (or more) higher than mean scores of the memory sub-tests in the presence of significant cognitive impairment. The pattern of increasingly poor performance on more demanding WMT tasks has been demonstrated as the typical pattern for clients with dementia (Green et al., 2011) but is not necessarily specific to the severely impaired and can be seen in some simulators (Armistead-Jehle & Denney, 2014).

Diagnostic Accuracy of Performance Validity Tests

Diagnostic accuracy describes the discriminative ability of a test to distinguish the presence and absence of a condition (Šimundić, 2009). Typical studies of diagnostic accuracy for specific tests/measures report the classification accuracy of the 'index test' in comparison to a reference test which is viewed as providing the 'true' diagnostic classification (Cohen et al., 2016). The validity of a diagnostic accuracy study therefore rests in a large part upon the accuracy and credibility of the reference test. It is common for diagnostic accuracy studies to refer to the reference test as 'gold' or 'silver' standard depending upon whether it provides excellent versus merely acceptable diagnostic classification. Within the field of performance validity testing, it is widely acknowledged that no such "gold standard" exists (Rogers, 2008). The lack of a gold standard reference test relates to the inability of PVT's to access the motivational basis of poor performance and to determine whether the respondent's performance was effortful. Accordingly, the motivational basis for the behaviour must be either imputed or experimentally manipulated. This is reflected in the two main methodologies of diagnostic accuracy for PVT's, namely known-group designs and analogue (simulator) designs (Rogers, 1997).

Known group designs

Within a known-group (also referred to as criterion-group) design, individuals who have a condition of interest are compared to those who are known not to have the condition or have another condition that the test will need to be insensitive to (Rogers, 1997). The accurate identification of these known-groups requires independent reference standards. In the absence of "gold standard" reference tests a "silver standard" reference criterion must be employed. This offers determination of diagnostic classification but with considerable error.

In the context of neurocognitive PVT's the known groups design typically involves (a) persons with legitimate cognitive impairments applying full effort to the PVT and (b) 'real world' clients dissimulating neurocognitive deficits. The derivation of known groups in PVT research is typically based upon Slick et al's (1999) criteria. However, fundamental to this criterion is the failure on at least one other validated PVT. Classification accuracy therefore depends upon the PVT's used to initially classify the groups. Use of a PVT or combination of PVT's with less than optimal sensitivity, then the non-simulating group will include simulators. This will then mean that the PVT under investigation, if more sensitive, will appear to have a lot of FP's.

Revised guidance (Sherman et al., 2020) suggests the use of at least two or three PVT's validated in clinical populations with a low false-positive rate (specificity of 0.90 or above) with a failure on two or more PVT's indicative of invalid performance. As the number of PVT's administered increases, the ratio of PVT failures to PVT's administered is suggested as the best indicator of invalid response (Sherman et al., 2020).

In differential prevalence designs of diagnostic accuracy of PVT's, the silver standard would typically involve comparing groups with a high and low probability of cognitive dissimulation, thus establishing the diagnostic accuracy of the PVT by comparison to the

"known prevalence" of dissimulation within these groups. Diagnostic accuracy therefore becomes a probabilistic statement, the validity of which is determined by the accuracy of the "known prevalence" in each group. For example, if litigating and non-litigating clients were compared although there would likely be a higher prevalence of dissimulation in the former, the design will be problematic as many litigating clients may not be dissimulating. Furthermore, the non-litigating group may have other incentives (e.g. sickness or unemployment benefits) that would provide a motivation for exaggeration.

The primary critique of known-groups designs using silver standard reference tests in relation to diagnostic accuracy is that it is not possible for us to determine the true positive (TP) rate of dissimulation as there is no way to ascertain if an individual is truly performing below their ability. Where true dissimulation cannot be determined apriori in genuine patient groups it may be determined with 100% accuracy by instructing participants to fake bad or to perform to the best of their abilities. This points to the use of analogue studies in exploring diagnostic accuracy of a test.

Analogue(simulator) designs

In contrast to the use of naturally-occurring "known groups", analogue studies are an experimental design. Participants are randomly assigned between an 'honest responding' and a 'simulating' condition. Due to the ease of application and high internal validity, simulation designs are the most used within the field of performance validity research (Gillard, 2010). Experimental allocation of participants to conditions allows researchers to know with high accuracy who is dissimulating and who is not. Group allocation serves as the "gold standard" reference criteria against which PVT performance is evaluated. To enhance the external validity of analogue studies, participants from clinical populations (where the test may be used in clinical practice) can be recruited and assigned to the genuine responding and simulating conditions alongside non-clinical participants (Rosenfeld & Penrod, 2011).

However, due to concerns about the ability of individuals with psychological/cognitive impairment to maintain dissimulation, the use of such individuals as experimental simulators has been advised against (Rosenfeld & Penrod, 2011).

Results from analogue studies can be cited as limited in their generalisability. It is unclear, whether analogue dissimulators behave in the same way as 'real life' dissimulators (Sweet et al., 2008). Namely, individuals who are instructed to simulate may have less motivation to perform in a convincing way than an individual who has a 'genuine' incentive to 'beat' a PVT, even with the introduction of additional external incentive (Tombaugh, 1997). Bianchini, et al. (2001) cite that few studies report the instructions given to their participants – either to feign or perform to best ability – to a level/extent which enables replication nor descriptions of any procedures in place to evaluate the level of compliance and understanding of those instructed to dissimulate.

Aims of this Review

A meta-analysis of diagnostic accuracy seeks to determine the effectiveness of the index test at classifying individuals correctly across multiple studies, synthesising this data into an overall description of the test accuracy. In the absence of a "gold standard" against which one can state an individual is dissimulating or not, it is not possible to establish the true detection rates across individual studies using known-groups designs. This meta-analysis will seek to synthesise the data from analogue studies using the Word Memory Test to explore its diagnostic accuracy.

As analogue studies have been critiqued for using potentially unmotivated and un-coached participants in the dissimulation condition, this review will conduct a subgroup analysis of those studies that used naïve and/or unmotivated participants and those that had received coaching or some reward contingent upon performance.

Method

Search of Electronic Databases

A systematic search of the literature was undertaken in July 2019 utilising electronic databases – PsycINFO, Medline, EMBASE and Web of Science. To obtain a comprehensive overview of published literature into the classification/diagnostic accuracy of the Word Memory Test, search terms were developed based upon previous guidelines (Devillé et al., 2002). Search terms of diagnostic accuracy and Word Memory Test were used as key constructs, from which truncations and synonyms/component terms were generated to maximise opportunity to obtain relevant papers (Table 2). The full search strategy can be found in Appendix 1.

Table 2

Search terms

Construct	Search Terms
Diagnostic accuracy	Sensitivity [OR] specificity [OR] test sensitivity (exp) [OR] test specificity (exp) [OR] valid* [OR] test validity (exp) [OR] psychometrics (exp) [OR] statistical validity (exp) [OR] predictive value* [OR] PPV [OR] NPV [OR] likelihood ratio* [OR] likelihood functions [OR] reference value* [OR] area under curve [OR] AUC [OR] receiver operating characteristic [OR] ROC [OR] diagnos* [OR] false positive [OR] false negative [OR] psychodiagnostic measures (exp) [OR] (diagnos* adj3 (performance* OR accurac* OR utilit* OR value* OR effectiveness OR efficien*))
	AND
Test of interest	Symptom validity [OR] performance validity [OR] effort test*
	AND
Word Memory Test	Word Memory Test [OR] WMT

Inclusion Criteria

Articles were screened according to inclusion/exclusion criteria (Table 3). Studies were included if their procedure consisted of the allocation of a sample of adult participants over the age of 18 years to an experimental condition in which they were instructed to simulate impairment. The classification accuracy of the WMT was compared to these experimentally defined groups.

Table 3

Criteria for the exclusion of studies

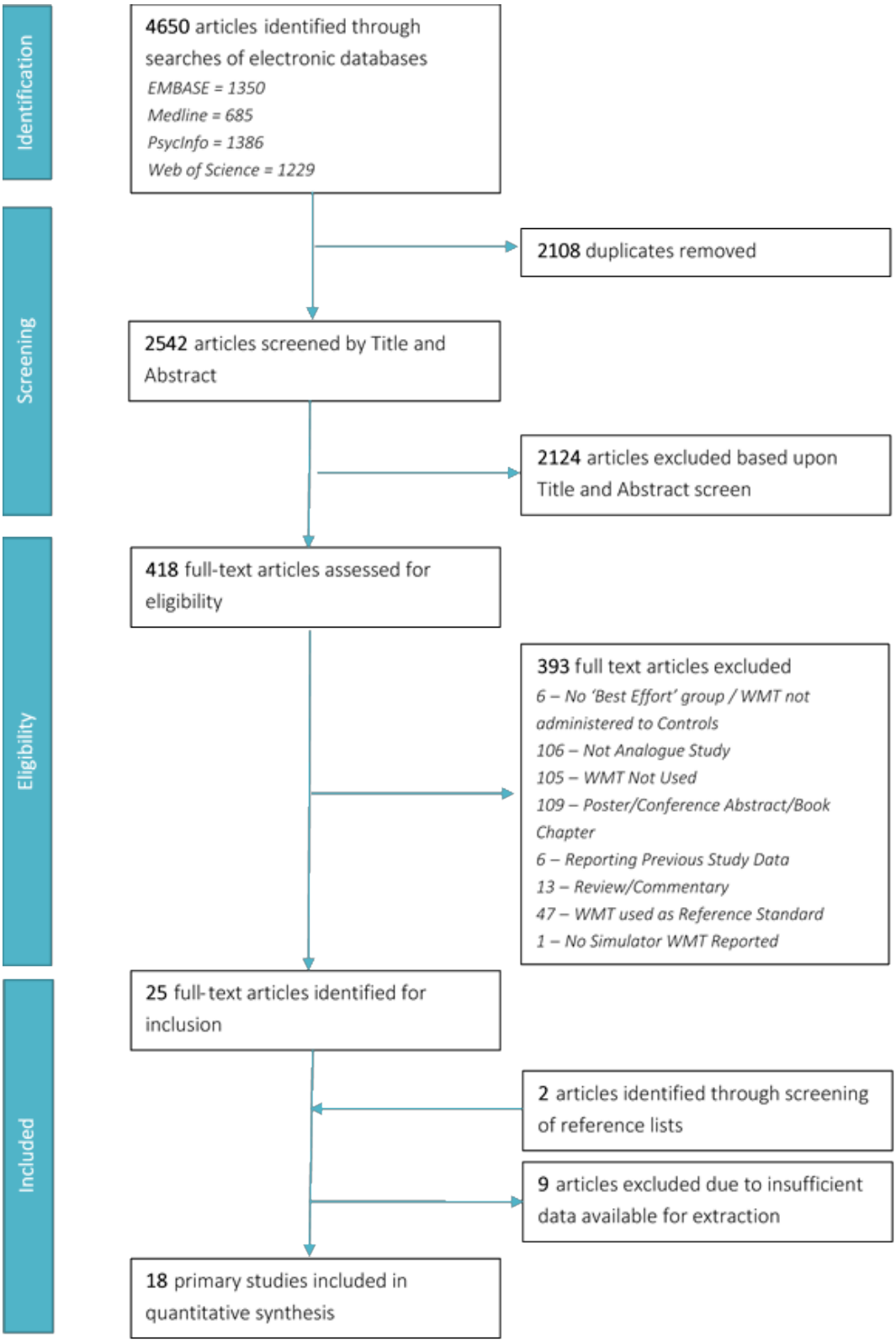
Exclusion Criteria	**Rationale / Justification**
Types of Study	
Studies in which group membership/allocation was not defined prior to the administration of the WMT.	This review seeks to explore the classification accuracy of the WMT therefore the WMT cannot be used as the criterion standard.
Instructions given to participants did not include either an instruction to simulate a condition or to perform to their best effort.	The absence of such instruction would not be indicative of an analogue methodology.
Participant Characteristics	
Study participants recruited from a paediatric/child sample	
Outcome Data	
Insufficient data provided to enable construction of 2x2 contingency table to allow for calculation of sensitivity, specificity, and positive and negative likelihood ratio values.	In order to be able to ascertain the diagnostic accuracy of the WMT, data to calculate these values was necessary.
Type of Article	
Books, book chapters, meta-analyses/systematic reviews, commentaries, letters to editors, conference abstracts, posters and 'grey literature' (e.g., unpublished doctoral dissertations)	These articles do not provide the type of data required for this review and may introduce risk of bias in the absence of full quality review.

Paper Selection

The PRISMA guidelines (Moher et al., 2009) for the selection of papers were followed (Figure 1). A total of 2542 articles were identified through the search. Duplicates were removed. Remaining articles were screened by title and abstract and removed if they met any exclusion criteria. Where this was not able to be determined, articles were reviewed in full texts against the exclusion criteria. At this stage, the most common reasons for exclusion were non-journal articles (including those reporting poster and conference abstracts – n=109), non-analogue study designs (n=106) and absence of use of the WMT (n=105). A total of 18 articles satisfied the full criteria for inclusion in this meta-analysis.

Figure 1

PRISMA chart summarising systematic search results and application of inclusion criteria



Data Extraction

Data were extracted by the corresponding author. Any ambiguity in the data was discussed with the other reviewers and where it was not possible to reach agreement, study authors were contacted for clarification. The initial anticipation was that data would be extracted in the form of True Positive (TP)/False Positive (FP) rates or sensitivity and specificity, The minimum requirement for data extraction was the reporting of numbers of WMT passes/failures and sample sizes for each of the two groups (simulator and control), from which sensitivity and specificity (including TP, True Negative (TN), FP, False Negative) rates could be derived. A full description of the information extracted from each paper can be found in Appendix 2.

Some studies removed participants who were not deemed to have performed to the authors expectations (Elbaum et al., 2019; Kirchhoff & Steinert, 2019; Lupu, et al. 2018; Merten, et al., 2004; Ortega et al., 2013; Tan, et al. 2002; Tydecks, et al. 2006). Unless these provided a convincing rationale and procedure for the removal of participants, this practice was viewed as a significant risk of bias as this may lead to the under-estimation of FP rates. Where studies reported that removal of participants who were not considered to have dissimulated to the expected level of competency or performed in line with expected testing parameters, an Intention to Treat (ITT) analysis was conducted. These participants were added back into the sample as TP results.

Where studies described multiple simulator or control groups (i.e., naïve/sophisticated simulators; healthy/patient controls), the duplication of participants across conditions or in multiple samples, there is a risk of artificial inflation within the overall sample size for this meta-analysis. Consequently, this inflation may reduce the confidence intervals in the calculation of diagnostic accuracy.

Missing data and zero frequency data

Event rates with a zero count can cause numerical problems when synthesising proportions, resulting in division by zero errors. These counts occur in studies where the sample size is insufficient to enable calculation of the event rate – e.g., where the WMT is reported with perfect sensitivity or specificity. Where this occurred a small constant of 0.5 was added to the zero-event rate to rectify the division by zero error.

The data analysis strategy follows the guidelines for the Centre for Applied Psychology, University of Birmingham and is paraphrased below.

Meta-analyses of diagnostic accuracy typically calculate pooled univariate measures such as sensitivity and specificity, positive and negative likelihood ratios and diagnostic odds ratios. However, these univariate measures cannot be considered in isolation from each other and the use of bivariate methods of meta-analysis are recommended (Zwindermann & Bossuyt, 2008). Bivariate methods calculate test sensitivity and specificity through the creation of a Receiver Operating Characteristic (ROC). A ROC curve maps the TP rate as a function of the FP rate for all possible cut-off values of a test. The resulting curve can therefore be interpreted as the average sensitivity across all possible specificities for the test in question.

The strategy for this meta-analysis of diagnostic accuracy will be to calculate the univariate meta analytic synthesis for sensitivity and specificity, likelihood ratios and the diagnostic odds ratio. Differences in the threshold for defining performance as 'valid' or 'invalid' may be a source of heterogeneity between studies which will impact upon the validity of the synthesis. Where there are differences in the thresholds used to classify performance on the WMT, these will be presented as separate syntheses. Univariate analyses will be supplemented by the calculation of the bivariate random effects model using the procedures described by Reitsma et al. (2005). This model will then be used to generate a summary

ROC (sROC) curve. An sROC graph plots test sensitivity (x-axis) against the FP rate (1-specificity; y-axis) at a 95% confidence interval (Jones, 2005).

Characteristics of Included Studies

Twenty-seven full text articles were assessed as eligible for inclusion in this analysis. Due to limitations in the reporting of primary outcomes required to undertake data extraction, only 18 unique papers (66.67%) provided the data required to complete the 2x2 diagnostic accuracy tables[2]. These 18 studies reported 57 indices of diagnostic accuracy across different coaching or motivational conditions. An overview of the characteristics of the included studies can be found in Table 4.

Sample Size

The total sample size within this meta-analysis was=1411 (range=10 – 242, mean=74.3, median=70). One paper (Brockhaus & Merten, 2004) reported outcomes from two separate studies utilising the same simulator group and therefore contributed two sets of data. To avoid duplication, the second simulator group (n=100) has not been counted in this total.

Simulator Population

The majority of the 18 studies (n=16) recruited participants from student populations or community volunteers. All studies used a between-subjects design, except for Larsen et al. (2010) for whom a smaller sample was recruited, and participants acted as their own control. Only two studies reported assignment of clinical participants to the simulating condition (Green et al., 2002; Kirchhoff & Steinert, 2019).

[2] The majority of these articles were excluded due to the reporting of mean scores for simulator and control groups, and rates of pass or fail detection.

Control Groups

Eight studies recruited participants with a known condition or impairment as the control group instructed to perform to best ability in testing (Brockhaus & Merten, 2004; Edmundson et al., 2017, Green et al., 2003; Kirchhoff & Steinert, 2019, Martins & Martins, 2010; Ortega et al., 2013; Lindstrom et al., 2009; Shandera et al., 2010). Six studies reported multiple control groups. These studies used separate groups of 'healthy' normal participants and patient populations as comparisons for simulating participants.

Patient populations included Learning Disability (LD) as defined by DSM-IV criteria (n=2, Brockhaus & Merten, 2004; Shandera et al., 2010) but also print-related LD (n=1, Lindstrom et al., 2009). One study (Kirchhoff & Steinert, 2019) recruited solely from a psychiatric population (inpatient and outpatients with anxiety and depression). Two studies reporting cognitive impairment described this as established based on neuropsychological assessment outside of the study itself (Martins & Martins, 2010; Ortega et al., 2013)

Simulator Instructions

Most included articles (n=10) instructed simulators to demonstrate cognitive impairment associated with TBI. These scenarios typically cited imagined involvement in an earlier car accident with the intention to simulate effectively to obtain financial compensation, except for Ortega and colleagues (2013) who were seeking additional health insurance benefits. Three articles instructed the simulation of Learning Disability (Lindstrom et al., 2009; Osmon et al., 2006; Shandera et al., 2010) but with differing motivations outlined – seeking of additional higher education support/resources and a forensic plea.

Level of sophistication and coaching of Simulators

Sophistication refers to the level of instruction and preparation provided to simulating participants prior to their engagement with the WMT. Appendix 3 contains a table summarising levels of coaching across studies. Ten of the 18 studies were classified as

having naïve or non-sophisticated simulators. These participants were provided with instruction/scenarios to simulate without guidance or prior knowledge of the simulated impairment. Amongst the remaining studies levels of sophistication and instruction varied, ranging from professional volunteers with an anticipated knowledge/understanding of cognitive impairment (Green et al., 2002) to the provision of information about symptoms associated with conditions (Shandera et al., 2010). Alongside coaching relating to symptoms, two studies included test-level coaching regarding strategies to perform well on PVT's (Lau et al., 2017; Larsen et al., 2010). One study (Tan et al., 2002) did not provide standard instructions/coaching to participants and instead encouraged self-preparation and research to overcome performance validity testing.

Mode and Administration of the WMT

Only three primary studies reported the oral administration of the WMT (Brockhaus & Merten, 2004(b); Shandera et al., 2010; Tydecks et al., 2006). All other studies administered the computer version of WMT, except for Davis et al. (2010), Edmundson et al. (2017) and Manderino and Gunstad (2018) who did not state their administration method.

Use of the full WMT (including memory sub-tests), which would allow use of the severe impairment profile, was reported in only five of the included studies (Edmundson et al., 2017; Gree et al., 2002; Martins & Martins, 2010; Osmon et al., 2006; Tydecks et al., 2006). Ten studies reported results based upon the primary effort indices of the WMT although for all except Davis et al. (2010), Kirchhoff and Steinert (2019), Lau et al. (2017), Manderino & Gunstad (2018) and Ortega et al. (2013) it is unclear if the additional sub-tests were administered. The two Hebrew studies (Elbaum et al., 2019; Lupu et al., 2018) and Larsen et al. (2010) both reported sole administration of the Immediate Recall (IR) and/or Delayed Recall indices respectively.

Cut-Off used for the WMT

Sixteen studies utilised the standardised cut-off of ≤82.5% on IR, DR or CNS subscales as indicative of dissimulation. Two studies (Manderino & Gunstad, 2018; Ortega et al., 2013) reported classification based on score <90%. Ortega et al (2013) also classified WMT failure based on an average score across the three primary effort subtests as opposed to standard administration which cite a score of 82.5% or below on any of the respective subtests. The below-chance classification (<37.5%) was only reported as a supplementary classification by Tan et al. (2002).

Language

Where there had been no mention of the language of use it may be assumed that most studies (n=12) have used English versions of the WMT. Within the included studies, the WMT has been translated into four languages. Translations into German (Brockhaus & Merten, 2004; Kirchhoff & Steinert, 2019) and Hebrew (Elbaum et al., 2019; Lupu et al., 2018) were most reported. Further translations into Portuguese (Martins & Martins, 2010) and Russian (Tydecks et al., 2006) have also been cited.

Table 4

Overview of Included Studies

	Sample Size	Simulator Population	Control	Instruction	Level of Sophistication	Motivation/ Incentive	Administration of WMT	Administration Mode	Cut-Off	Language
Brockhaus & Merten (2004)	a) 127 b) 32	Healthy volunteers	a) Healthy b) LD[a]	Cognitive impairment	Naive	Not stated	IR, DR & CNS only reported	a) Computer b) Oral	<82.5%	German
Davis, Wall, Ramos, Whitney & Barisa (2010)	82	University students	Healthy controls	Post-accident symptoms	Naïve & Coached	Not stated	IR, DR & CNS only	Not stated	<82.5%	Not stated
Edmundson, Berry, Combs et al. (2017)	76	Undergraduate students	Patient (ADHD)[b]	Feign ADHD	Naïve & Coached	Cash incentive	Full WMT	Not stated	<82.5%	Not stated
Elbaum, Golan, Lupu, Wagner & Braw (2019)	60	Undergraduate students	Healthy controls	Cognitive impairment	Naive	Not stated	IR only	Computer	<82.5%	Hebrew
Green, Lees-Haley & Allen (2002)	105	a) Clinician vols. b) Patient[c]	a) Healthy b) Patient[c]	Memory impairment	Sophisticated	Not stated	Full WMT	Computer	<82.5%	Not stated
Kirchhoff & Steinert (2019)	60	Psychiatric population[d]	Psychiatric population[d]	Exaggeration of medical condition	Sophisticated	Not stated	IR, DR & CNS only	Computer	<82.5%	German
Larsen, Allen, Bigler, Goodrich-Hunsaker & Hopkins (2010)	10[g]	Community volunteers	Own controls	Memory impairment	Coached	Not stated	IR and DR only	Computer	<82.5%	Not stated
Lau, Basso, Estevis, Miller, Whiteside, Combs & Artensen (2017)	80	Undergraduate students	Healthy controls	Cognitive impairment post mild TBI	Multiple Coached	Not stated	IR, DR & CNS only + composite score	Computer	<82.5%	Not stated
Lindstrom, Lindstrom, Coleman, Nelson & Gregg (2009)	75	Undergraduate students	Patient[e] and healthy controls	Learning disability	Naive	No incentive	IR, DR, CNS only reported	Computer	<82.5%	Not stated
Lupu, Elbaum, Wagner & Braw (2018)	92	Undergraduate students	Healthy controls	Cognitive impairment	Naive	Cash incentive	IR only	Computer	<82.5%	Hebrew

	Sample Size	Simulator Population	Control	Instruction	Level of Sophistication	Motivation/ Incentive	Administration of WMT	Administration Mode	Cut-Off	Language
Manderino & Gunstad (2018)	242	Undergraduate psychology students	Healthy controls	Concussion	Naive	Gift card raffle	IR, DR & CNS only + composite	Not stated	<90%	Not stated
Martins & Martins (2010)	70	Undergraduate students & relatives	Patient(MCI)[f] and healthy controls	Retirement for medical reasons	Naive	No incentive	Full WMT	Computer	<82.5%	Portuguese
Merten, Henry & Hilsabeck (2004)	20	Volunteers	Healthy controls	Cognitive impairment	Coached	Book incentive	IR, DR & CNS only reported	Computer	<82.5%	Not stated
Ortega, Labrenz, Markowitsch & Piefke (2013)	60	Healthy volunteers	Patient[g] and healthy controls	Cognitive impairment	Naive	Cash incentive	IR, DR & CNS only	Computer	<90%	Not stated
Osmon, Plambeck, Klein & Mano (2006)	84	College students	Healthy controls	Reading/ cognitive speed	Naive	Cash incentive	Full WMT	Computer	<82.5%	Not stated
Shandera, Berry, Clark, Schipper, Graue & Harp (2010)	59	Community volunteers	Patient[h] and healthy controls	LD in forensic setting	Coached	Cash incentive	IR, DR & CNS only reported	Oral	<82.5%	Not stated
Tan, Slick, Strauss & Hultsch (2002)	56	Undergraduate students	Healthy controls	Cognitive impairment	Self-coached	Not stated	IR, DR & CNS only reported	Computer	<82.5% & <37.5%	Not stated
Tydecks, Merten & Gubbay (2006)	21	Volunteers	Healthy controls	Cognitive impairment	Coached	Voucher incentive	Full WMT	Oral	<82.5%	Russian

[a] Patients meeting DSM-IV criteria for Learning Disability [b] ADHD diagnosis established prior to age 18 through screening interview by mental health practitioner [c] Patients referred for neuropsychological assessment who passed WMT effort scores during regular examination (head injuries, neurological, psychiatric, orthopedic and other conditions) [d] Anxiety and depression inpatient and outpatient psychiatric [e] Students with print-related LD (dyslexia/dysgraphia), 10+ hours of neuropsychological assessment as part of diagnosis [f] Amnestic MCI according to Petersen criteria (Petersen et al., 2001) [g] Neurological patients with moderate cognitive impairment recruited from inpatient and outpatient rehabilitation centres – further neuropsychological assessment battery within study [h] Adults with LD diagnosed according to DSM-IV criteria, co-morbid psychiatric not excluded

Assessment of Bias

To assess the potential risk of bias within the literature, a set of quality criteria were developed (Table 5) based upon the validated Quality Assessment of Diagnostic Accuracy Studies checklist (QUADAS-2; Whiting et al., 2011), adapted to meet the requirements for a review of the Word Memory Test. For each of the four domains considered by the QUADAS, a rating was allocated to classify the potential extent of associated risk – High, Low or Unclear (Table 6). Ratings of concern associated with the applicability of the primary studies was also rated following the same grading. For the purposes of rating, where one aspect of the quality criterion was rated as 'High Risk of Bias' then, following discussion with supervisors, the overall rating for this domain was also rated as 'High'.

Table 5

Criteria for the rating of risk of bias (adapted from Whiting et al. 2011)

Domain	Details	Risk of Bias
Patient/Participant Selection – Risk of Bias	Was a case–control design avoided? A study ideally should be a consecutive or random sample of individuals to prevent the potential for bias. Studies that make inappropriate exclusions (for example, not including "difficult-to-diagnose" patients) may result in overestimation of diagnostic accuracy. Studies enrolling participants with known impairment or motivation to dissimulate and an honest responding control group may similarly exaggerate diagnostic accuracy. Did the study avoid inappropriate exclusions? Excluding patients with "red flags" for dissimulation who may be easier to identify may lead to underestimation of diagnostic accuracy.	High Risk - No random sampling. Inappropriate exclusions. Participants in simulated group have prior knowledge/experience (e.g., participants who have previously had Brain Injury). Exclusion on the basis of effort test scoring. Unclear Risk - Insufficient information on sampling method and inclusion/exclusion criteria. Low Risk - Random or consecutive sampling. No inappropriate exclusions. No prior knowledge or experience of the instructed "condition" (e.g., brain injury).
Patient/Participant Selection – Applicability Concerns	Concerns about applicability may exist if patients included in the study differ from those targeted by the review question in terms of severity of the target condition, demographic features, presence of differential diagnosis or comorbid conditions, setting of the study, and previous testing protocols.	High Concern - Individuals with the condition instructed to simulate (e.g., cognitive impairment) Unclear Concern - Psychology undergraduates Low Concern - Members of the public with or without coaching.
Index Test – Risk of Bias	The index test results should be interpreted without knowledge of the results of the reference standard? This criteria is similar to "blinding" in intervention studies. Knowledge of the reference standard may influence interpretation of index test results. The potential for bias is related to the subjectivity of interpreting index test and the order of testing. If the index test is always conducted and interpreted before the reference standard, this item can be rated "low risk of bias." If a threshold was used, was it prespecified? Selecting the test threshold to optimize sensitivity and/or specificity may lead to overestimation of test performance. Test performance is likely to be poorer in an independent sample of patients in whom the same threshold is used.	High Risk - No blinding of researchers to group membership Unclear Risk - Insufficient information on "blinding" of research facilitators to group membership Low Risk - Clinical/comparison groups defined prior to administration of Word Memory Test. Researchers were blinded to group membership.
Index Test – Applicability Concerns	Variations in test technology, execution, language translation or interpretation may affect estimates of the diagnostic accuracy of a test. If index test methods vary from those specified in the review question, concerns about applicability may exist.	High Concern - Non-standard administration of WMT. Translated version. Use of non-standardised cut-offs / reporting of multiple cut-offs. Unclear Concern - Word Memory Test not administered in full - IR, DR and CNS only. Low Concern - Word Memory Test administered as per manual instructions - alone, standardised format. Cut-offs

Domain	Details	Risk of Bias
		pre-determined. Use of Profile analysis.
Reference Standard – Risk of Bias	Is the reference standard likely to correctly classify the target condition? Estimates of test accuracy are based on the assumption that the reference standard is 100% sensitive and that specific disagreements between the reference standard and index test result from incorrect classification by the index test. Were the reference standard results interpreted without knowledge of the results of the index test? The potential for bias is related to the potential influence of previous knowledge on the interpretation of the reference standard. For the purpose of this analysis, the reference standard was determined as the level of information provided within studies about the instruction/information and coaching provided to participants and the level to which this is able to be replicated/validated.	High Risk - No information provided regarding the instructions given to participants. Unclear Risk – Insufficient information on the instructions provided to simulating and control participants. Low Risk – Clear information regarding the instructions (and information/coaching) given to participants in simulator and control conditions. Instructions sufficient for participants to understand task demands.
Reference Standard – Applicability Concerns	The reference standard may be free of bias, but the target condition that it defines may differ from the target condition specified in the review question. For example, when defining dissimulation, the reference standard is generally based on a specific test however, the threshold above which a result is considered positive may vary. For the purpose of this analysis the assessment of applicability concerns was based upon the level of information provided regarding the confirmation/validation of participant understanding and compliance with instructions to dissimulate or perform to best effort.	High Concern – No check/confirmation with participants regarding understanding and conduct of instructions Unclear Concern – Insufficient information on how participants understood and followed instructions in control and simulator conditions Low Concern - Evaluation and confirmation that participants followed instructions given.
Flow and Timing	Was there an appropriate interval between the index test and reference standard? Results of the index test and reference standard are ideally collected on the same patients at the same time. If a delay occurs or if treatment begins between the index test and the reference standard, recovery or deterioration of the condition may cause misclassification. The interval leading to a high risk of bias varies among conditions. A delay of a few days may not be problematic for patients with chronic conditions, but it could be problematic for patients with acute infectious diseases. Conversely, a reference standard that involves follow-up may require a minimum follow-up period to assess whether the target condition is present. For example, to evaluate magnetic resonance imaging for early diagnosis of multiple sclerosis, a minimum follow-up period of approximately 10 years is required to be confident that all patients who will fulfil the diagnostic criteria for multiple sclerosis will have done so. Did all patients receive the same reference standard? Verification bias occurs when only a proportion of the study group receives confirmation of the diagnosis by the reference standard, or if some patients receive a different reference standard. If the results of the index test influence the decision on whether to perform the reference standard or which reference standard is used, estimated diagnostic accuracy may be biased. Were all patients included in the analysis? All participants recruited into the study should be included in the analysis. A potential for bias exists if the number of patients enrolled differs from the number of	High Risk - Post-test removal from either group without apparent or valid justification. Not all patients received same instructions. Unclear Risk - Insufficient information on time period between provision of instructions and conduct of WMT. Post-test removal of participants with apparent justification. Low Risk - Individuals were given instructions to malinger and then undertook the WMT. All participants who were recruited were included in analysis, or appropriate reasons for exclusion or description of difference between any who were excluded

patients included in the 2 x 2 table of results, because patients lost to follow-up differ systematically from those who remain.

Table 6

Risk of bias in primary studies

Study	Selection Bias	Selection Applicability	Index Test Bias	Index Test Applicability	Reference Standard Bias	Reference Standard Applicability	Flow and Timing
Brockhaus & Merten – 2004 – *Healthy Control*							
Brockhaus & Merten – 2004 – *LD Control*							
Brockhaus & Merten – 2004 – *Mixed Control*							
Davis, Wall, Ramos, Whitney & Barisa – 2010 – *Naïve Simulator*							
Davis, Wall, Ramos, Whitney & Barisa – 2010 – *Soph. Simulator*							
Edmundson, Berry, Combs et al.. 2017 – *IR Index, Naïve Simulator*							
Edmundson, Berry, Combs et al.. – 2017 – *DR Index, Naïve Simulator*							
Edmundson, Berry, Combs et al.. – 2017 - *CNS Index, Naïve Simulator*							
Edmundson, Berry, Combs et al.. – 2017 – *IR Index, Soph. Simulator*							
Edmundson, Berry, Combs et al.. - 2017 – *DR Index, Soph. Simulator*							
Edmundson, Berry, Combs et al.. - 2017 – *CNS Index, Soph. Simulator*							
Elbaum, Golan, Lupu, Wagner & Braw - 2019							
Green, Lees-Haley & Allen – 2002 –*Patient Sim. & Control*							
Green, Lees-Haley & Allen – 2002 – *Professional Sim. & Healthy Control*							
Kirchhoff & Steinert (2019)							
Larsen, Allen, Bigler, Goodrich-Hunsaker & Hopkins (2010)							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *IR Index, Naïve Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *IR Index, Soph. Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *DR Index, Naïve Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *DR Index, Soph. Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *CNS Index, Naïve Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *CNS Index, Soph. Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *Overall Index, Naïve Sim.*							
Lau, Basso, Estevis, Miller, Whiteside, Combs & Arentsen – 2017 – *Overall Index, Soph. Sim.*							
Lindstrom, Lindstrom, Coleman, Nelson & Gregg – 2009 – *Naïve Sim., Normal Control*							
Lindstrom, Lindstrom, Coleman, Nelson & Gregg – 2009 – *Naïve Sim., LD Control*							
Lindstrom, Lindstrom, Coleman, Nelson & Gregg – 2009 – *Naïve Sim., Mixed Control*							
Lupu, Elbaum, Wagner & Braw (2018)							
Manderino & Gunstad – 2018 – *IR Index*							
Manderino & Gunstad – 2018 – *DR Index*							
Manderino & Gunstad – 2018 – *CNS Index*							
Manderino & Gunstad – 2018 – *Overall Index*							
Martins & Martins – 2010 – *Naïve Sim, Healthy Control*							
Martins & Martins – 2010 – *Naïve Sim, MCI Control*							
Martins & Martins – 2010 – *Naïve Sim, MCI Control, Profile Index*							
Martins & Martins – 2010 – *Naïve Sim, Mixed Control, Profile Index*							
Martins & Martins – 2010 – *Naïve Sim, Mixed Control*							
Merten, Henry & Hilsabeck – 2004 – *IR Index*							
Merten, Henry & Hilsabeck – 2004 – *DR Index*							

Study	Selection Bias	Selection Applicability Concern	Index Test Bias	Index Test Applicability	Reference Standard Bias	Reference Standard Applicability	Flow and Timing
Merten, Henry & Hilsabeck – 2004 – *CNS Index*							
Ortega, Labrenz, Markowitsch & Piefke – 2013 – *Cognitive Imp. Control*							
Ortega, Labrenz, Markowitsch & Piefke – 2013 – *Healthy Control*							
Osmon, Plambeck, Klein & Mano – 2006 – *Naïve Sim. Dyslexia*							
Osmon, Plambeck, Klein & Mano – 2006 – *Naïve Sim. LD*							
Osmon, Plambeck, Klein & Mano – 2006 – *Mixed Sim*							
Shandera, Berry, Clark, Schipper, Graue & Harp – 2010 – *IR Index, Healthy Control*							
Shandera, Berry, Clark, Schipper, Graue & Harp – 2010 – *DR Index, Healthy Control*							
Shandera, Berry, Clark, Schipper, Graue & Harp – 2010 – *CNS Index, Healthy Control*							
Shandera, Berry, Clark, Schipper, Graue & Harp – 2010 – *IR Index, LD Control*							
Shandera, Berry, Clark, Schipper, Graue & Harp – 2010 – *DR Index, LD Control*							
Shandera, Berry, Clark, Schipper, Graue & Harp – 2010 – *CNS Index, LD Control*							
Tan, Slick, Strauss & Hultsch - 2002							
Tan, Slick, Strauss & Hultsch – 2002 – *Below Chance*							
Tydecks, Merten & Gubbay – 2006 – *IR Index*							
Tydecks, Merten & Gubbay – 2006 – *DR Index*							
Tydecks, Merten & Gubbay – 2006 – *CNS Index*							
Tydecks, Merten & Gubbay – 2006 - *Overall*							

Participant Selection – Risk of Bias

Overall, the selection and allocation of participants to simulator conditions was most frequently rated as a 'Low' risk of bias (n=12). Strengths within this domain were in the sampling of participants from community populations and the use of random allocation to assign simulator and control groups. Two studies were rated as 'High' risk of bias due to the exclusion of participants based upon WMT scores (Lupu et al., 2018; Tan et al., 2002). The remaining studies (n=5) were deemed 'Unclear' due to the recruitment of participants who may have had experience or knowledge of the simulated condition which could impact upon their ability to simulate based solely according to instruction or the absence of sufficient information regarding recruitment (e.g., Tydecks et al., 2006).

Participant Selection – Applicability Concerns

Thirteen studies recruited students and were therefore rated as 'Unclear' in relation to the applicability of their results to a wider population. The limitations that this recruitment may pose upon the external validity of findings is uncertain as it is unknown how similar these individuals may be to the wider population. Two studies (Green et al., 2002; Kirchhoff & Steinert, 2019) recruited from solely patient populations (groups with known diagnoses/conditions). One study (Green et al., 2002) was rated as 'High' risk of bias due recruiting medical professionals as their simulator group whom it may be anticipated have some knowledge of cognitive impairment and how this may present.

Index Test – Risk of Bias

Across the included studies the level of blinding for researchers was mixed. Three studies were rated as 'High' risk of bias within this domain as the group membership was already known by those responsible for WMT scoring and interpretation (Green et al., 2002; Lau et al., 2017; Lindstrom et al., 2009). A total of nine studies were rated as 'Low' risk of bias with group allocation defined prior to WMT administration and interpretation, and researchers blinded to membership. The absence of reporting of blinding of researchers to group status was the cause of 'Unclear' risk of bias ratings given to the remaining studies (n=10)

Index Test – Applicability Concerns

The applicability of the WMT was the area of greatest quality concern overall. Just three of the included studies were rated as 'Low' risk. Translation out of original language was given as a reason for 'High' risk ratings for six studies (Brockhaus & Merten, 2004; Elbaum et al., 2019; Kirchhoff & Steinert, 2019; Lupu et al., 2018; Martins & Martins, 2010;

Tydecks et al., 2006)[3]. Three studies reported the use of non-standardised cut-scores (Manderino & Gunstad, 2018; Ortega et al., 2013; Tan et al., 2002). The use of IR, DR and CNS subscales only resulted in the rating of four articles as 'Unclear' risk of bias.

Reference Standard – Risk of Bias

Overall, the included studies provided good information participant instructions and levels of coaching. Eleven studies were rated as 'Low' risk of bias in this domain. One study (Tan et al., 2002) was rated as 'High' risk due to the use of idiosyncratic instructions (i.e., participants undertaking their own research to prepare for the dissimulation condition).

Reference Standard – Applicability Concerns

Overall, most studies were rated as 'High' or 'Unclear' risk (n=13). Four studies (Lau et al., 2017; Lupu et al., 2018; Manderino & Gunstad, 2018; Shandera et al., 2010) cited use of multiple methods to check the adherence of participants to instructions, these were not described in a way sufficient to determine their likely effectiveness. Four studies were rated as 'Low' risk. This was due to the administration of an exit interview with participants to explore the strategies used to avoid detection on PVT's (Ortega et al., 2013) or conduct of a quiz/repetition of instructions to researchers to validate understanding.

Flow and Timing

The quality of studies in relation to the flow and timing of the WMT administration and participants across reported studies was variable. Seven studies were rated as 'High' risk due to the exclusion or removal of participants from analysis without an adequate reason being provided, but simply reporting being evaluated as not having followed instructions

[3] The author recognises that the WMT computer application has multiple languages and normative data has been produced for these languages however it is unclear from these papers whether this has been used. Due to issues of psychometric standardisation this was rated as 'high risk'. It has however been noted that studies have shown that the WMT can be passed by those whose first language is not English and achieve equivalent scores (Rienstra et al., 2009).

adequately (Elbaum et al., 2019; Kirchhoff & Steinert, 2019; Lupu et al., 2018; Merten et al., 2004; Ortega et al., 2013; Tan et al., 2002; Tydecks et al., 2006). This was corrected through ITT analysis. One study was also rated as 'High' risk due to the one-week time lag between assignment to simulation and control groups and administration of the WMT (Tan et al., 2002).

Summary of Bias Assessment

In summary, the quality of the studies within this meta-analysis was mixed. Only three of the 18 papers (Davis et al., 2010; Edmundson et al., 2017; Shandera et al., 2010) did not receive a rating of 'High' risk in any of the domains. Across the review there were substantial areas of bias and in the domain of index test and reference standard application. The quality of participant selection and therefore generalisability was variable across studies. On this basis the results of this meta-analysis warrant some caution due to the limitations in the quality and administration of the primary studies. Despite these concerns these papers reflect the currently available literature and these quality reviews provide some scope for future studies to expand and improve this area of the literature.

Meta-Analytic Syntheses

Univariate Indicators of Diagnostic Accuracy

The majority of PVT's apply a cut-off value to classify results as either positive or negative or pass or fail. This approach avoids some of the challenges associated with the continuous outcomes which may be obtained on PVT's (i.e., numerical score). Using dichotomous variables (pass/fail) enables comparison between a PVT and the reference standard. Common indicators of test performance are the sensitivity of the test, its specificity, the positive and negative predictive values, and the positive and negative likelihood ratios. These univariate test performance indicators are described in Table 7.

Test sensitivity and specificity are the most reported indicators of test accuracy. They relate to the proportion of test respondents who are likely to fail an PVT if they are dissimulating (sensitivity) or the proportion of test respondents who would pass a PVT if they are responding genuinely (specificity).

These statistics relate to the proportions of test responders correctly identified and do not inform the test administrator of the *probability* of dissimulation or genuine responding. The probability that a PVT failure indicates dissimulation should be calculated using Bayesian posterior probabilities with the population base rate for dissimulation as the apriori probability. Although the sensitivity and specificity of a PVT can be used to calculate these posterior probabilities, the sensitivity and specificity of a test do not directly provide information about the accuracy of an individual's test performance.

The positive and negative predictive power of a PVT is similarly focused on the proportional accuracy and presents the same interpretive limitations as sensitivity and specificity. Alternatively, the positive and negative likelihood ratios provide a measure of diagnostic accuracy independent of prevalence.

Table 7

Univariate test performance indicators

	Formula	Definition
True Positive	TP	The number of people faking bad who also obtain a positive test result
True Negative	TN	The number of genuine responders who obtain a negative test result
False Positive	FP	The number of genuine responders who obtain a positive test result
False Negative	FN	The number of people faking bad would obtain a negative test result
Sensitivity (true positive rate, TPR)	TP/(TP + FN)	Proportion positive test results among those faking bad
Specificity (true negative rate, TNR)	TN/(TN + FP)	Proportion negative test results among the valid responders
Positive predictive value (PPV)	TP/(TP + FP)	Proportion identified as faking bad among subjects with a positive test result
Negative predictive value (NPV)	TN/(TN + FN)	Proportion of genuine responders among subjects with a negative test result
Likelihood ratio of a positive test result (LR+)	$\frac{\text{sensitivity}}{(1-\text{specificity})}$	Ratio of a positive test result among dissimulators to the same result in the genuine responders
Likelihood ratio of a negative test result (LR-)	$\frac{(1-\text{sensitivity})}{specificity}$	Ratio of a negative test result among dissimulators to the same result in the genuine responders

Word Memory Test: Univariate Test Performance Indicators

Quantitative synthesis was undertaken for primary studies which reported sensitivity and specificity based on an Overall index on the WMT (n=13). Analysis of classification based upon the WMT subscales (IR, DR and CNS) will be reported subsequently. Three of the included studies (Manderino & Gunstad, 2018; Ortega et al., 2013; Tan et al., 2002)

reported WMT failure on the basis of non-standardised cut-offs (≤82.5% on the IR, DR or CNS subscales; Green, 2003) Analysis was conducted to control for variance in test interpretation. Psychometric data for the thirteen studies (reporting 25 effects) within this analysis can be found in Table 8.

Table 8

Overview of validity data reported by studies using WMT (Overall Index)

Study	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)
Manual Cut-Off (≤82.5%)					
Brockhaus & Merten (2009a)	0.995 (0.954 - 0.999)	0.982 (0.847 - 0.998)	54.726 (3.511 - 853.098)	0.005 (0.000-0.080)	10800.00 (209.431-556938.303)
Brockhaus & Merten (2009b)	0.995 (0.954 - 0.999)	0.969 (0.843 - 0.994)	31.841 (4.626 - 219.180)	0.005 (0.000-0.082)	6200.000 (203.105-189261.981)
Brockhaus & Merten (2009c)	0.995 (0.954 - 0.999)	0.983 (0.910 - 0.997)	58.706 (8.408 - 409.887)	0.005 (0.000 – 0.080)	11600.000 (383.226 – 351124.355)
Davis, Wall, Whitney & Barisa (2017a)	0.786 (0.605 – 0.898)	0.981 (0.842 – 0.998)	41.643 (2.656 – 652.930)	0.218 (0.107 – 0.445)	190.667 (10.076 – 3607.800)
Davis, Wall, Whitney & Barisa (2017b)	0.714 (0.529 – 0.847)	0.981 (0.842 – 0.998)	37.857 (2.407 – 595.457)	0.291 (0.162 – 0.524)	130.000 (7.040 – 2400.690)
Green, Lees-Haley & Allen (2002a)	0.976 (0.804 – 0.997)	0.976 (0.804 – 0.997)	40.000 (2.586 – 618.645)	0.025 (0.002 – 0.387)	1600.000 (30.239 – 84658.729)
Green, Lees-Haley & Allen (2002b)	0.960 (0.805 – 0.993)	0.988 (0.892 – 0.999)	77.760 (4.942 – 1223.454)	0.041 (0.006 – 0.276)	1920.000 (62.042 – 59417.848)
Kirchhoff & Steinert (2019)	0.400 (0.246 - 0.577)	0.933 (0.787 - 0.982)	6.000 (1.467 – 24.547)	0.643 (0.473 – 0.874)	9.333 (1.866 – 46.684)
Lau et al. (2017g)	0.976 (0.804 - 0.997)	0.976 (0.804 - 0.997)	40.000 (2.586 – 618.645)	0.025 (0.002 – 0.387)	1600.000 (30.239 – 84658.729)
Lau et al. (2017h)	0.925 (0.801 - 0.974)	0.976 (0.804 - 0.997)	37.925 (2.451 – 586.885)	0.077 (0.026 – 0.229)	493.333 (23.530 – 10343.307)
Lindstrom et al. (2009a)	0.920 (0.750 - 0.978)	0.980 (0.837 - 0.998)	46.920 (3.009 – 731.704)	0.082 (0.022 – 0.309)	575.000 (24.633 – 13422.125)
Lindstrom et al. (2009b)	0.920 (0.750 - 0.978)	0.920 (0.750 - 0.978)	11.500 (3.028 – 43.671)	0.087 (0.023 – 0.330)	132.250 (17.138 – 1020.551)
Lindstrom et al. (2009c)	0.920 (0.750 - 0.978)	0.960 (0.865 - 0.989)	23.000 (5.887 – 89.865)	0.083 (0.022 -0.315)	276.000 (36.540 – 2084.756)
Martins & Martins (2010a)	0.950 (0.764 - 0.991)	0.983 (0.857 - 0.998)	56.050 (3.583 – 876.819)	0.051 (0.008 – 0.344)	1102.000 (35.205 – 34495.189)
Martins & Martins (2010b)	0.950 (0.764 - 0.991)	0.333 (0.172 - 0.546)	1.425 (1.036 – 1.960)	0.150 (0.020 – 1.113)	9.500 (1.046 – 86.261)
Martins & Martins (2010c)	0.950 (0.764 - 0.991)	0.720 (0.583 - 0.825)	3.393 (2.151 – 5.352)	0.069 (0.010 – 0.473)	48.857 (5.962 – 400.377)
Osman, Plambeck, Klein & Mano (2006a)	0.643 (0.458 - 0.793)	0.964 (0.823 - 0.994)	18.000 (2.575 – 125.803)	0.370 (0.224 – 0.612)	48.600 (5.716 – 413.236)
Osman, Plambeck, Klein & Mano (2006b)	0.643 (0.458 - 0.793)	0.964 (0.823 - 0.994)	18.000 (2.575 – 125.803)	0.370 (0.224 – 0.612)	48.600 (5.716 – 413.236)
Osman, Plambeck, Klein & Mano (2006c)	0.643 (0.512 - 0.755)	0.964 (0.823 - 0.994)	18.000 (2.601 – 124.573)	0.370 (0.259 – 0.530)	48.600 (6.136 – 384.925)
Tan, Slick, Strauss & Hultsch (2002a)	0.920 (0.750 - 0.978)	0.964 (0.823 - 0.994)	25.760 (3.746 – 177.139)	0.083 (0.022 – 0.314)	310.500 (26.421 – 3649.050)
Tydecks, Merten & Gubbay (2006d)	0.818 (0.523 – 0.949)	0.952 (0.667 – 0.995)	17.182 (1.133 – 260.647)	0.191 (0.054 – 0.674)	90.000 (3.570 – 2268.738)
Cut off <90%					
Manderino & Gunstad (2018)	0.676 (0.582 - 0.758)	0.847 (0.773 - 0.900)	4.413 (2.858 – 6.814)	0.382 (0.287 – 0.509)	11.540 (6.103 – 21.822)
Ortega, Labrenz, Markowitsch & Piefke (2013a)	0.950 (0.764 – 0.991)	0.750 (0.531 – 0888)	3.800 (1.767 – 8.172)	0.067 (0.010 – 0.458)	57.000 (6.000 - 541.466)
Ortega, Labrenz, Markowitsch & Piefke (2013b)	0.950 (0.764 – 0.991)	0.976 (0.804 – 0.997)	38.950 (2.516 – 603.002)	0.051 (0.008 – 0.347)	760.000 (24.070 - 23996.314)

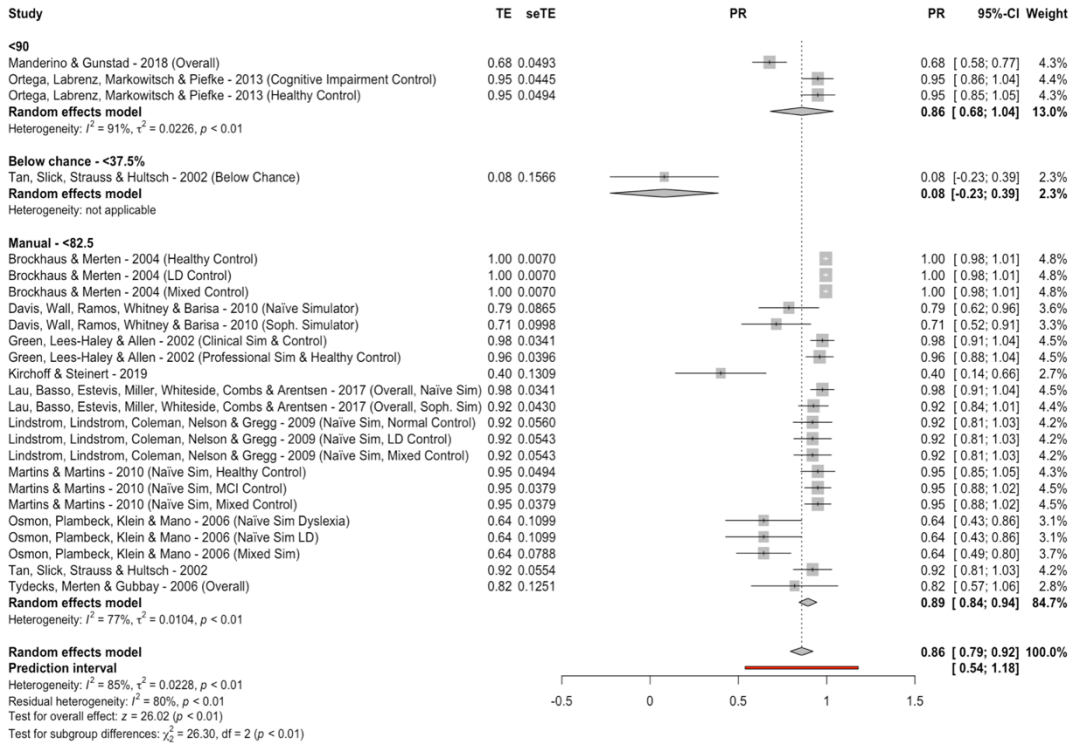
Study	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)
Below Chance Cut off - <37.5%					
Tan, Slick, Strauss & Hultsch (2002b)	0.080 (0.022 - 0.250)	0.964 (0.823 - 0.994)	2.240 (0.216 - 23.233)	0.954 (0.833 - 1.093)	2.348 (0.200 - 27.592)

Sensitivity and Specificity of the Word Memory Test

The random-effects model for sensitivity and specificity were calculated using the Generic Inverse Variance method. Figure 2 shows the forest plot for sensitivity. The model suggested an overall pooled sensitivity of 0.86 (z=26.02, p<0.01) with a 95% confidence interval between 0.79-0.92. A test of difference between subgroups according to classification cut-off identified significant differences in sensitivity (X^2_2=26.30, df=2, p <0.01). Within the studies reporting use of the standardised cut-off significant heterogeneity was identified (Higgins' I^2=77%, Q=88.56, df=20, p <0.01). Individual studies found to exert undue influence upon overall meta-analytic effect can be identified through use of a "leave one out" analysis. Removal of Kirchhoff & Steinert (2019) was found to hold the greatest influence on overall effect amongst studies using standard cut-offs which produced a revised I^2 of 72% (Q=68.39, df=19, p <0.01).

Figure 2

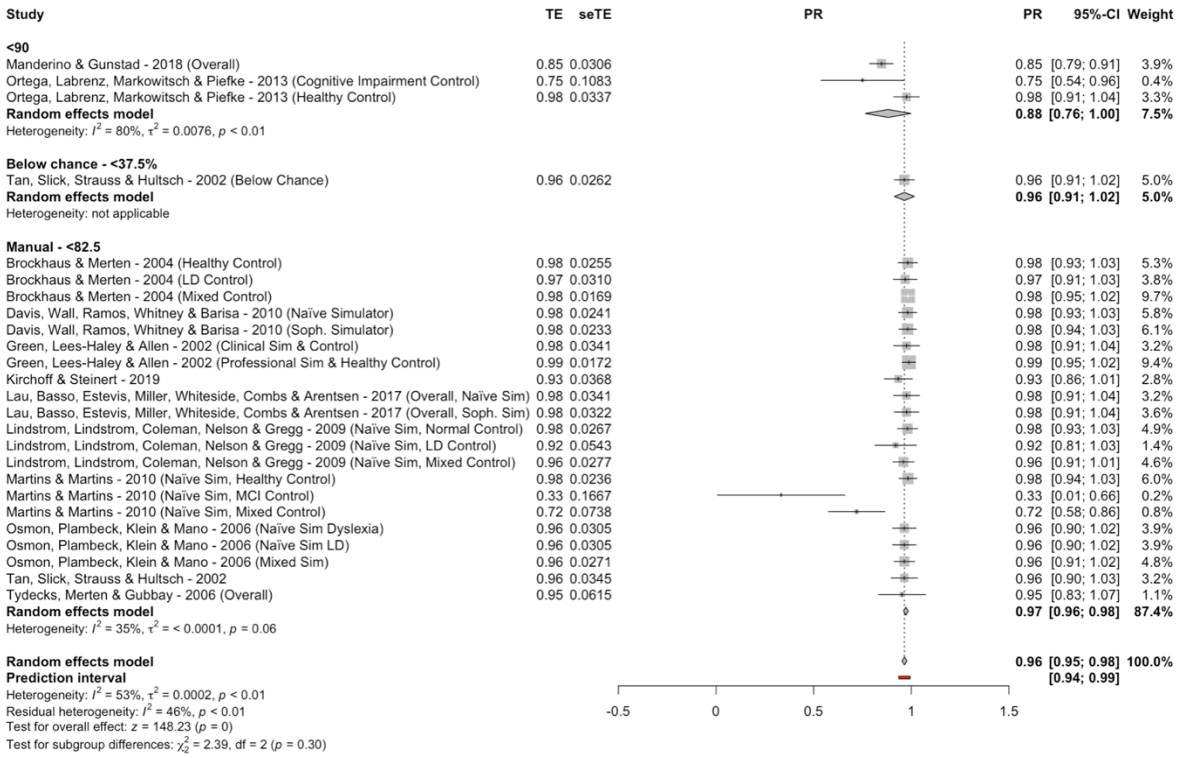
Forest Plot of Sensitivity for Overall Index by Cut-Off



The pooled specificity (Figure 3) demonstrates an overall test specificity of 0.96 (z=148.23, p=0, 95% CI 0.95-0.98). A moderate level of heterogeneity was observed within studies using the standardised classification cut-off (Higgins' I^2=35.4%, Q= 30.97, df=20, p=0.06). "Leave one out" analysis once again identified Kirchhoff and Steinert (2019) as holding greatest influence although the removal of this study did not result in a substantial reduction in heterogeneity (I^2=36.3%, Q=29.82, df=19, p=0.05).

Figure 3

Forest Plot of Specificity for Overall Index by cut-off



The use and independent reporting of univariate measures to describe the discriminatory performance of a PVT is subject to significant limitations. As previously summarised, both sensitivity and specificity and predictive power are dependent upon

reference to externally defined groups ('dissimulating'/'honest responding') without integration across these to acknowledge the presence of FP's (sensitivity) and FN's (specificity). At their heart sensitivity and specificity are highly correlated, with test sensitivity increasing as the specificity of the test decreases and vice versa.

An alternative is to examine the positive and negative likelihood ratios. These indicators provide a measure of test accuracy relative to a measure of test misclassification. The likelihood ratio of a positive test result (LR+) reports the ratio of a positive test result among dissimulators to the same result in genuine responders. The likelihood ratio of a negative test result (LR-) describes the ratio of a negative test result among dissimulators to the same result in genuine responders.

Likelihood ratio measures of the Word Memory Test

Figure 4 and Figure 5 present the meta-analysis of the positive and negative likelihood ratios for the WMT. The pooled positive likelihood in this meta-analysis was found to be 12.93 (95% CI 7.50–22.29) with no significant heterogeneity observed (Higgins' $I^{2=}$0%, Q=18.371, df=24, p=0.785). As the positive likelihood ratio increases, the probability of dissimulation if the WMT was positive increases, with a value above 10 considered a large and significant increase in probability (Egger, Davey-Smith & Altman, 2008). A value of 12.93 in this meta-analysis indicates a significant increase in the probability of dissimulation given a positive screen using the WMT. The pooled negative likelihood ratio was 0.153 (95% CI 0.101-0.232) and there was moderate heterogeneity noted (Higgins' I^2=29.06%, Q=33.83, df=24, p=0.088). A smaller negative likelihood ratio indicates a more informative test, with a value below 0.2 indicating strong evidence for the diagnostic test (Egger et al., 2008). Therefore, the value of 0.153 suggests good evidence for the value of the WMT in detecting dissimulation.

Figure 4

Forest Plot of log-transformed Positive Likelihood Ratios

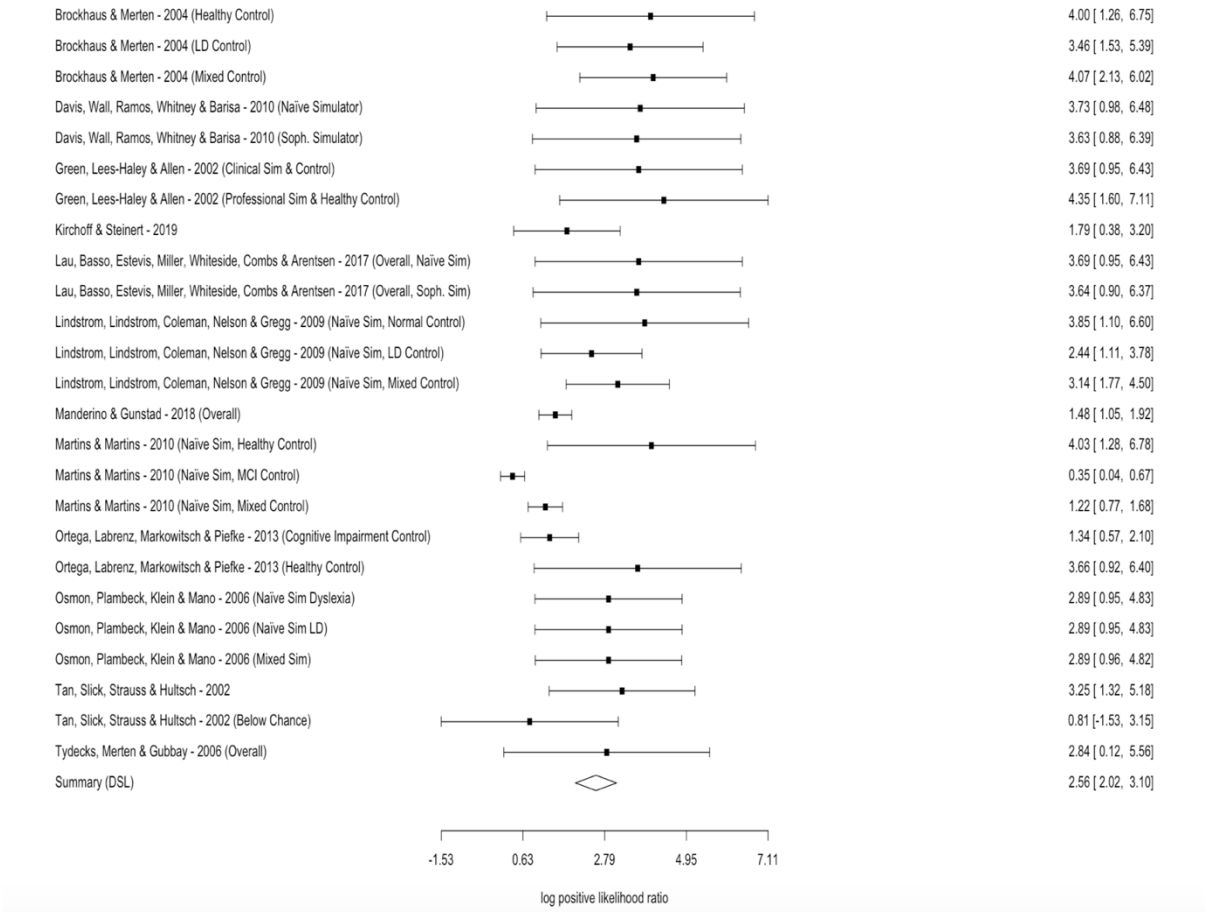
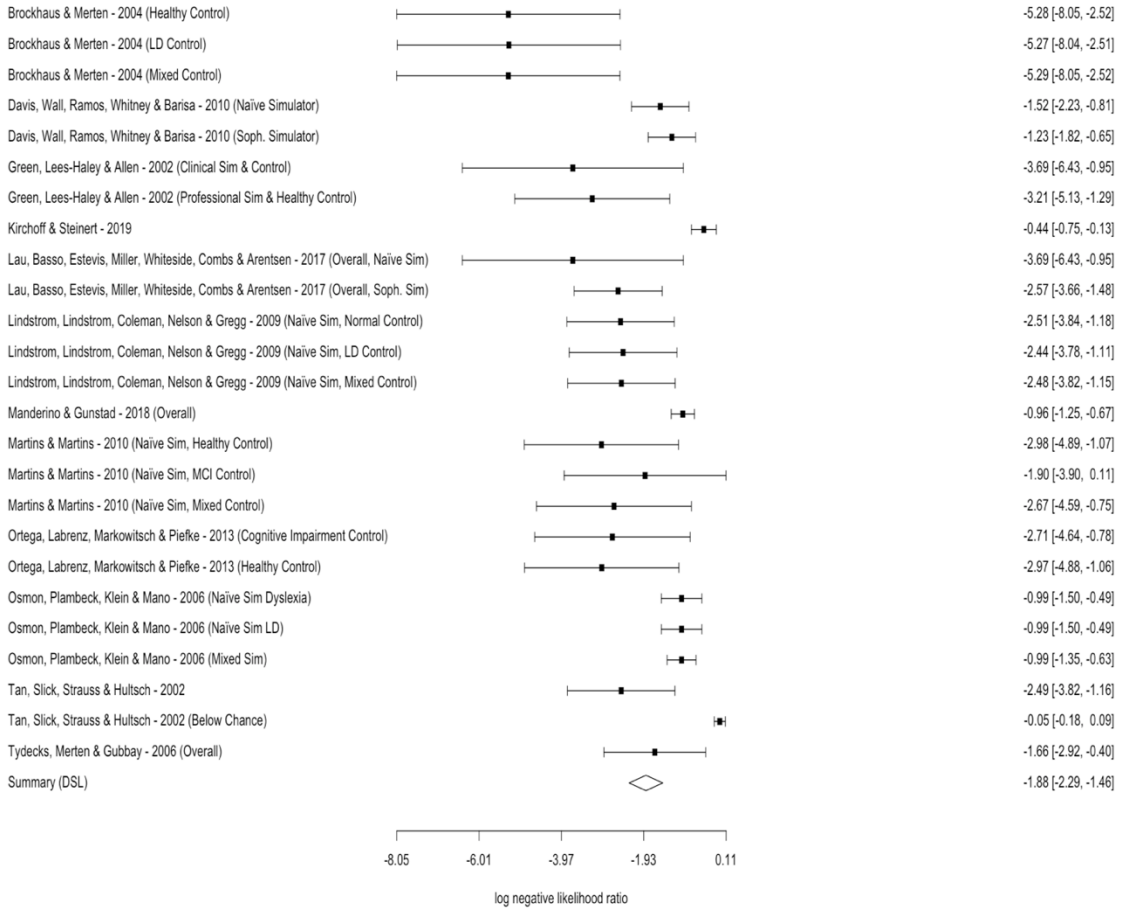


Figure 5

Forest Plot of log-transformed Negative Likelihood Ratios



Due to limitations in reporting of univariate measures, in order to review test accuracy based on a single index, a global measure of test performance is required (Table 9). “Overall accuracy” is obtained through the mathematical aggregation of sensitivity and specificity (or positive and negative predictive values and likelihood ratios) into one statement of the proportion of correctly identified subjects. Once again disease prevalence can impact upon the “overall accuracy” of a test, with sensitive tests obtaining greater accuracy in high prevalence populations and specific tests achieving this within low prevalence populations.

Table 9

Global measures of test performance

	Formula	Definition
Overall Accuracy	(TP + TN)/(TP + TN + FP + FN)	Proportion correctly identified subjects
Diagnostic Odds Ratio	$\frac{TP}{FN}\Big/\frac{FP}{TN} = \left(\frac{PPV}{(1-PPV)}\right)\Big/\left(\frac{(1-NPV)}{NPV}\right) = \frac{LR+}{LR-}$	DOR is the ratio of the odds of test failure in dissimulators relative to the odds of test failure in the honest responders. Alternatively, the DOR can be viewed as the ratio of the odds of dissimulation in PVT failers relative to the odds of dissimulation in PVT passers

Note: The abbreviations TP, FP, FN, and TN denote true positives, false positives, false negatives, and true negatives.

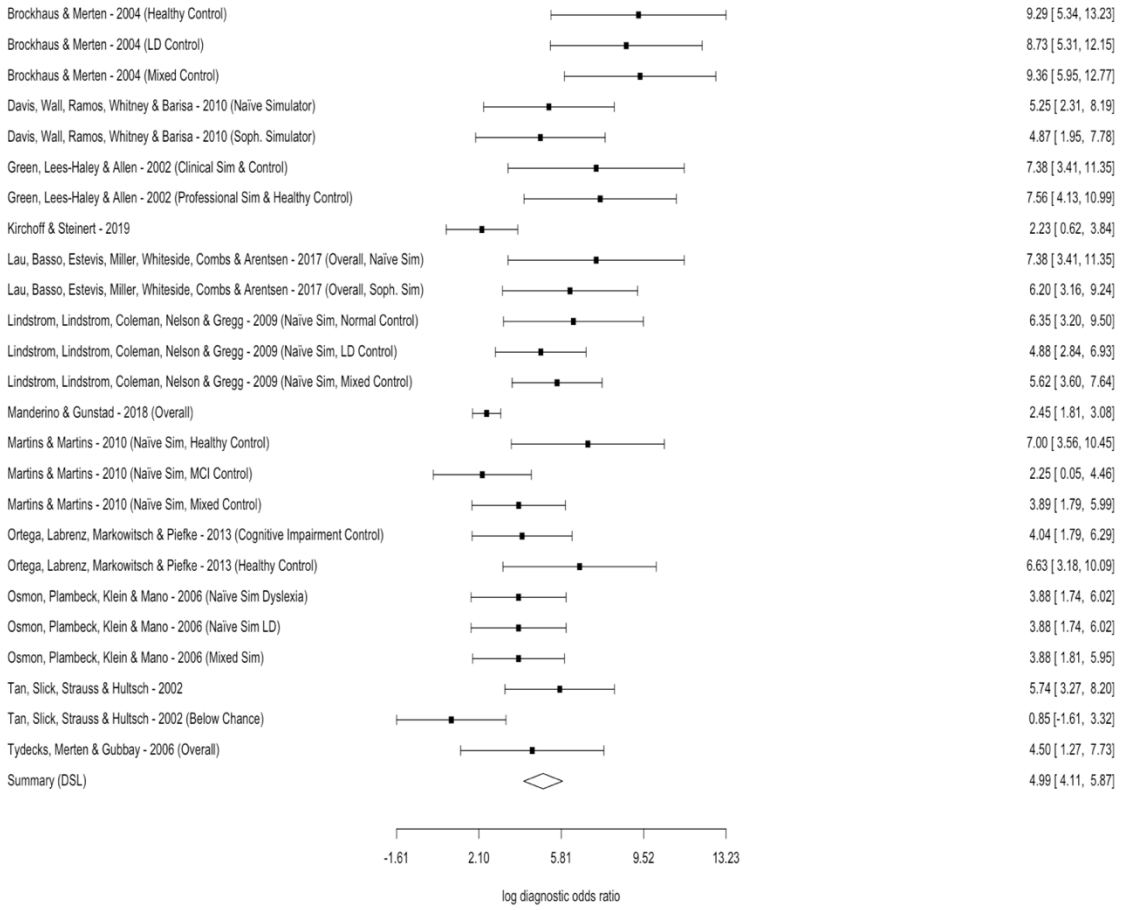
Diagnostic Odds Ratio of the Word Memory Test

The Diagnostic Odds Ratio (DOR) offers a measure of global test performance that is not dependent upon the prevalence of a condition (Glas et al., 2003). The DOR ranges from zero to infinity. A DOR of one indicates that a test is as likely to predict a positive outcome regardless of the condition (i.e., the test provides no information). For useful tests, a DOR is greater than one, with higher DOR's are indicative of better test performance.

Figure 6 represents the log transformed DOR for the twelve studies where test performance was classified based on an Overall index. Log transformation was undertaken to normalise the distribution and stabilise between study variance. The log(DOR) was found to be 4.99 (95% CI 4.11–5.87). Based on these results it is suggested that individuals who were identified by the WMT as simulating suboptimal performance are one hundred and forty-seven times more likely to be demonstrating suboptimal effort than control participants.

Figure 6

Forest plot of the log(Diagnostic Odds Ratio)



Subgroup Analyses

Control Group

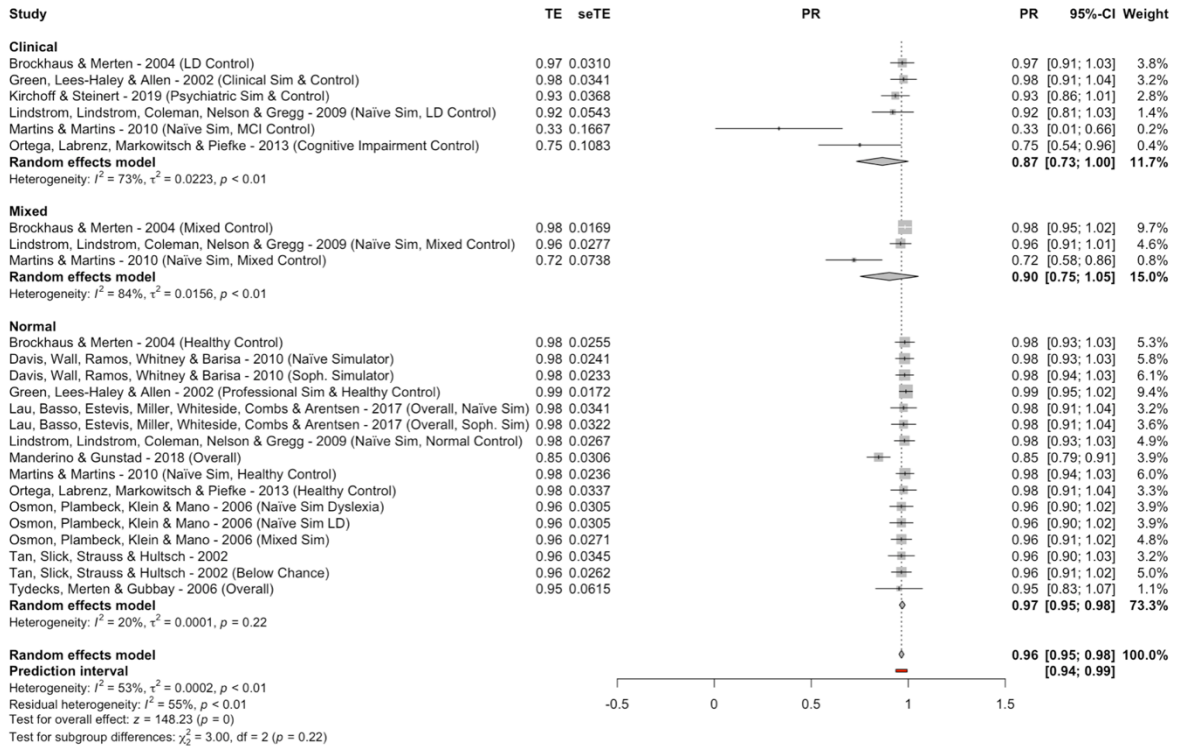
The selection of participants from patient and non-patient populations is suggested to have implications upon the specificity of the WMT[4]. To explore this, subgroup analyses of specificity and DOR were conducted on studies which used patient and non-patient (healthy) control groups. Across all cut-offs of the WMT analysis identified no significant difference in

[4] Sollman & Berry (2011) cite lower WMT specificity when patient control groups are used.

specificity between patient control groups (n=6, specificity=0.87, 95% CI 0.73-1.00) and normal controls (n=16, specificity=0.97, 95% CI 0.95-0.98)[5]. Findings suggest that individuals identified as displaying suboptimal effort were ninety times more likely to be identified as dissimulating than patient controls (DOR=90.645, 95% CI 14.416–569.965) and one hundred and fifty times more likely than non-patient controls (DOR=150.82, 95% CI 48.760–466.492).

Figure 7

Forest plot of Specificity for Overall index for all cut-offs by Control population



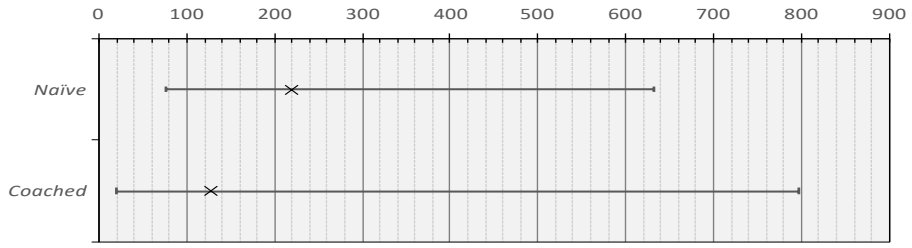
[5] Despite the absence of significant difference, it was noted that this specificity value falls below the typically accepted value of 0.9 for a PVT. Further scrutiny identifies that the low identification cited by Martins & Martins (2010) may contribute to these values. It was noted however that when the authors applied the GMIP to their results, Martins & Martins (2010) obtained a specificity of 0.95.

Simulator Coaching and Motivation

As analogue studies have been critiqued for using potentially unmotivated and un-coached participants in the dissimulation condition, subgroup analyses were conducted on studies which used naïve and coached participants (Figure 8) and those who were unmotivated or received reward contingent upon performance (Figure 9). As might be expected, there was greater detection accuracy in the naïve simulators, however, there was no significant difference in effects between naïve (n=17; DOR=176.651, 95% CI 60.464-516.107) and coached simulators (n=8; DOR=104.875, 95% CI 18.713-587.768) on the WMT.

Figure 8

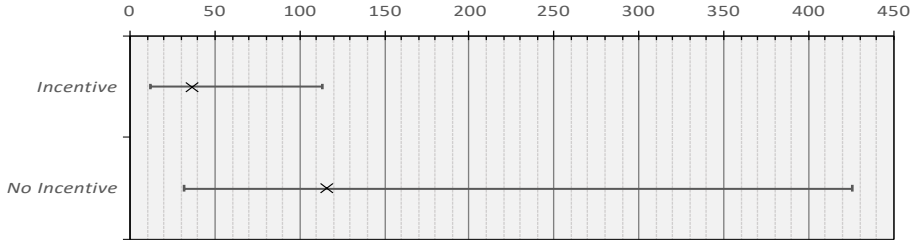
Subgroup analysis of diagnostic odds ratios and confidence intervals for Naïve and Coached simulators from studies reporting failure at standard (≤82.5%) cut-off



Comparison of the DOR and confidence intervals of studies offering incentives (n=7) and non-incentivised simulators (n=6) was undertaken. There was no significant difference in effect observed between incentivised (DOR=39.925, 95% CI 14.662–108.716) and non-incentivised simulators (DOR=115.852, 95% CI 31.483-426.311).

Figure 9

Subgroup analysis of diagnostic odds ratios and confidence intervals for incentivised and non-incentivised simulators from studies reporting failure at standard (≤82.5%) cut-offs



A DOR, whilst not influenced by disease prevalence in the same way as other univariate measures also holds limitations in its accuracy. Due to its dependency upon sensitivity and specificity of the test in question, the obtained DOR is therefore subject to change according to the criteria used to define dissimulation (i.e., the threshold effect). In addition, the DOR cannot be used to determine the TP and the TN rates of a test. It is therefore possible for two tests with the same DOR to possess quite different sensitivities and specificities.

Bivariate indicators of Diagnostic Accuracy

As cited above, each of the univariate meta-analysis strategies present with interpretive difficulties. To address some of this critique the bivariate random effects model was calculated. The Reitsma bivariate model has been demonstrated to be equivalent to the HSROC linear mixed effects model proposed by Rutter & Gatsonis (2001; Harbord et al., 2007). From this, an average trajectory of point estimates (weighted with a measure of study precision) for the original data range can be obtained which is then extrapolated for values beyond the original data set.

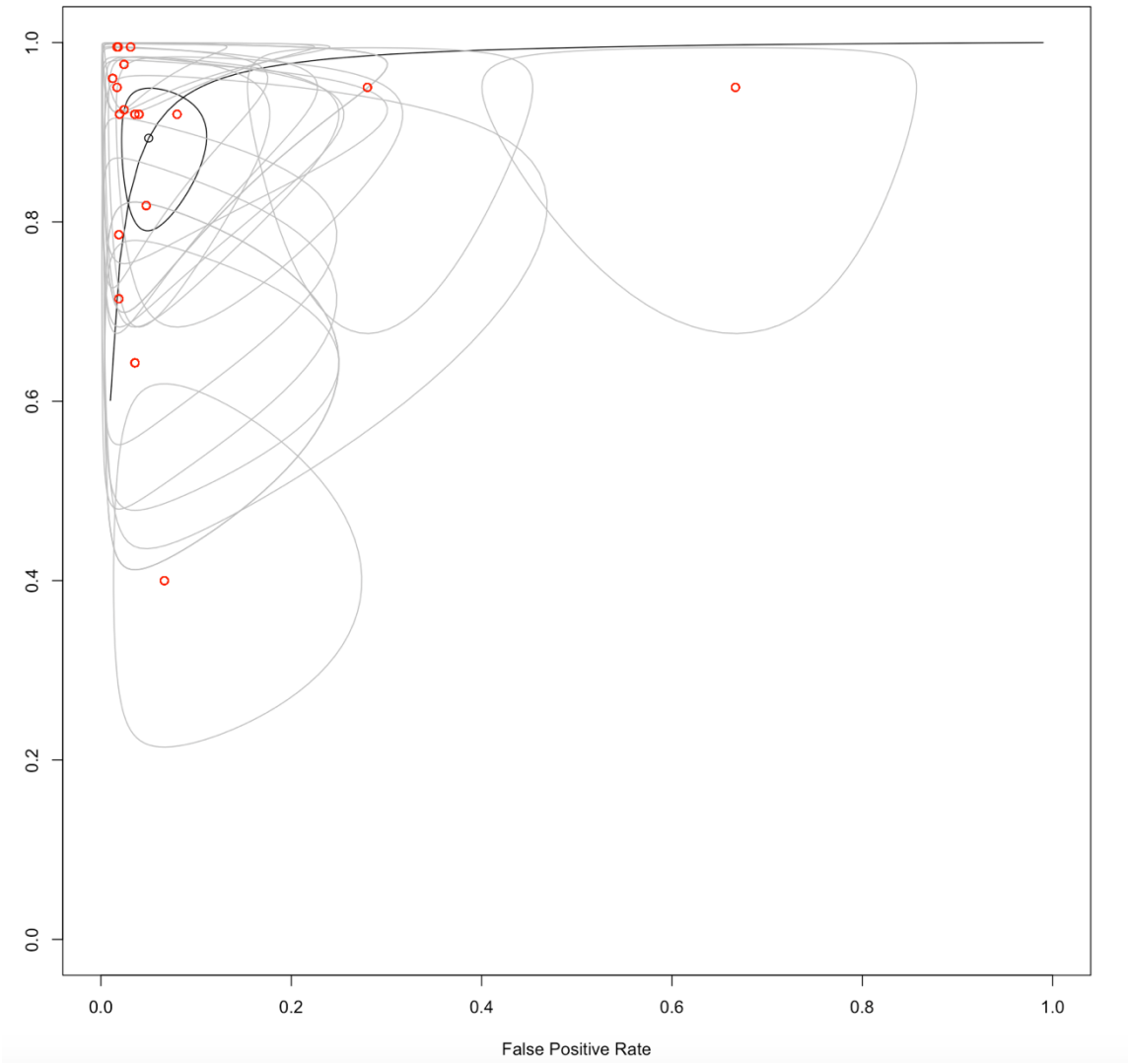
Summary Receiver Operating Characteristic of the Word Memory Test

The bivariate random effects model considers the effects of sensitivity and specificity simultaneously the calculation of a receiver operator characteristic (ROC) curve, plotting TP rate as a function of the FP rate for all possible cut-off values of the test. The obtained area under the curve (AUC) can used as an alternative global indicator of test performance and can be interpreted as the average sensitivity across all possible specificities.

The bivariate random effects meta-analysis for studies utilising the standardised cut-off indicated an AUC value of 0.97 (extrapolated across all possible FP rates) and a partial area under the curve of 0.963 (restricted to the observed FP rates). AUC values closer to 1 are deemed indicative of high test accuracy (Glas et al., 2003), therefore a value of 0.97 indicates that the WMT has excellent overall diagnostic accuracy at the recommended cut-off of 82.5% or below (Figure 10). This AUC value demonstrated little variation when calculated for studies using a patient and non-patient control group (0.96 and 0.98 respectively).

Figure 10

sROC plot of Sensitivity and False Positive Rate for studies reporting Overall Index failure at Standard cut-off



The SROC curve indicates that the average sensitivity at all levels of specificity was 0.893 (95% CI 0.815 – 0.941) and the average specificity at all levels of sensitivity was 0.95 (95% CI 0.90–0.974). Table 10 presents the positive and negative likelihood ratio and DORs through the Markov Chain Monte Carlo (MCMC; Zwindermann & Bossuyt, 2008) procedure

which are acknowledged to be more accurate than estimate derived from univariate meta analyses (Reitsma et al., 2005).

Table 10

Bivariate random effects model for positive and negative likelihood ratios and diagnostic odds ratio

	Mean	Median	95% Confidence Interval
Positive Likelihood Ratio	18.800	17.70	9.320 – 34.500
Negative Likelihood Ratio	0.117	0.113	0.0624 – 0.195
Diagnostic Odds Ratio	177.000	159.000	62.700 – 401.000

Analysis of Individual Sub-Tests of Word Memory Test

Table 11 displays pooled univariate and bivariate measures of diagnostic accuracy for the three primary effort sub-tests of the WMT at or below 82.5% cut-off. A total of seven primary studies (10 data sets) reported classification based on the Immediate Recall portion of the WMT (Appendix 4 & 5 contain Tables and sROC charts). Two studies did not report Delayed Recall and Consistency indices (Elbaum et al., 2019; Lupu et al., 2018) and therefore were not used for meta-analysis of these subtests.

Table 11

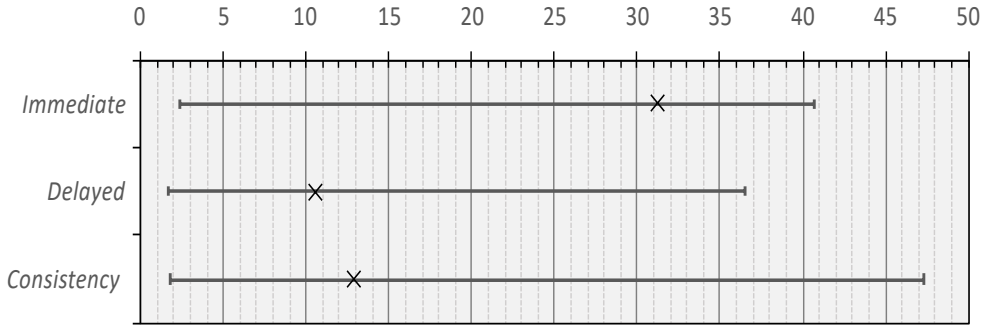
Bivariate random effects model for individual sub-tests of the WMT

	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)	AUC
Immediate (n = 10)	0.773 (0.602 – 0.885)	0.9 (0.74 – 0.966)	9.050 (2.730 – 23.600)	0.267 (0.126 – 0.474)	42.000 (6.470 – 145.000)	0.9
Immediate (n = 8)	0.698 (0.534 – 0.824)	0.885 (0.661 – 0.968)	7.580 (1.920 – 22.500)	0.359 (0.196 – 0.589)	25.100 (3.510 – 90.000)	0.829
Delayed (n = 8)	0.600 (0.426 – 0.751)	0.885 (0.661 – 0.968)	6.450 (1.650 – 19.100)	0.471 (0.279 – 0.721)	15.800 (2.390 – 55.000)	0.769
Consistency (n = 8)	0.761 (0.604 – 0.870)	0.823 (0.52 – 0.952)	5.370 (1.530 – 15.900)	0.315 (0.156 – 0.569)	21.000 (2.870 – 76.300)	0.834

The meta-analysis suggests that the Immediate Recall subtest holds the strongest diagnostic accuracy across reported studies however the calculated likelihood ratios and DOR's did not necessarily indicate strong evidence for the independent validity of the WMT indices. Due to limitations in the data reporting across primary studies it is not possible to look at failure on the basis of any one, two or three of the effort subtests, as the test would be used in practice. Bivariate analysis provided AUC values between 0.751 – 0.892 offering a moderate (DR) to strong (IR) diagnostic accuracy. Despite the apparent decrease in effectiveness of the DR sub-test compared to the IR sub-test, subgroup analysis did not find significant differences in obtained DOR's (Figure 11). Removal of the two additional studies which used the IR alone produced a revised DOR of 16.400 (95% CI 2.150–60.00).

Figure 11

Subgroup analysis of diagnostic odds ratios and confidence intervals for WMT sub-tests



Discussion

Summary of findings

This meta-analysis examined the diagnostic accuracy of the Word Memory Test (WMT) in detecting dissimulation within analogue studies. A total of 57 datasets were identified as reporting diagnostic accuracy of the WMT across different motivational and coached conditions and indices of WMT performance. Quantitative synthesis of the data using both univariate and bivariate indicators of test performance was undertaken

Univariate meta-analysis indicated that the WMT provides an accurate measure of dissimulation in analogue malingerers when test failure is determined according to an Overall index. A pooled DOR of 146.986 (95% CI 61.138 – 353.378) indicates that an individual who obtains a positive WMT result (identified as providing suboptimal effort) is one hundred and forty-seven times more likely to be dissimulating than a control participant. Using the more accurate bivariate model and restricted to failure as classified according to standard ≤82.5% cut-off, this DOR increases to 177.000 (95% CI 62.700 – 401.000).

Measures of positive and negative likelihood ratio also demonstrated favourable diagnostic properties for the WMT. The WMT (≤82.5% cut-off) was found to have a pooled positive likelihood ratio of 18.800 (95% CI 9.320 – 34.500) indicating a significant probability of dissimulation when a positive result is obtained, well above the advised threshold value of 10 cited by Egger and colleagues (2008). Likewise, the pooled negative likelihood ratio of 0.117 (95% CI 0.062 – 0.195) suggests that it is ten times less likely that an individual who is dissimulating will not be identified as doing so (FN).

Bivariate meta-analysis produced an overall sensitivity of 0.89 and specificity of 0.95. An area under the curve of 0.97 showed high levels of diagnostic accuracy. These findings

suggest that the WMT can be used as a reliable and accurate measure for the detection of dissimulation of neurocognitive deficit between groups of experimentally created groups.

Subgroup analyses found no significant difference in diagnostic capabilities of the WMT (standard cut-off) between coached and naïve simulator groups, although the presence of coaching did reduce detection in the expected direction. This lack of significance likely reflects the large variation in performance of naïve simulators across studies. Likewise, the observed decrease in detection accuracy for incentivised simulators was also not deemed significant.

No significant differences were identified in the detection of simulators against patient (DOR 90.645, 95% CI 14.416 – 569.965) and non-patient control groups (DOR=150.82, 95% CI 48.760 – 466.492). Limited conclusions can be drawn due to the absence of severe impairment groups within study participants. Whilst the WMT has demonstrated good specificity in LD, anxiety and depression and MCI control populations, the absence of control groups containing individuals with severe impairment (e.g., dementia) is a limitation of the available literature.

On exploration of the diagnostic accuracy of individual WMT indices the IR sub-test was identified as holding strong diagnostic ability across the ten studies which reported performance on this basis. According to bivariate modelling DR score displayed a lower accuracy with a DOR of 10.600 (95% CI 1.660 – 36.500) compared to DOR 31.200 (95% CI 2.380 – 21.700) for IR sub-test. Whilst this result appears anomalous, it was not found to be significant. Only eight of the included studies reported classification based on DR results. As demonstrated in the removal of the additional studies reporting the IR sub-tests alone, it is possible that publication of further studies which report the use of the DR sub-test may produce more consistent results. It is important to note that the test publishers do not

advocate the use of any of the test sub-tests in isolation and failure on the WMT is classified as a score of 82.5% or below on any one of the effort subtests.

Review of the quality of the included studies highlighted many areas of concern in relation to bias in the areas of index test and application of the reference tests. As discussed, the quality of participant selection and therefore generalisability was variable across studies which is reflective of the currently available literature. The results of this meta-analysis may warrant some caution in their interpretation however, this may provide some indication of the scope for future studies to expand and improve upon.

Comparison to other reviews

To the authors' knowledge, this is the first review to quantitatively synthesise the diagnostic accuracy of the WMT within an analogue sample, and it is therefore difficult to draw comparisons to other reviews which have explored the WMT validity. The results obtained in this review do however appear to indicate greater sensitivity and specificity than those previously reported (75.1% and 69.4% respectively; Sollman & Berry, 2011). This was reported on the basis of known-group and simulation research methodologies. Secondly, there are limited studies within Sollman & Berry's analysis which likely requires updating to reflect the developing literature. Two studies not included in the present meta-analysis were excluded due to being dissertations not within published journals, and one due to data not being presented in a way suitable for extraction (Batt et al, 2008).

A recently published meta-analysis of the Test of Memory Malingering (TOMM; Martin et al., 2019) explored the diagnostic accuracy of the TOMM within adult populations. Analysis of simulator/control samples demonstrated that individuals instructed to simulate impairment performed worse on the TOMM than individuals with known-impairment and incentive to

dissimulate (sensitivity=0.41-0.66; Martin et al., 2019). In comparison to the TOMM, the WMT presents with greater sensitivity and similar levels of specificity.

Strengths and limitations of current review

In addition to being the first known meta-analysis of the diagnostic accuracy of the Word Memory Test, a strength of the current review is in the use of the bivariate random-effects analysis model, this has been deemed the optimum method of meta-analysis (Reitsma et al., 2005).

It is also worth acknowledging that this review may be subject to methodological limitations. Data extraction and quality review were undertaken by the author alone which may therefore introduce some bias into the results. Due to this review being undertaken as part of a doctoral thesis there was limited capacity for obtaining a second opinion.

Furthermore, the variable quality of the included studies may introduce further limitations in the ability to interpret results as representative of the true performance of the Word Memory Test. It may also be noted that several studies which may be able to provide evidence of the diagnostic accuracy and validation of the WMT were not published in format which would have warranted inclusion in this meta-analysis.

As previously highlighted, the extraction of duplicated data for samples which employed multiple simulator or control groups may have led to increased risk of artificial inflation within the overall sample size thus reducing the obtained confidence intervals. The implementation of an ITT analysis for studies who removed participants who did not perform/attempt to simulate impairment is contrary to suggestions made by Bianchini and colleagues (2001) that inclusion of participants in analysis whom report minimal attempt to dissimulate in debriefing may increase risk of highlighting inaccuracy in group assignment and underestimating test sensitivity.

Implications for clinical practice and future research

Analysis of the diagnostic accuracy of a test is dependent upon comparison of the results of the index test, with those of an accepted reference standard (Trikalinos & Balion, 2012). Within the field of performance validity, the absence of a universally accepted "gold standard" for identifying dissimulation produces significant difficulty in determining diagnostic accuracy (Millis, 2009). Indeed, in the analysis and validation of PVT's, the WMT has often been used as one of the reference standards for determining dissimulation status (e.g., Henry et al., 2014; McCaul et al., 2018).

Using experimentally-created groups offers the opportunity to look at diagnostic accuracy against a clearly operationalised group status (i.e., simulator or honest performance). This review focused upon studies which used the WMT as a measure of detecting dissimulation between experimental groups. Whilst analogue research is important in the development of PVT's, such designs carry limitations in their generalisability as it is unknown if the persons in those groups behave similarly to those who would dissimulate neurocognitive deficit within clinical settings (Heilbronner et al., 2009). Sollman & Berry (2011) report that patient simulator-normal control contrasts found the largest mean effect size difference, significantly greater than known-group and normal simulator-patient control designs, although they did not quote an effect size.

To overcome these limitations, use of known-impairment groups has become increasingly prevalent in the literature pertaining to WMT specificity. Whilst extensive, this research continues to have trouble in defining participants as displaying dissimulation, relying on the use of other PVT's as per the previously discussed Slick criteria (1999). This creates a circularity in the categorisation of known-groups - a limitation of previous research. This testing is often conducted in settings where participants may be deemed to have greater motivation to feign impairment (e.g., forensic or litigation settings). One alternative method of

exploring specificity in known-impairment groups may be through the comparison of WMT fail rates between groups with high levels of expected impairment (e.g., dementia) against alow anticipated impairment group such as mild TBI (Flaro et al., 2007; Green & Flaro, 2019). Interpreting failure on the WMT as indicative of dissimulation is dependent on the consistency of these fail rates with the validated impairment. It may be hypothesised that a lower positive detection rate within a severe impairment group (without motivation to dissimulate) would suggest the overall specificity of the WMT and reinforces the interpretation of this test as a measure of effort and not one of cognition. Future research may look to synthesise findings of known-group literature to explore the diagnostic accuracy of the WMT.

As stated, this analysis does not include synthesis of studies reporting use of the Genuine Memory Impairment Profile (GMIP). When using the WMT in practice failure to consider the presence of GMIP with severely impaired groups will impact upon diagnostic accuracy. It is suggested that using the GMIP as an exclusion criterion for poor validity can increase the WMT specificity in discriminating between severely impaired groups and simulators (e.g., Green et al., 2011), but that this may be at the expense of sensitivity. Whilst Martins & Martins (2010) noted a specificity increase to 0.95 when using GMIP, it was noted that 10% of simulators (n=2) were able to produce this profile, resulting in their misclassification. Literature has reported that up to 50% of simulators can produce a GMIP on at least one out of three PVT's (including the WMT) but none on all three administered (Armistead & Denney, 2014). This appears to suggest a mitigating role for the use of multiple PVT's to increase sensitivity which is supported by Sherman et al (2020). Whilst the so-called 'chaining' of multiple PVT's to increase sensitivity has been subject to significant

debate[6] (see Fredericks, 2015 & Larrabee et al., 2019 for further discussion), updated criteria for malingered neurocognitive dysfunction cites failure of two or more PVT's if small number of tests administered but advises consideration of the ratio of test failures to tests administered as the latter increases (Sherman et al., 2020).

The recommendation from test authors, and this current analysis would be for future studies to employ the full WMT, and to document the easy-hard subtest scores of individual patients, as well as their failure on the IR, DR or CNS subscales. The test publisher further states that in the presence of a GMIP, there should be corroborative evidence of significant impairment or disability (Green et al., 2005; Sherman et al., 2020). The use of the GMIP with groups with severe impairment requires further exploration particularly relating to the potential effect of decreasing sensitivity when the WMT is used in isolation. At this juncture the available literature either did not administer or did not report the full WMT profile with sufficient frequency to allow the application of meta-analytic methodology.

PVT literature and this meta-analysis is predicated on the assumption that healthy individuals with no neurological conditions and those who have experienced mild TBI (mTBI) should not be failing PVT's. Reasons for PVT failure away from intentional poor performance have been proposed to include diagnosis threat, perceived injustice, and expectation for recovery. However, in a recent exploration of the role of self-efficacy and health locus of control in individuals with mTBI found no associations between these constructs and failure on PVT's (Armistead-Jehle et al., 2020). It is suggested that greater understanding of factors

[6] Fredericks (2015) suggests that interpretation of the likelihood of a person who fails multiple PVTs as demonstrating suboptimal effort will differ depending upon proportion of tests administered to failed. However, others have cited the importance of considering all possible combinations of pass/fail to influence how criterion groups are defined (e.g., Larrabee et al., 2019).

influencing PVT performance outside of motivation/conscious intention may increase the validity of neuropsychological evaluation of clinical presentations.

Given the potential implications attached to the mistaken attribution of poor test performance to suboptimal effort, it is important that clinicians access reliable and valid measures of performance validity. Fundamentally, literature continues to cite the importance of clinician test selection including the use of both stand-alone and embedded PVT's (Heilbronner et al., 2009; Miller et al, 2017) and the critical appraisal of reported test accuracy in the population in question (Lippa, 2018). The widespread implementation and use of PVT's must therefore come with clear guidelines for clinicians to support them in this practice and the acknowledgement that not all PVT's are created equally.

Conclusions

This meta-analysis has shown that when used in line with guidance from test publishers, the WMT does demonstrate strong diagnostic accuracy in the detection of dissimulation within analogue study designs. However, there is need identified for improvements in the quality of data reporting to enable thorough exploration and synthesis of literature pertaining to the WMT.

References

Armistead-Jehle, P., & Denney, R. L. (2014). The Detection of Feigned Impairment using the WMT, MSVT and NV-MSVT. *Applied Neuropsychology: Adult, 22*(2), 147-155. https://doi.org/10.1080/23279095.2014.880842

Armistead-Jehle, P., Lippa, S. M., & Grills, C. E. (2020). The Impact of Self-Efficacy and Health Locus of Control on Performance Validity Testing. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acaa027

Batt, K., Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society*, *14*(6), 1074-1080. https://doi.org/10.1017/S135561770808137X

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom Validity Testing: A Critical Review. *The Clinical Neuropsychologist, 15*(1), 19-45. https://doi.org/10.1076/clin.15.1.19.1907

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28*(13-14), 1623-1638. https://doi.org/10.3109/02699052.2014.947627

Brockhaus, R., & Merten, T. (2004). Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test. *Nervenarzt, 75*, 882-887. https://doi.org/10.1007/s00115-003-1592-4

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*, 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., Irwig, L., Levine, D., Reitsma, J. B., de Vet, H. C. W., & Bossuyt, P. M. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open, 6, e*012799. https://doi.org/10.1136/bmjopen-2016-012799

Davis, J. J., Wall, J. R., Ramos, C. K., Whitney, K. A., & Barisa, M. T. (2010). Using Grip Strength Force Curves to Detect Simulation: A Preliminary Investigation. *Archives of Clinical Neuropsychology, 25,* 204-211. https://doi.org/10.1093/arclin/acq002

Devillé, W. L., Buntinx, F., Bouter, L. M., Montori, V. M., de Vet, H. C., van der Windt, D. AWM., & Bezemer, D. (2002). Conducting systematic reviews of diagnostic studies:

didactic guidelines. *BMC Medical Research Methodology, 2* (9), 1-13. https://doi.org/10.11861471-2288-2-9

Edmundson, M., Berry, D. T. R., Combs, H. L., Brothers, S. L., Harp, J. P., Williams, A., Rojas, S. L., Saleh, A.)K., & Scott, A. B. (2017). The Effects of Symptom Information Coaching on the Feigning of Adult ADHD. *Psychological Assessment, 29* (12), 1429-1436. https://doi.org/10.1037/pas0000478

Egger, M., Davey-Smith, G., & Altman, D. G. (2008). *Systematic Reviews in Health Care: Meta-Analysis in Context* (2nd ed.). BMJ.

Elbaum, T., Golan, L., Lupu, T., Wagner, M., & Braw, Y. (2019). Establishing supplementary response time validity indicators in the Word Memory Test (WMT) and directions for future research. *Applied Neuropsychology: Adult.* https://doi.org/10.1080/23279095.2018.1555161

Fazio, R. L., Sanders, J. F., & Denney, R. L. (2015). Comparison of the Test of Memory Malingering and Word Memory Test in a criminal forensic sample. *Archives of Clinical Neuropsychology,* 30(4), 293-301. https://doi.org/10.1093/arclin/acv024

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury*, *21*(4), 373–383. https://doi.org/10.1080/02699050701311133

Frederick, R. (2015). *Too much information: Problems with using multiple malingering tests [Invited lecture].* American Psychology-Law Conference, San Diego, CA.

Gillard, N. (2010). *Methodological issues in malingering research: The use of simulation designs.* University of North Texas

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The Diagnostic Odds Ratio: A Single Indicator of Test Performance. *Journal of Clinical Epidemiology, 56*(11), 1129-1135. https://doi.org/10.1016/s0895-4356(03)00177-x

Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word Memory Test performance in amnesic patients with hippocampal damage. *Neuropsychology, 23*(4), 529 – 534. https://doi.org/10.1037/a0015444

Green, P. (2003) *Green's Word Memory Test for Microsoft Windows: User's Manual.* Green's Publications Incorporated

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A User's Guide to the Oral and Computer Administered Forms*, US Version 1.1.

Green, P., & Flaro, L. (2019). Performance validity test failure predicts suppression of neuropsychological test results in developmentally disabled children. *Applied Neuropsychology: Child,* 1–17. https://doi.org/10.1080/21622965.2019.1604342

Green, P., Lees-Haley, P. R., & Allen III, L. M. (2003). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, *2*(3-4), 97-124.

Green, P., & Merten, T. (2013). Noncredible explanations of noncredible performance on symptom validity tests. In Bush, S. S., & Carone D. A. (Eds). *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp 73-100). Springer Publishing.

Green, P., Montijo, J., & Brockhaus, R. (2011). High Specificity of the Word Memory Test and Medical Symptom Validity Test in Groups with Severe Verbal Memory Impairment. *Applied Neuropsychology, 18*(2), 86-94. https://doi.org/10.1080/09084282.2010.523389

Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P., & Sterne, J. A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*, *8*(2), 239-251. https://doi.org/10.1093/biostatistics/kxl004

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias and Malingering. *The Clinical Neuropsychologist, 23*(7), 1093-1129. https://doi.org/10.1080/13854040903155063

Henry, G. K., Heilbronner, R. L., Mittenberg, W., Hellemann, G., & Myers, A. (2014). Development of the MMPI-2 cognitive complaints scale as an embedded measure of symptom validity. *Brain Injury, 28*(3), 357-363. https://doi.org/10.3109/02699052.2013.865272

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal, 327*, 557-560.

Jones, C. M., & Athanasiou, T. (2005). Summary Receiver Operating Characteristic Curve Analysis Techniques in the Evaluation of Diagnostic Tests. *The Annals of Thoracic Surgery, 79*(1), 16-20. https://doi.org/10.1016/j.athoracsur.2004.09.040

Kirchhoff, L., & Steinert, T. (2019). Klassifikationsgüte der Beschwedevalidierungstests Word Memory Test und Strukturierter Fragebogen Simulierter Symptome. *Nervenarzt, 90*, 516-522. https://doi.org/10.1007/s00115-018-0637-7

Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist, 33*(8), 1354-1372. https://doi.org/10.1080/13854046.2019.1614227

Larsen, J. D., Allen, M. D., Bigler, E. D., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the Word Memory Test. *Brain Injury, 24*(2), 89-99. https://doi.org/10.3109.02699050903508218

Lau, L., Basso, M. R., Estevis, E., Miller, A., Whiteside, D. M., Combs, D., & Arentsen, T. J. (2017). Detecting coached neuropsychological dysfunction: a simulation experiment regarding mild traumatic brain injury. *The Clinical Neuropsychologist, 31*(8), 1412-1431. https://doi.org/10.1080/13854046.2017.1318954

Lindstrom, W. A., Lindstrom, J. H., Coleman, C., Nelson, J., & Gregg, N. (2009). The Diagnostic Accuracy of Symptom Validity Tests when Used with Postsecondary Students with Learning Disabilities: A Preliminary Investigation. *Archives of Clinical Neuropsychology, 24*, 659-669. https://doi.org/10.1093/arclin/acp071

Lippa, S. M. (2018). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist, 32*(3), 391-421. https://doi.org/10.1080/13854046.2017.1406146

Lupu, T., Elbaum, T., Wagner, M., & Braw, Y. (2018). Enhanced detection of feigned cognitive impairment using per item response time measurements in the Word Memory Test. *Applied Neuropsychology: Adult, 25*(6), 532-542. https://doi.org/10.1080/23279095.2017.1341410

Manderino, L. M., & Gunstad, J. (2018). Performance of the Immediate Post-Concussion Assessment and Cognitive Testing Protocol Validity Indices. *Archives of Clinical Neuropsychology, 33*, 596-605. https://doi.org/10.1093/arclin/acx102

Martins, M., & Martins, I. P. (2010). Memory Malingering: Evaluating WMT Criteria. *Applied Neuropsychology, 17*(3), 177-182. https://doi.org/10.1080/09084281003715709

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2019). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist, 34*(1), 88-119. https://doi.org/10.1080.13854046.2019.1637027

McCaul, C., Boone, K. B., Ermshar, A., Cottingham, M., Victor, T. L., Ziegler, E., Zeller, M. A., & Wright, M. (2018). Cross-validation of the Dot Counting Test in a large sample of credible and non-credible patients referred from neuropsychological testing. *The Clinical Neuropsychologist, 32*(6), 1054-1067. https://doi.org/10.1080/13854046.2018.1425481

McMillan, T. M., Anderson, S., Baker, G., Berger, M., Powell, G. E., & Knight, R. (2009). *Assessment of effort in clinical testing of cognitive functioning for adults.* British Psychological Society

Merten, T., Henry, M., & Hilsabeck, R. (2004). Symptomvalidierungstests in der neuropsychologischen Diagnostik:eine Analogstudie. *Zeitschrift für Neuropsychologie, 15*(2), 81-90. https://doi.org/10.1024/1016-264X.15.2.81

Millis, S. R. (2009). *What clinicians really need to know about symptom exaggeration, insufficient effort, and malingering: Statistical and measurement matters.* In J. E. Morgan & J. J. Sweet (Eds.), *American Academy of Clinical Neuropsychology/Psychology Press continuing education series. Neuropsychology of malingering casebook* (p. 21–37). Psychology Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal/pmed.1000097

Ortega, A., Labrenz, S., Markowitsch, H. J., & Piefke, M. (2013). Diagnostic Accuracy of a Bayesian Latent Group Analysis for the Detection of Malingering-Related Poor Effort. *The Clinical Neuropsychologist, 27*(6), 1019-1042. https://doi.org/10.1080/13854046.2013.806677

Osmon, D. C., Plambeck, E., Klein, L., & Mano, Q. (2006). The Word Reading Test of Effort in Adult Learning Disability: A Simulation Study. *The Clinical Neuropsychologist, 20*(2), 315-324. https://doi.org/10.1080/13854040590947434

Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Scholten, R. J. P. M., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology, 58*(10), 982-990. https://doi.org/10.1016/j.jclinepi.2005.02.022

Rienstra, A., Spaan, P. E. J., & Schmand, B. (2009). Reference Data for the Word Memory Test. *Archives of Clinical Neuropsychology, 24*, 255-262. https://doi.org/10.1093/arclin/acp035

Rogers, R. (Ed.). (1997) *Clinical assessment of malingering and deception* (2nd ed). Guildford Press.

Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception* (3rd ed). Guildford Press.

Rosenfeld, B., & Penrod, S. D. (2011). *Research methods in forensic psychology.* John Wiley & Sons.

Rutter, C. M., & Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in medicine*, *20*(19), 2865-2884. https://doi.org/10.1002/sim.942

Shandera, A. L., Berry, D. T. R., Clark, J. A., Schipper, L. J., Graue, L. O., & Harp, J. P. (2010). Detection of Malingered Mental Retardation. *Psychological Assessment, 22*(1), 50-56. https://doi.org/10.1037.a0016585

Sherman, E. M., Slick, D. J., & Iverson, G. L. (2020). Multidimensional Malingering Criteria for Neuropsychological Assessment: A 20-Year Update of the Malingered Neuropsychological Dysfunction Criteria. *Archives of Clinical Neuropsychology, 0*, 1 – 30. https://doi.org/10.1093/arclin/acaa019

Šimundić, A. (2009). Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC, 19*(4), 203-211.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545-561. https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Sollman, M. J., & Berry, D. T. R. (2011). Detection of Inadequate Effort on Neuropsychological Testing: A Meta-Analytic Update and Extension. *Archives of Clinical Neuropsychology, 26*, 774-789. https://doi.org/10.1093/arclin/acr066

Sweet, J. J., Condit, D. C., & Nelson, N. W. (2008). Feigned Amnesia and Memory Loss. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 218-237). Guilford Press.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd They Do It? Malingering Strategies on Symptom Validity Tests. *The Clinical Neuropsychologist, 16*(4), 495-505. https://doi.org/10.1076/clin.16.4.495.13909

Tombaugh, T. N. (1997). The Test of Memory Malingering: Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260-268. https://doi.org/10.1037/1040-3590.9.3.260

Trikalinos, T. A., Balion, C. M., Coleman, C. I., Griffith, L., Santaguida, P. L., Vandermeer, B., & Fu, R. (2012). Meta-analysis of test performance when there is a "gold

standard". *Journal of general internal medicine*, *27*(1), 56-66. https://doi.org/10.1007/s11606-012-2029-1

Tydecks, S., Merten, T., & Gubbay, J. (2006). The Word Memory Test and the One-in-Five-Test in an Analogue Study with Russian Speaking Participants. *International Journal of Forensic Psychology, 1*(3), 29-37.

Wechsler, D. (2009). *Wechsler Memory Scale (WMS-IV) technical and interpretative manual* (Fourth). Pearson.

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., Bossuyt, P. M. M., & QUADAS-2 Group. (2011). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine, 155*(8), 529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Zwinderman, A. H., & Bossuyt, P. M. (2007). We should not pool diagnostic likelihood ratios in systematic reviews. *Statistics in Medicine, 27*(5), 687-697. https://doi.org/10.1002/sim.2992

Addendum

The author would like to express their thanks to Dr Paul Green for his insights and advice in the conceptualisation of this analysis, and lifelong contribution to the field of neuropsychology.

Chapter II

Empirical Paper

Exploring clinician experiences of Shared Decision-Making in Multiple Sclerosis: An Interpretative Phenomenological Analysis

Abstract

Background

Multiple Sclerosis (MS) presents both patients and their clinicians with significant challenges due to the unpredictable but cumulative impact upon quality of life. Shared decision-making is widely advocated in healthcare settings and is suggested as preferable in consultations involving preference-sensitive decisions – such as the initiation of disease-modifying therapies (DMT's). To date, little research has explored the perspectives of clinicians working with MS about this process. This study therefore aimed to develop an understanding of the lived experiences of clinicians as they engaged in shared treatment decisions with their patients.

Method

Four consultants were interviewed regarding their experiences of shared decision-making with patients with MS. Interviews were analysed using Interpretative Phenomenological Analysis.

Results

Three superordinate themes were identified through analysis; 'Navigating the therapeutic relationship', 'Knowing what I don't know' and 'Values, self and the system'. Through these themes participants shared the tension which they experience during consultations with their patients. Whilst identifying significant value and pursuit of shared decision-making, the influence of both internal (emotional) and external processes on their experience were described. Acknowledging the influence of ongoing tension upon clinician wellbeing may increase clinician capacity to engage ongoingly with shared decision-making.

Conclusions

Whilst tentative, implications for the provision of increased psychological resource to support clinician wellbeing are suggested. This, and future synthesis with existing patient qualitative research may provide further insight into this dyadic process.

Introduction

Multiple Sclerosis

Multiple Sclerosis (MS) is a chronic autoimmune disorder resulting from inflammation of the central nervous system, typically the brain and spinal cord (Compston & Coles, 2008; Inojosa et al., 2020). According to prevalence data, an estimated 2.2 million individuals globally are living with MS, with over 100,000 of these in the UK (Wallin & GBD 2016 Multiple Sclerosis Collaborators, 2019). Neurologically, MS symptoms can present in variable ways, including changes in cognitive function, fatigue, limb weakness, bladder dysfunction and gait ataxia amongst others (Wallin & GBD 2016 Multiple Sclerosis Collaborators, 2019). The accumulation of disease in MS and its treatments are associated with significant levels of physical and mental disability and have a substantial impact on quality of life (Jongen, 2017). MS has been described as one of the most common causes of non-traumatic disability in young adults, with typical age of diagnosis peaking between 20-40 years old (Kamm et al., 2014). It is widely acknowledged that the unpredictable nature of MS and the potential for long-term effects following a relapse can have a significant impact upon the mental state of patients, with depression and anxiety resulting from high levels of uncertainty (Kroencke et al., 2001). This enduring uncertainty produces significant challenges for both patients and clinicians.

Approximately 85-90% of individuals who are diagnosed with MS are diagnosed with Relapse-Remitting MS (RRMS), whereby they experience acute episodes of worsening symptoms which may last for weeks or months, before making at least a partial recovery over a similar period (Compston & Coles, 2008; Jongen, 2017). Individuals with Primary Progressive MS (PPMS) will experience an ongoing worsening of their symptoms over time, albeit with periods of apparent stability. For many patients with RRMS, the level of recovery they experience from relapses can decline, at which time they may be

diagnosed with Secondary Progressive MS (SPMS). As with symptomatology and individual experiences, the therapeutic management of MS can vary significantly. Management of progressive MS types (PPMS and SPMS) typically focuses upon using steroid treatment to provide symptomatic relief to improve quality of life and the ability to maintain work (Kamm et al., 2014). Treatments for RRMS involve the use of proactive treatments to reduce the frequency and intensity of relapses, known as Disease Modifying Therapies (DMTs). Although variable in their precise operation, these therapies seek to reduce the scarring (lesions) and damage to the cells which make up the myelin sheath (coating on nerve fibres which transmit messages), associated with relapses and thus slow the accumulation of ongoing disability (Vargas & Tyor, 2017). DMTs are widely categorised as both first-line therapy and second-line escalation therapies. Typically, one may commence with a first-line medication and switch to an alternative therapy if this initial therapy is ineffective or only partially effective (Fenu et al, 2015). As of 2019, more than 30 DMTs were in development (Ha Vinh et al., 2019).

Whilst DMTs have been shown to reduce the frequency and severity of relapses in MS and have a positive effect on life expectancy (Kingwell et al., 2019), the treatment of MS presents a complex decision for patients. A decision between partially effective medications, each carrying differing side effects, and which may have a significant impact upon quality of life. These adverse side effects range from nausea, respiratory infections and abdominal pain, to more potentially life-threatening side effects such as progressive multifocal leukoencephalopathy (PML[7]) and leukaemia (Torkildsen et al., 2015). DMTs

[7] PML is a progressive viral disease of the brain can cause substantial damage to the brain in many areas leading to wide-ranging symptoms. Potential consequences of PML can include those similar to MS (e.g., weakness, visual changes, loss of language skills, personality changes and memory problems), progressing to further complications including dementia, seizures or coma. Mortality from PML ranges from between 30-50% of cases diagnosed in the first few months. (Pietrangelo, 2020)

have, however, been shown to slow the progression of MS and reduce relapse rates when taken as prescribed, thus delaying potential disability (Li et al., 2020; Freedman, 2013)., Developing their understanding of the disease and the options available, and adjusting to the unpredictable nature of their MS can add a significant psychological burden for patients (van Capelle, 2017). Despite the potential for a positive outcome, the numbers of patients who agree to commence DMTs can vary significantly from 50% to 75% (e.g., Greb, 2012; Remington et al., 2013). Those who have not initiated DMT treatment cite disease phase and severity, neurologist advice and uncertainty about the effects of DMTs, including potential side effects, as contributing to their decisions (Visser & van der Zande, 2011).

Shared Decision-Making

Shared decision-making (SDM) is described as an “interpersonal and interdependent process” (Legare & Witternan, 2013, p 277) for making decisions about care and treatment. Within this process the relationship between patients, professionals and their caregivers, and their influence upon each other is acknowledged. Its collaborative stance demonstrates an appreciation of the respective expertise of both the clinician and the patient, as displayed in Table 12 (Kings Fund, 2011 as adapted in National Institute for Health and Care Excellence, 2019).

Table 12

Complementary expertise of patient and healthcare professionals in Shared Decision Making (adapted from Kings Fund 2011 as cited by NICE 2019)

Expertise of healthcare professionals	Expertise held by patients
- Diagnosis	- Experience of illness
- Aetiology of disease	- Social circumstances
- Prognosis	- Attitudes to risk
- Treatment options	- Values
- Probability of outcomes	- Preferences

Whilst the involvement of patients in SDM is strongly advocated within healthcare systems (e.g., NICE, 2019; Shafir, 2012), previous systematic reviews of SDM have identified as many as 161 definitions of SDM, with limited consistency relating to what this means in practice (Makoul & Clayman, 2006). The model proposed by Makoul and Clayman (2006) sought to integrate these definitions and concepts into key components, a summary of which are presented in Table 13.

Elwyn (2017) proposed the 'Three-Talk-Model of SDM. This cites three conversational steps necessary within patient-doctor interactions: (i) Team talk (describing choices, offering support, defining goals); (ii) Option talk (discussing and comparing possible options); (iii) Decision talk (seeking informed preferences and making decisions based upon these). (See NICE, 2019).

SDM has been cited as especially valuable in preference-sensitive decisions, where more than one treatment option is available and the choice is based upon the individual evaluating their options (Wennberg et al., 2010). It is suggested that arriving at

a decision using a collaborative approach may have a positive effect on treatment continuation, which relies upon patient motivation e.g., making lifestyle changes, attending regular appointments, taking medication (e.g., Joosten et al., 2008; Samalin, 2018).

Table 13

Behaviours identified as essential in the Shared Decision-Making process (Makoul & Clayman, 2006)

	Element of shared decision making
1	**Definition and explanation of the healthcare problem**
2	**Presentation of available options**
3	**Exploration of the relative pros and cons of options (including benefits, risks and costs)**
4	**Clarification of patient values and preferences**
5	**Discussion of patient ability and self-efficacy in relation to health/health behaviours**
6	**Clinician presentation and sharing of knowledge and recommendations**
7	**Clarification of patient understanding**
8	**The making or deferral of a decision**
9	**Arrangement of follow up**

Shared Decision-Making in MS

Due to the chronic and intermittent fluctuations and the different medications available, MS is one such condition where an SDM approach is advocated. Even amongst patients who initiate DMTs treatment adherence can be as low as 40% (Treadaway et al.,

2009). Factors influencing treatment initiation and adherence include challenges in obtaining and remembering to take medications, perceived self-efficacy to administer medications (especially injectables), as well as patient perceptions of the respective risks and efficacy of medication and perceived ability to cope with adverse events (Costello et al., 2008; Remington et al., 2013). The relationship and communication between a patient and their clinician have also been cited as an influential factor upon engagement with DMT treatment and highlights the importance of SDM (Remington et al., 2013; Ben-Zacharia et al., 2018).

The influence of individual patients' cognitive styles, personality traits and cognitive competency (within the context of disease-related processes) have all been reported to be under-researched (Heesen et al., 2013). Similar individual differences have been found amongst clinicians. In a systematic review of cognitive biases and personality traits associated with errors in diagnosis and/or the management of conditions, factors including overconfidence, information and availability bias, tolerance to risk and ambiguity emerged as associative (Saposnik, et al., 2016).

It has been demonstrated that patients and clinicians report differences in their assessment/tolerance of risk of side effects (Heesen et al., 2010; 2017). Within the early stages of their disease, MS patients may be willing to accept severe side effects for a significant improvement in their symptoms (Wilson et al., 2014). However, Bruce and colleagues (2016; 2018) found that patients who had progressive MS, or who had previously taken a DMT, were more likely to report a willingness to take a hypothetical drug with a high probability of mild, moderate and severe side effects despite limited treatment efficacy. For patients recently diagnosed with MS, it is suggested that the process of making decisions was a product of confronting the disease and the potential for decline alongside their hopes for delaying disease progression (van Capelle et al., 2017).

Clinician Experiences

There has been little research published to date which explores the perspectives of clinicians working with MS patients on the SDM approach. Therefore, it is necessary to draw upon the literature which exists within other chronic health conditions. Qualitative studies have reported clinician perspectives on sharing treatment decisions in heart disease (Borg Xuereb et al., 2015) and kidney disease (Noble et al., 2017). Within these studies, clinicians reflected upon the significance of the decisions they and their patients are faced with, as well as the influence of systemic factors upon the decision-making process and outcomes. Brogan, Hasson and McIlfatrick (2017) explored the use of SDM in focus groups undertaken with community palliative care professionals. They found that whilst the benefits of SDM were universally acknowledged, professionals identified several barriers to its implementation in practice. Firstly, conceptual confusion regarding SDM, including what it was and who was responsible for initiating it was recognised. Due to the varied interpretation of what constitutes SDM amongst both clinicians and patients, understanding the process and its effectiveness is challenging (Ben-Zacharia et al., 2018). In addition, the uncertainty experienced in the process of SDM through unpredictable illness trajectory and patient/family knowledge of end of life care was a further source of challenge. Finally, organisational challenges, including disconnected service delivery models and high workload pressures (Brogan et al., 2017) were cited. Within an increasingly pressured health system, it is suggested that an accumulation of these factors in treatment decision making may be associated with effects upon clinician wellbeing and increased prevalence of burnout (Imo, 2017).

Rationale for Current Study

Given the various treatment options available to people with MS, and the complexities of the possible outcomes, the current study will focus on exploring SDM in

MS, focussing specifically on the experience of clinicians. The aim of the current research is to focus on how clinicians involved in treatment discussions experience the process of treatment-focussed interactions with patients. This provides an opportunity to gain a better understanding of their experiences and how they make sense of this in the context of their personal, professional and social world. To date there has been limited research which has explored how clinicians make sense of the experiences of sharing decisions with their patients, specifically in the context of treatment of MS. This study therefore sought to gain insight into these experiences with a view to aiding and informing future research and synthesis of psychological factors which may influence or be a product of this decision making.

Method

Design

This study adopted a qualitative design using an Interpretative Phenomenological Analysis (IPA) methodology. Semi-structured interviews were undertaken with participants to obtain rich data about their experience of SDM within the context of MS.

Rationale for use of Interpretative Phenomenological Analysis

IPA has been widely cited as a valuable adjunct to traditionally used quantitative research approaches within the area of health psychology (e.g., Brocki & Wearden, 2006). IPA is an approach which is concerned with the detailed exploration of the human lived experience (Smith et al., 2009). This in-depth methodology attends closely to the detailed and nuanced phenomenology as articulated by individuals. Through IPA, a researcher seeks to make sense of the meaning which a participant attributes to their experience of a given phenomenon - this is known as the double hermeneutic stance. IPA was chosen as the methodology for this research due to the intention to explore the experiences of clinicians as they engage in the socially constructed process of decision-making.

Procedure

Ethical approval

This study received sponsorship and approval from the University of Birmingham (Appendix 6). The research was approved by the Health Research Authority (HRA) (Appendix 7) and by the Research and Development Team of the recruiting NHS Trust (Appendix 8). Prior to taking part in the study, participants were provided with a Participant Information Sheet (Appendix 9). Informed consent was sought (Appendix 10) and participants were informed of their right to withdraw.

Recruitment

Participants were recruited through a purposive sampling method. Potential participants were identified by the Clinical Supervisor from within an MS team working in a large NHS teaching hospital. These participants were provided with brief details of the research and were asked to provide their consent to be contacted by the author. On receipt of the Information Sheet and after an agreed period (at least 48 hours) during which participants were offered the opportunity to ask any questions about the research, the author contacted participants and, a time for interview was arranged.

Interview procedure

A semi-structured interview schedule (Appendix 11) was developed in line with the guidelines outlined by Smith et al. (2009). This schedule was reviewed by a Consultant Neurologist and two Clinical Psychologists in their capacity as research supervisors during the development of the proposal. A semi-structured interview is the favoured mode of data collection in IPA, as this enables the researcher to maintain a flexible and responsive rapport with participants, whilst providing the space and time for a participant to share their experiences and facilitate exploration of their lifeworld (Smith et al., 2009). Participants were interviewed at a mutually convenient time in a private room located within the hospital. Each interview lasted between 36 and 48 minutes (Mean=41 minutes), guided by the same interview schedule.

Ethical considerations and information storage

Participants were informed and reminded of their right to withdraw without need for justification from the study prior to, during and after the interview had taken place. Due to difficulties in withdrawing data from IPA analysis once it has been analysed, participants

were contacted after a two-week reflection period post-interview to confirm their consent for their data to be included.

On entry into the study participants were anonymised through the allocation of an alphanumeric ID and, at transcription stage, a pseudonym. Interviews were transcribed verbatim by the author and, following transcription, interview recordings were destroyed. All participants were informed that anonymised quotes from their transcripts may be used as part of the thesis and any future publication in a scientific journal.

Participant pseudonyms/identifiers and associated demographics (consisting only of names, role and contact email address) were stored on a password protected and encrypted computer database. Any hard copies containing personal identifiable information (e.g., consent forms) were stored in a locked filing cabinet on the hospital site accessible only to the researchers and their supervisor. Following study completion research data and personal identifiable information will be transferred to the University of Birmingham and stored for 10 years in line with University of Birmingham policies.

Participants

Four participants were recruited and interviewed between December 2019 and March 2020. Due to the onset of the global Coronavirus (COVID-19) pandemic in early 2020, recruitment was halted by the NHS Trust involved in the study before the planned target of six to ten participants had been recruited. For details of participants, see Table 14 below. For the purposes of ensuring confidentiality only brief details are presented.

Table 14

Participant Demographic Information

Pseudonym	Gender	Role	Experience with MS (Years)
Charles	Male	Consultant	13
Asif	Male	Consultant	12
Michael	Male	Consultant	27
Asha	Female	Consultant	11

N.B. All Consultants worked within the Neurology speciality

Analysis of data using IPA

IPA analysis was undertaken according to the common processes cited by Smith et al. (2009). Following verbatim transcription of the audio-recorded interviews, transcripts were read and re-read twice, before reading through a third time whilst listening to the original audio-recording. This was conducted to support the author to engage with the interview data of individual subjects in an immersive and attentive way. The five steps of IPA analysis as described in Table 15 were then undertaken on a participant by participant basis.

Table 15

Outline and description of the steps of IPA as outlined in Smith et al. (2009)

Step Number	Step Name	Process
1	Reading and re-reading	Transcripts were read and re-read, alongside the audio-recording, in order to engage actively with the data.
2	Initial noting	Initial thoughts, comments and reflections were noted on the original transcript. These notes included the author's comments on the content (what the participant said), language (how this was said, the language used) and concepts (the beginnings of more interpretative comment about the participant's understanding)
3	Development of emergent themes	Through review of the initial notes and the transcript, these were combined into emergent themes
4	Identifying connections between themes	Emergent themes were explored for connections and collated to develop a thematic structure for each participant including subordinate and super-ordinate themes.
5	Identifying connection and patterns across subjects	Thematic structures for each participant were recreated and looked at simultaneously to explore connections between identified themes.

Following detailed line-by-line coding of the initial transcripts during which the author commented upon specific linguistic, descriptive and conceptual interrogation of the data (Appendix 12), initial prototype/emergent themes were developed. These emergent themes were then organised by the author into clusters based upon connections and similarities between themes (Appendix 13). The identification of corresponding transcript excerpts were

then grouped together under these emergent themes, enabling the author to check the internal validity of the theme against the research data (Appendix 13). Emergent themes and the corresponding excerpts were then collated into super-ordinate themes and their corresponding sub-themes to create an initial thematic structure for each participant. After analysis of individual participants, a brief narrative summary was written in which the author captured key points and reflections from the analysis and the thematic structure for each participant.

Individual thematic structures were then written out onto post-it notes for the researcher to be able to collate shared themes/connections across participants more easily (Appendix 14). These final groupings included the integration of supporting transcript excerpts to develop the overall final thematic structure, including super-ordinate and sub-ordinate themes.

Credibility and validity

To maintain a check on the quality of the research process, the author was guided by the guidelines developed by Yardley (2000) in which characteristics of good qualitative research are set out. These guidelines and the steps taken to address them are summarised in Table 16.

Table 16

Principles of validity of qualitative research as cited by Yardley (2000).

Quality/validity principle	How this was addressed in the present research.
Sensitivity to context	- Familiarisation with existing literature on SDM and consultation with clinical and research supervisors (including Consultant Neurologist) during project development. - Use of semi-structured interviews supporting space for participants to share their experiences - In-depth analysis including integration of verbatim quotations in supporting developed themes ensured participant voice remained central.
Commitment and rigour	- Consistent engagement across the design and seeking of ethical approval through to the analysis and write-up of findings. - Regular supervision, engagement in teaching and additional reading facilitated methodological rigour - Triangulation of interpretations was undertaken at all stages of analysis using peer supervision with fellow Trainee Clinical Psychologists and research supervision with academic and clinical supervisors. - Reflective journal and narrative summaries kept throughout recruitment and analysis to facilitate authors' awareness of their role in the double hermeneutic.
Transparency and coherence	- Participant information was provided in writing using a clear information sheet which had received appropriate ethical approval and validity checks. - The analysis process was undertaken using steps outlined previously, with regular checking from research supervisors. - Clear audit trail of how themes were developed (including dated document versions and photographs) supports research transparency and coherence.
Impact and importance	- Discussion of potential clinical applications and implications of this analysis will be included within this paper to support the reader to consider the meaningfulness of findings and influence for future practice.

Researcher context – Reflexivity

Shaw (2010) cites reflexivity as a crucial component of the qualitative research process within psychology, in supporting the context of the research process. Holding awareness of the context of oneself, the way in which we interpret the experience of the research process, and our relationship with research and participants means that one can never solely enter analysis as a 'blank slate' (Shaw, 2010). The reflexive account was written in the form of a first-person narrative beginning from the time of the first interviews and throughout the analysis process.

> *"I am a 30 year old, White-British female with a longstanding professional interest in the area of long-term health conditions and the influence on wellbeing and quality of life. Over the course of clinical psychology training the topic of treatment choice/decision-making has been one which I have been increasingly part of due to experiences of close family members. I have been increasingly confronted by the unpredictability of health and its impact on hopes and expectations. Throughout the coding of my data it has been important for me to maintain my reflective log and to consider how my personal familial context including recent significant loss and family illness may have impacted on the position from which I was interpreting my data. However, reflecting upon this and having access to regular supervision has been integral in supporting me to be flexible and open to alternative interpretations which may be drawn from the data. It is of interest to me that my findings and the implications of these, may reflect some of the value which I have appreciated.*
>
> *Reflecting upon my approach to data analysis and guiding my approach to the interviews was my epistemological position. As a psychologist I am typically drawn to a critical realist position in that whilst there may be a more objective or measurable reality, there are nuances in the way in which this is perceived and described by*

individuals. One way in which I interpret this is that phenomena are constructed and subjectively experienced within a social context. I therefore fall somewhere between social constructionist and critical realist positions."

Results

A total of three superordinate themes were identified from the analysis: 'Navigating the therapeutic relationship'; 'Knowing what I don't know'; and 'Values, self and the system'. As displayed in Table 17, these superordinate themes had a total of eight subordinate-themes. These themes are described narratively below, along with example participant quotes to facilitate their illustration.

Table 17

Overview of superordinate and subthemes identified through analysis

Superordinate Theme	Subordinate Theme	Participants Contributing
Navigating the therapeutic relationship	1.1 Collaborating with, not doing to	All participants except Charles
	1.2 Individual in context	All participants
	1.3 To let go or step in?	All participants
Knowing what I don't know	2.1 Knowledge is not straightforward	All participants
	2.2. The unknown future	All participants
Values, self and the System	3.1 Holding values and acknowledging "emotional accidents"	All participants
	3.2 Protection and containment	All participants
	3.3 Restricting my agency	All participants

Superordinate theme 1: Navigating the therapeutic relationship

All participants reflected upon the considerations and nuances of the relationship between clinician and patient in the decision-making process. Analysis of this theme reflected firstly upon the importance that clinicians placed upon maintaining a therapeutic alliance and how they position themselves in working with their patients. Secondly, participants expressed their acknowledgement of the individual circumstances and context which patients bring to a decision-making consultation and the influence that this may have on the relationship. Thirdly, clinicians identified a need to make a judgement call in the decision to permit patients to opt against a preferred treatment or to step in and assert their recommendation.

Theme 1.1 Collaborating with, not doing to

All participants except Charles reflected upon the value and importance of maintaining a therapeutic alliance with their patients. For Asha this involved the explicit expression of intent to situate herself alongside her patients: *"Ye-es I think it's really important I think it's really important for them to feel that as a doctor I'm on their side…"* ***(Asha)***. Asha describes approaching consultations from the imagined position of her patients and using this to guide their discussion:

> *"Mmm no I-I-I think I mean in neurology in general I always sort of put myself in the patient's shoes as much as possible and I think if I was sitting in that chair on that chair what would I want from my doctor how what kind of conversation would I ha- would I want to have…"* ***(Asha)***

Throughout their accounts, participants emphasised the importance of being perceived as working with their patients rather than holding the position of a directive and paternalistic figure. Here Asif illustrates this: *"…there's no point pullin-ting the entire doctor*

patient relationship just to force people into doing something they don't want cos they're not gonna comply anyway" ***(Asif).*** This was one of many examples where participants cited a preference for prioritising the wellbeing of the patient over a clinical position and acknowledging the role of motivation in behaviour. Michael provides further illustration:

> *"…I've always taken that view and as a-because it's far more important to meet th- to maintain the quality of the relationship with the patient [mm] rather than stand on some didactic point er in a sort of scientific integrity which y'know would perhaps I-well it's just a value judgement isn't it really"* ***(Michael)***

Michael identified the potential for significant damage within the therapeutic relationship if patients perceive clinicians as obstructing their access to treatment opportunities:

> *"…um and then when they didn't fulfil criteria would um become aggrieved and it would sour the whole professional relationship so the-the-the appearance that one was as a clinician as a physician um obstructing patient care [mm] and therefore behaving not in the best interests of the patient…"* ***(Michael)***

Interestingly, all three clinicians who contributed to this sub-theme described citing the influence of external regulation (such as funding criteria, processes, and guidelines) in their consultations with patients. In Asif's case this was a very explicit way of declining requests from his patients without being viewed as obstructive:***"…****urm lots of people might know someone that's been on this drug or that drug and it's it's fairly easy enough to say but y'know you're not eligible for that for this reason and people understand that.."* ***(Asif)***

> *"…'cause a lot of patients will actually come and say can I not have this drug so so the first thing I have to apologise [nervous tone] and then I have to sa- I have to be frank and say based on the evidence based on the literature based on the bodies of*

guidance we have that w- that we work with this is the choices we have available this may change in the future urm but this is what ur this is what I can offer you so [sigh]…" ***(Asha)***

Asha describes a sense of regret or guilt associated with having to decline a request from her patients whilst simultaneously appearing to try and offer hope. The sigh at the end of this quotation further contributed to a sense of despair and fatigue. Whilst the avoidance of paternalism resonated strongly throughout this theme, there was also a recognition that there were times where the protection of the patient may be more important than maintaining the relationship. As described by Michael:

"…I think I'm very comfortable with saying no it's too dangerous for you [laugh] [yeah] because I'm sorry that's paternalistic [yeah] but that actually is my job to make sure that the patient has understood that that's the way life is…" ***(Michael).***

Alongside narratives of the necessity and drive to protect the therapeutic relationship there was a recognition that any alliance is dependent upon the clinician's ability to hold in mind the individual circumstances and context which patients bring to the decision-making consultation. This is described in the next subtheme.

Theme 1.2 Individual in context

All participants expressed a recognition of the individual differences between patients and their influence within the therapeutic consultation. Acceptance and adjustment to the diagnosis of MS was one such factor. Specifically, the acceptance or otherwise of diagnosis was viewed by participants as having the potential to impact upon even the earliest stage of SDM. As Asif describes:

"…at that point I think she hit a brick wall cos she wasn't prepared for this [mm mm] I think she'd put all the eggs in the 'it's never gonna be MS' basket [mm] it just happens to other people doesn't it…" ***(Asif)***

Here, an individual's perceptions and appraisals of the likelihood of MS to affect them can lead to feelings of shock and become a barrier in the immediate discussion of SDM. Asif explains how this apparent denial continues throughout this patient's interactions with him. Acceptance and adjustment was identified as something that the patient did not undertake alone, and the responses of those around them in close relationships provided a further sense of uncertainty for participants:

"…on the one hand you've got um partners or significant others who are totally bought into the urm what I would call a restoration um agenda so everything's gotta go back to normal…you've got significant others who have become very resentful and very urm negative about the experience and are processing really in terms of urm how destructive it is and urm that can of course become quite difficult relationships break down [mm] isolation occurs…" ***(Michael)***

Fundamental to this process, many of the participants described a sense of awareness of the impact of an MS diagnosis and the implications it may have upon a patient's imagined future hopes and expectations. Asif highlights the fear that this can bring and this is supported by Asha: *"I dunno people probably don't hear that..[mm] and they will know someone in a wheelchair..[right]..so it's always a a mental image they come pre-preconceived ideas about imminent disability"* ***(Asif)***

"…the first thing that they think about is do I need a treatment or because most patients don- that I've met anyway don't want to be reminded of their MS they want to live a normal life so for most patients when you talk about treatment the first thing

they I imagine they would think about is what impact is this gonna have on my life…am I gonna be able to work travel provide for my family have children live my full life with a normal life expectancy…" ***(Asha)***

All participants who contributed to this sub-theme recognised the significance of the decision which patients were making and the potential implications for their lives. Charles was able to illustrate the breadth and spectrum of considerations which a patient may hold in the decision-making process:

"…some people are become fixed upon efficacy data, some people are very concerned about the side-effect profile, some er people some patients will urm look at the treatment burden you know is this something I need to take every day or every other day or once a month you know what are the requirements in terms of monitoring…" ***(Charles)***

Whilst a complex myriad of factors may feature in the equation for an individual patient, all participants were able to recognise the importance of attempting to understand the priorities of the patient in their context at that time. This awareness of the context was cited as influencing the way in which the clinician communicates with their patients. As summarised by Asif:

"…different people are different some people are very risk-averse some people just want an effective treatment and I think that it's precisely to avoid difficulties I try and factor that in [mm] I-I will rather they get on a treatment they're happy with sooner than try to force them to go on something that they're not happy with…" ***(Asif)***

To maintain the therapeutic relationship, the clinician is required to make a series of moment-by-moment assessments to ensure that they are meeting the patient in their context. This includes making decisions about what, where and when to share information, advice

and guidance in a digestible way to the client. Sometimes this may mean that the clinician must decide to 'not' intervene as is described in the next sub-theme.

Theme 1.3 To let go or to step in

Through their interviews, all participants expressed a tension in balancing the autonomy of their patients with their drive to protect and offer the 'best possible' care. These experiences encapsulated a recognition of the clinicians' perceived role and the way in which they manage when a patient makes a decision that is not the preferred option. For Michael, the MS patient group was one which was very proactive and invested in seeking or understanding the treatment opportunities which were available to them: *"…so we have to accept that patients will come with all of their [hesitation] um-research in their heads or usually in their big folders [m-hm] and tell us urm what they want [mm]…"* ***(Michael)***. Whilst expressing a belief that this was necessary for the clinician to accept and work with, Michael also highlighted that this may leave patients vulnerable and that he held a responsibility to intervene. At times this may override the aforementioned therapeutic relationship, however Michael described that he moderated such interactions through explanation:

> *"…I've always emphasised to patients …that we would support them whatever they wanted to do [mm] it wasn't my job to police but simply-y'know as long as it wasn't overtly dangerous or I hadn't got reasonable grounds to exp- to think that they were being exploited…"* ***(Michael)***

For Asha, the experience of being prevented from being able to help the patient in front of her was one which held a strong position in her memory. The intense emotion which she describes as coming from the patient was also reflected in the sadness and frustration that she expresses:

"…that case is particularly challenging to me because a) it was of the anger and the frustration that she had b) the fact that I knew that her disease progression was gonna continue and I wanted to help and c) because she didn't want to take that help…" ***(Asha)***

Asha does however reflect upon how these challenges are magnified by apparent 'misjudgement' of the patient context (patient previously "*fell through the gaps").* For example, Asha recognised that her attempts to encourage treatment initiation and alleviate suffering was perceived as imposing on them or trying to save them by the patient. This sense of needing to bide your time and pick the moment to intervene was shared by Asif:

"…once you get beyond that it's this...this concern they have that err all these treatments might have side effects and oo do I really need it anyway [mhm mm] aand then it's trying to rationalise with them what their relapse rate risk scans have been to explain why it will be beneficial and then you know if you've told them everything and they still don't want to you then have to just look for the next opportunity with them be open-minded… it might it might meant that they don't want to listen to you until they've had another relapse…" ***(Asif)***

Holding this position of 'sitting with' was not a comfortable one however, and Asif highlighted that this often carried a lot of uncertainty and a sense of needing to compensate:

"…then I will be most concerned about those ones that I think that are going for a treatment that's not gonna hold them…"…in a way it just means that I have to put more effort and energy into monitoring them clinically and radiologically cause I don't I don't want them to just be yeah relapsing silently in the background.." ***(Asif)***

Conversely, Charles appeared to experience much less challenge in this domain, citing that *"rarely they will do something completely opposite"* to that recommended by

clinicians. Collectively these experiences highlight the challenges of allowing individuals to make their own decisions, and parallels were noticed (explicitly in Michael's case) between the content of these discussions and principles within the Mental Capacity Act about an individual's right to make what may be viewed as an unwise decision in the eyes of the professionals.

Achieving the balance between respecting a patient's right to choose and the ongoing wish to 'care for' appeared crucial in the experiences of participants and impacted upon the ways in which they approach their patients. This was also found to be heavily related to the sharing of information and knowledge and the dilemmas associated with this.

Superordinate theme 2: Knowing what I don't know

Fundamental to the SDM process was the role of information. The relationship with information/knowledge however was not without complexity for clinicians as the first theme captures. The second theme highlights the experience of working with the uncertainty and unpredictability of MS 'the condition'.

Theme 2.1 Knowledge is not straightforward

For all participants, having knowledge and information was fundamental for the making of an informed decision, not only for the clinician but for the patient. Indeed, giving information was a significant part of the decision-making consultation and was emphasised in all the interviews as the foundation to this process and a key role for clinicians with the support of colleagues. Initially however they need to assess what the patient already knows, as well as their willingness and ability to receive information. For Asha this was sometimes starting from a position of informing patients about the changes in approach to MS management *"…in the past there was no treatmen*t…" and opening opportunities. Not providing this information/education was viewed as having significant negative

consequences for clinicians and patients' confidence in treatment and wellbeing. As described by Asif:

> *"so I think then they have a lot of stress and anxiety around this cos they think the sky's falling down and their MS is uncontrolled [mm] whereas actually people that understand t-he treatments and their limitations and what things are relapses and what things aren't I-I think it takes a fair bit fair deal of their stress away so they can actually get on with their lives"* ***(Asif)***

The possession and exchange of knowledge was recognised by all participants as an important part of the decision-making process. However, the relationship with knowledge was not straightforward. Charles describes the stages and need to obtain *"…a neurological explanation of your findings…"* to proceed. Through medical investigation, including scans, clinicians seek an objective and quantifiable picture of their patients' illness. Furthermore, for clinicians possessing a knowledge and understanding of the evidence base and having this available to them, carried a sense of protection and security which they could turn to. This was highlighted by Asif:

> *"…I think there is evidence that if you need to escalate so long as you do it within the first two years you can preserve outcomes so I'm willing to take that if it means that they do go on treatment…" (****Asif)***

Use of trusted sources of information including patient information leaflets and websites were important in these discussions, as were structures available within the service, such as discussion groups facilitated by specialist nursing colleagues. In the absence of clarity and/or being given information, participants recognised the likelihood of patients seeking information for themselves, as described in theme 1.3. Whilst recognising the value and importance of knowledge, Michael identified that information itself can be fallible:

"…the abuse that occurs that- the- th- for which it's quite difficult to prevent but the the social media stuff that is is so inaccurate or led by enthusiasts and urm again I-I wish there was a way we could just make that side of things more responsible…" ***(Michael)***

The increased accessibility of information outside of the control of clinicians was identified as having the potential to lead to negative consequences for their relationships with patients through building false expectations. Whilst experiencing empathy for their patients, this was also identified as posing threat to the therapeutic relationship as discussed in theme 1.1, leading also to some sense of irritation.

Alongside the danger of misinformation, knowledge was also viewed as having the potential to overwhelm their patients– *"there's a lot I think there's a lot to take in…"* ***(Asif)***. Thus, clinicians cited the need to adjust and adapt the amount and way in which they provided information to their patients. The necessity for openness and transparency in relation to the diagnosis and medications was highlighted by all participants. Asha stated:

"…then when I when I g-get an idea about the patient I then y'know say to them 'look y'know these are the choices we have' and I'm very open about it cos I always tell them look y'know you can't have all twenty drugs y'know these are the three that you can chose from…" ***(Asha)***

However, for Michael there was a recognition that information about the extent and possibility of significant side-effects was beyond the scope of communication *"…I wonder if the patients can fully take on board just how awful it is if you get PML [Ok] and I don't know if we tell the truth …"* ***(Michael).*** Some of this was attributed to having a limited time available to express and communicate the nature of these consequences. Questioning whether one "*tells the truth"* about such matters suggests potential for an avoidance of these

conversations due to the challenge in communicating this to patients. Alternatively, it may be that the apparent rarity of these consequences means that they are and can be less present in the consultation. The weighing up of clinician's wish for proactive treatment versus the potential for side-effects is a complex equation and this is may influence the depth of information which is shared.

The possession of knowledge and information was viewed as a double-edged sword by participants. However, having knowledge appears preferable over the alternative, which is the acceptance and recognition of the uncertainty associated with the future.

Theme 2.2 The unknown future

Whilst participants possess a substantial amount of knowledge, obtained through their training and status as Consultants specialising in the field of MS, all participants acknowledged the uncertainty and unpredictability associated with a chronic condition. More than this however, was a sense that the presentation of MS specifically threatened any hopes for certainty. Michael cites the nature of disease as one of the reasons behind this uncertainty: "…*remitting relapsing nature of the disease any treatment um that you might propose in MS can seem to have an effect unless you do a properly conducted trial…"* ***(Michael).*** Here Michael describes how without the use of highly controlled research studies, the fluctuating nature of MS symptoms may lead to incorrect associations between trialled medications and symptom reduction. However, when one favours holding knowledge and objective information, the experience of managing uncertainty is one which can create dissonance for participants. For Asha the experience was contrary to her hopes as a clinician:

> *"…so we're kind of making an assumption and it's very hard for me as a doctor in particular who wants to make a decision based on lots of evidence a lot of MS is still driven very much by best guess [m-hm m-hm m-hm] and I find that very hard"* ***(Asha)***

Asha goes on to describe her hopes that knowledge and evidence can guide expectations of herself and her patients. Being unable to meet the standards which they aspire to within practice and not being able to offer the reassurance that their patients seek about the future may pose significant challenge to clinicians.

For some participants, relaying the unpredictability directly to their patients was a feature of the clinical consultation *"…I say that urm 'look if err if this is your first and only attack..we don't know if you're ever gonna have another one…'"* ***(Asif)*****.** However, for others managing this uncertainty and limited control may have felt like a greater challenge to their competence which they found harder to address directly within their interviews. This sense was demonstrated by Charles:

> *"…I think I'd say maybe ten percent [mmm] where there is uncertainty where there is no b- you know it's not a black and white scenario you know it's grey kind of zone why you don't really know you know what's the best…."* ***(Charles)***

Whilst acknowledging the possibility of uncertainty, Charles appears to move to rationalise this experience through reassuring himself that this is not a common experience but also describing this as the pragmatic nature of working in medicine. It may be that this strategy is a consequence of experience in the profession and a need to separate oneself from doubt and encourage recognition of 'majority' experiences. Charles goes on to suggest that his approach being guided by a more general treatment principle *"to be proactive…"* ***(line 96)*** offers him a sense of security within this uncertainty. Exploring a personal relationship with knowledge and certainty is likely to have roots or implications for the way in which clinicians view themselves and their roles within their system.

Superordinate theme 3: Values, self and the system

All participants contributed to ideas of complexity in the integration of themselves as clinicians with the wider systems and regulations around MS. Analysis firstly reflected upon the relationship between participants' values and the challenges which sometimes confront these. Secondly, participants described the containing and supportive value of working in a team. However, participants acknowledged some of the systemic challenges which they experienced as inhibiting their ability to act for their patients.

Theme 3.1 Holding values and acknowledging "emotional accidents"

All participants articulated a sense of an internalised view of themselves and a motivation to make a difference and reduce the suffering experienced by their patients. When acknowledging the influence of their role as Asha says *"…here as a professional this is my duty of care to offer the best possible treatment that's available…".* She highlights how this position as a professional appears to have become a part of her identity and indicates the necessary requirements which this carries. Reducing suffering and being able to make a difference are values which appear to be strongly linked to the notion of holding hope for their patients and clinicians' ability to perceive a personal sense of agency and being able to *"be proactive you know er rather than… inactive…"* ***(Charles).*** For Asif, his values and hopes for his patients appeared to have had a significant influence on his personal decisions:

> *"…so the whole point is that they should be stable so they can get on with their lives… well I think if you can get through that and explain that y'know the whole point of treatment is that they stay stable so that they can get old happy with everyone else… and I think that's that's why I chose to do MS in the first place it's not as depressing as a lot of other neurological specialities"* ***(Asif)***

Not only does Asif highlight the value he perceives in being able to be 'do' something, but he also alludes to the negative emotions which are experienced when this is inhibited. He describes making a deliberate choice to work within MS because it is suggested that this is a hope that he can make the difference for his patients and play a role in their living a long and fulfilled life. This is echoed by Asha:

> *"…and y'know for some extent as a doctor you want to protect that patient y'know I want to look after them and you can't- you can't [participant laugh] and that is really upsetting for both the parties so [laughter ongoing]"* ***(Asha)***

Asha's laughter in this extract was striking as this appeared incongruent with the emotion she was describing. This appeared to suggest a level of discomfort in articulating and focusing upon the experience of powerlessness/hopelessness and the difficult emotion which this may evoke. Similarly, it is possible that Asif has selected to work in MS in order to avoid experiencing this discomfort too frequently. These emotional experiences appear to evidence the extent to which the identity as a member of the 'helping profession' has been internalised and aligned with participants' values. This leads to emotional distress when they are unable to act congruently with these values.

Inhibition of action was not the only threat to values experienced by participants. For Michael and Charles, confronting the potential for treatment to cause harm appeared to be a precipitant for difficult emotion. This was demonstrated by Michael who said: *"…I have to say that some of the patients we've seen with PML who … have life-changing urm neurological deficits you just go back and think [sigh] was that really worth the risk…"* ***(Michael)***. Throughout this discussion Michael reflected upon the poignancy of witnessing the loss associated with MS progression but also how an awareness of the possible harm associated

with treatment may influence how he would discuss this with patients now. For Charles, the experience of "*regret"* was one which he appeared to find particularly difficult to engage with:

> *"…there is regret you know what if …not so much with the second with the third patient because she's done well but with the first and second patient you see..well actually not so much with the second patient it's just the first patient who developed it so he was in the process of treatment change and then blablabla so"* ***(Charles)***

The tentative expression of the emotion appeared to be followed by rationalisation that the patients concerned had not experienced the level of long-term harm that they may have. The potentially dismissive use of the phrase 'blablabla' may also provide interesting insight into Charles' reluctance to engage any further with what had had happened to his first patient, which may have included a description of the negative consequences. Charles goes on to acknowledge the influence of age and experience in reducing the intensity/frequency of emotion experienced in his role: *"..the longer you do it I guess the less emotional you are going to be with there'll be y'know little accident- emotional accidents …x number of years down the line…"* ***(Charles)***

The description of emotions as accidents appeared particularly powerful and suggestive of how these are perceived as something which needed to be contained or avoided. The apparent unpredictability of emotion appears in contrast to the more valued rational state. This provides a suggestion of striving to keep his own emotions 'in check' within a professional context. Likewise, the change in person from 'I' to 'you' was of interest and appeared to serve a function of distancing himself from this experience. Throughout this subtheme, participants appear to be balancing a tension between outwardly appearing/being a professional caring but rational clinician who wants and can help their patients, and their own internal emotional experience. The tension they experience internally also appeared to

be reflected in their experiences of the system around them as explored in the remaining subthemes.

Theme 3.2 Protection and containment

The relationship between participants, their team and the wider health system features prominently throughout the decision-making process. Part of the perceived value in being part of a team was articulated in relation to the positive impact upon the patient. As Asha described:

> ***"…****and I think the patient also likes to know that there is a body of people that is urm going to be looking after them y'know urm if you come back to a consultan- urm to a patient discussion and you say 'I've had a chat with the MDT the group of other MS expert and nurses' they feel more reassured aswell the patient feels like 'okay so a body of experts have had this discussion that's what they..."* ***(Asha)***

Whilst use of the term 'experts' alludes to a sense of power and authority which comes from the professional knowledge, this passage also suggested a protective experience of having more than one perspective. It is unclear however how much of this experience is mirrored in the experience of the clinician, leading them to feel reassured when they have the agreement of their colleagues. Specialist nursing colleagues were highlighted as being particularly valuable to clinicians due to their ongoing and consistent relationship with the patient. Within the SDM process this was especially important in supporting the thorough communication of information and options to patients which Consultants were unable to do. As described by Asif:

> *"…then there's a lot more that you could do but there isn't time and that I think that's where the MS nurses are pivotal cos when they bring them to a new patient appointment they go through lots of things in detail…"* ***(Asif)***

A sense of gratitude was apparent in discussing the input of colleagues, and for Asif, the above quotation suggested that this alleviated distress when they perceived barriers to being able to offer more to their patients. The gratitude which clinicians highlighted towards their supportive staff was contained to some extent by the backdrop of regulation. Regulation and standardisation because of external governance was a prominent feature in the decision-making process. The fundamental presence and role of funding, through influencing the number and type of options open to patients, was recognised by all. Likewise, the role of external regulation in mandating the need for MDT discussion prior to initiation/escalation of DMT's was a significant change for clinicians. Indeed, for participants, MDT was a forum for agreement and consensus as summarised by Asif:

> *"…I think attending the MDT regularly means that you kind of become familiar with the way the decision works … it means that a lot of our decisions probably tend to converge cos we all know what each others gonna approve of the MDT so there's no point bringing things that are crazy [mm] you already know it's not gon-u it's not gonna fly what's the point …"* ***(Asif)***

Asif also appears to allude to processes of social influence, with participation in the MDT holding a conforming function for its members – i.e., making sure that any decisions and suggestions which are made are acceptable and fair. Alternatively, this may be interpreted as being perceived as restrictive for Asif, especially if ideas may be perceived as "crazy" by those around him. He goes on to express his perception that "..*it's a tight ship…"* but cites this as positive in reducing false hope for his patients. Inevitably within the professional team, whilst identifying value in the roles of others, the relationships between professions and individuals was not always equal. Whilst working collaboratively within an MDT, Charles appears to be highlighting the difference in perceived power between oneself and their colleagues as described in his description of the multidisciplinary (MDT) meeting:

"…and then you essentially recommend treatment and er and ask your colleagues to agree urm not er to disagree…" ***(Charles)***

The benefit of predictability and standardisation which was achieved through external regulation and standardisation was, however, viewed as coming at an emotional cost for clinicians. The 'flip side' of this containment was explored further through the final subtheme.

Theme 3.3 Restricting my agency

Whilst reflecting upon the containment offered through operating within a wider health system and integrated team, participants also identified challenges that this posed for them as they went through the decision-making process. Participants were united in their recognition of the practical barriers which impacted upon their ability to fulfil the role they hoped for. This is demonstrated by Asif who described how the desired proactivity in treatment initiation and delivery is impeded by *"lack of capacity"***.** The size and volume of clinicians' caseloads, and the consequences of this upon the length and breadth of the discussions that can be had with patients was reflected across participants. For Charles, the influence of this on the decision itself was clear:

"..going back to your original question sometimes that decision-making is based on a er the quickest treatment perhaps er rather than the best y'know err I wouldn't say the best more er the most desirable treatment y'know…" ***(Charles)***

This experience of offering a perceived 'less than ideal' was one which could present some internal challenge to clinicians as they seek to intervene in the context of variables outside of their control. Indeed for Charles, this requires him to accept the reality of these limitations and *"just carry on"* because to seek changes was *"just unrealistic"***.** Barriers were also experienced as originating from outside of the immediate team/locality and instead

arising from broader systemic levels. For Asha however this experience was much more acutely expressed in interview:

> *"…one difficulty I have is sometimes I know that the patient would um benefit from a higher efficacy drug but I don't have the leverage to offer that choice which means I'm fully aware that that patient will progress and become more disabled and that's very hard for me to see knowing that there are other things available but there's I can't provide that choice because there won't be the funding available for that patient for that that's quite hard for me…."* ***(Asha)***

Having an awareness of the processes behind the development of funding decisions and appreciation of their necessity did little to support clinician empowerment. Indeed, at times it appeared to devalue and lead to distress for clinicians as their practice was restricted. Observations relating to the perceived validity and applicability of medications in clinical settings further introduced a sense of feeling restrained or impeded for clinicians. For Michael this was explained:

> *"…the criteria are based on the trial [m-hm] but that is to apply wi-with legal rigidity a research process that was never meant to be an accurate portrayal of real-life…"* ***(Michael)***

For Michael, the experience of having to deny treatments for his patients due to their perceived ineligibility was a direct challenge to hopes of offering individual person-centred care. As reflected through this theme the perceived level of personal agency which came with the professional position and training of these clinicians was dependent upon the operational context in which they were acting.

Discussion

Summary of findings

This study sought to explore the lived experience of MS consultants as they engage in the process of SDM with their patients. Through understanding how professionals make sense of this process within their personal and social context, further exploration of the psychological factors which may influence decision-making may be undertaken. Due to the global Coronavirus (COVID-19) pandemic, recruitment was halted with a total of four clinicians. From qualitative analysis using IPA, three superordinate themes were identified: 'Navigating the therapeutic relationship', 'Knowing what I don't know' and 'Values, self and the system'. These themes had a total of eight sub-themes which provided a valuable insight into the ongoing complexities and tensions clinicians encounter as they explore treatment options with their patients. These findings will be positioned in the context of existing literature before the strengths and limitations of the study will be discussed. Finally, recommendations for future research and clinical practice will be explored.

Research findings in context

These findings bare several similarities to the existing literature exploring SDM in the context of chronic and/or life-limiting medical conditions. Previous meta-synthesis of qualitative studies identified three categories of influences on SDM as reported by both patients and providers of healthcare: 'Patient-centred care', 'Bridging the knowledge gap' and 'Dimensions of decision-making' (Truglio-Londrigan et al., 2014). As in the current study, these categories reflect the interactions between the relational, informational, and systemic components of treatment decisions.

During the writing of this paper, the authors became aware of a recent thematic framework analysis exploring factors influencing DMT prescribing decisions amongst a large

sample of Neurologists and specialist MS nurses across the UK (Cameron et al., 2019). A number of similarities were noted within the themes presented in this and the present study, including the influence of guidelines, individual differences in familiarity and experience and the role of peers/cultures in influencing decisions on the part of clinicians (Cameron et al., 2019). The focus of this present study on clinicians' phenomenological experience of SDM through the use of IPA was a significant difference and thus maintains a novel position.

The above complexity experienced by clinicians mirrors that experienced by patients with MS in a recent study (Matthews, n.d.). The requirement for patients to acknowledge a need for medication to protect against unknown future consequences in the absence of present symptoms may reflect some of the uncertainty expressed by clinicians as they share 'the Unknown Future'. Within this study, patients also appeared to reflect upon the importance of a positive relationship with professionals, specifically their MS nurses - which seemed to evolve from initial support in accepting the diagnosis to the development of a collaborative working relationship as they engage/shared the treatment decision. Other themes included the integration of MS and treatment into personal identity and roles, and the imagined future with and without MS treatment (Matthews, n.d.). Whilst noting the relationship between their findings and behavioural economic theories of decision-making (e.g., Kahneman & Tversky, 1979; Jarmolowicz et al., 2018), the author described a values-driven equation for patients based upon the perceived practical 'fit' of treatment with their life. The findings of the present study suggest a developing awareness and depth of insight of clinicians into the context of patients as highlighted in 'Individual in context'.

Dennison et al. (2016) reported that patients with MS displayed a level of ambivalence to the uncertainty of prognosis, expressing little desire or attempt to seek more information about the future progression of their disease. The findings from this qualitative study of 15 patients contrasts the wider literature emphasising the importance for greater empowerment and knowledge for patients. Authors questioned whether this in itself reflected

patient denial against the threat of future MS progression (Dennison et al, 2016). The exchange of information is a fundamental part of the SDM process (Edwards et al., 2009). The present study provided some insight into clinicians' interpretations of how patients may differ in their experience of information being shared within the theme 'Knowledge is not straightforward' – either seeking more and undertaking their own research in the absence of clinicians having a 'better option' or having the potential to find information as overwhelming.

Making decisions within healthcare is described as a task dominated by uncertainty, with decisions often being made based on incomplete information and probabilistic outcomes (Groot Koerkamp et al., 2010). The present study's finding relating to the challenge clinicians experience in being unable to offer certainty or reassurance for their patients about the future ('Unknown Future') echoed those identified within oncology (Dong et al, 2016), end of life care for dementia (Lamahewa et al, 2017) and palliative care for chronic heart disease (Brannstrom et al., 2011). Lamahewa and colleagues (2017) identified that clinician fears about explaining or communicating to family members the unpredictability of disease progression/trajectory may result in a loss of trust. This may lead to implications for future practice and support for clinicians to manage the experience being unable to reassure their patients.

Across the wider literature, the importance of the physician-patient relationship in facilitating collaborative decision-making is commonly cited (e.g., Ben-Zacharia, 2018), and is echoed in the theme 'Navigating the therapeutic relationship'. The development and maintenance of a therapeutic alliance is reliant upon communication and the ability of the clinician to become attuned to, gather and assess patient context, perspectives, and expectations (Dong et al, 2016; Frerichs et al, 2016). Patient perceptions about their own role and expertise relative to clinicians, trust and clinician interpersonal style were found to be important interactional factors in the development of a therapeutic working relationship which facilitates SDM (Joseph-Williams et al., 2014).

Literature has also reflected upon the importance of the way in which a diagnosis is communicated and the impact this has upon a patient's subsequent adjustment and acceptance (Truglio-Londrigan et al, 2014), with patients typically seeking clear unambiguous information as early as possible (Solari et al, 2007; Messina et al., 2015). Within the present study, patient acceptance of their diagnosis was one factor which clinicians displayed awareness of, influencing the amount and timing of the information they share with their patients. This demonstrated the positive ability of clinicians to attune to their patients and adjust their practice.

Findings relating to the experience of clinicians as they try to adapt and balance their personal values and beliefs with the external regulation and barriers of the system within 'Values, self and the system' are highlighted by Lemahewa et al. (2017). The importance of clinician's self-efficacy, perceived ability to make a difference and their professional fulfilment has been found to influence the level of engagement and participation of clinicians in developing healthcare services (Lindgren et al., 2013). It is suggested that clinicians may benefit from opportunities to explore the way in which personal values and expectations may influence their relationship with patients and the decisions they make.

The challenge of balancing apparent competing demands between the wish to protect and provide care for their patients, in the context of systemic pressures, is highlighted within these experiences. Literature suggests that clinicians may be vulnerable to high levels of self-criticism and guilt when they perceive themselves as not being able to live up to the personal and role expectations. Childers & Arnold (2019) identify that the emphasis upon empathic exploration of intense emotions experienced by patients and their families may place additional burden on clinicians. Balancing this alongside the rational, clinical and evidence-driven perspective can lead to significant challenge, especially in the absence of space and time to process and validate their own emotional reactions. Whilst clinicians in the

present study identified a value in working as a team, this appeared to stem from the practical validation/support provided by colleagues and less explicitly, the emotional containment.

Strengths and limitations of study

It is important to consider both the strengths and limitations of this current study. Firstly, the selection of an IPA methodology with its idiographic and individualised focus has encouraged the collection of rich, in-depth data which has been able to provide insight into the lived experience of the clinicians who participated in this study. However, a limitation of this design and approach is in the interpretative nature of data analysis. It is acknowledged that the interviews and their content has been analysed through the personal lens of the researcher, as a Trainee Clinical Psychologist with significant recent familial experiences serious and life-threatening health difficulties. Therefore, the interpretations presented here may be different to those of another with differing personal and professional perspectives. Due to these recognised limitations measures were put in place to monitor validity and credibility.

This study targeted the recruitment of a small sample of specific clinicians working closely with patients with MS. This was curtailed by the onset of the global Coronavirus pandemic, meaning that recruitment did not meet the minimum target of six. A strength of the IPA approach is in the potential to engage with participants and to obtain a rich account and insight into the meaning and sense-making. This was able to be achieved, and was perhaps enhanced due to the limited numbers. Whilst the researcher remains cautious in the stating of conclusions and implications for clinical practice, some tentative suggestions are offered below.

Whilst all participants were Consultant Neurologists a significant level of heterogeneity was identified as both a limitation and a strength of this study. This heterogeneity arose mainly from the varying levels of experience, as well as sociodemographic backgrounds which were less able to be commented upon in any depth due to confidentiality.

Implications for clinical practice and future research

Due to the limited sample size, the exploratory nature of this study, and the context within which it was conducted, the author would express caution in drawing conclusions and implications without further research. However, at a local level, there may be implications which could support clinicians in their engagement with SDM. In a healthcare system which typically prioritises evidence-based practice and standardisation, there are well-established processes and guidance for treatment decisions. The role of emotions within clinical decision-making is an area of increasing interest within medical education literature and which may benefit from further research and exploration (Kozwolski et al, 2017). NICE guidelines for Supportive and Palliative care (2004) state that all staff should be provided with opportunities to explore their personal and professional responses to loss through clinical supervision and/or reflective practice. However, it is noted that this recommendation no longer appears in the updated guidance from October 2019 (NICE).

Whilst acknowledging that detachment from emotions is a necessary position in enabling medical professionals to undertake their roles (e.g., Kerasidou & Horn, 2016), the opportunity to engage with and develop an awareness of the emotional impact of working in health settings, and specifically with those with MS, may be of benefit to staff. No more so than currently as the NHS manages the emotional impact of the coronavirus pandemic. Self-compassion is viewed to be an integral component of compassionate healthcare delivery. Establishing a safe and reflective space to explore their personal expectations and values,

and the perceived emotional unpredictability of caring for people with MS may encourage greater staff wellbeing and patient satisfaction. Alongside clinical supervision and reflective practice facilitated by Clinical Psychologists, Schwartz Rounds an initiative largely seen within palliative care settings in the UK may be one such forum. Positive effects from witnessing/sharing vulnerability have been reported in relation to staff connectedness (Reed et al., 2015) and, facilitating integration and letting go of emotions for hospice staff (Almeida, 2017). Increased access to a Clinical Psychologist within local Neurology departments may serve two functions. Firstly, offering direct therapeutic support and intervention for patients during the decision-making process as discussed by Matthews (unpublished thesis). Secondly to provide support to clinicians to combat the psychological difficulties that may occur because of or which may influence SDM in MS.

Future research may look to explore the perspectives of other professional groups involved within the process of MS DMT decisions and there may be scope for an integrative study of patient and clinician perspectives to establish similarities, and differences, in their experiences.

Conclusion

This research provides a valuable addition to literature in relation to the experiences of clinicians as they participate in SDM. Implications for future service development to enhance the wellbeing of clinicians and thus their ability to engage compassionately and collaboratively with their patients in the context of medication decisions may be suggested as valuable. However further research and synthesis of these findings with those of MS patients may provide additional insight and improvements to staff and patient care.

References

Almeida, J. C. (2017) *"Oh you're part of the human gang too" the experience of Schwartz Center Rounds ®: an initiative to promote compassionate healthcare.* (Unpublished doctoral thesis). University of Birmingham. Retrieved from https://etheses.bham.ac.uk//id/eprint/6952/

Ben-Zacharia, A., Adamson, M., Boyd, A., Hardeman, P., Smrtka, J., Walker, B., & Walker, T. (2018). Impact of Shared Decision Making on Disease-Modifying Drug Adherence in Multiple Sclerosis. *International Journal of MS care*, *20*(6), 287-297. https://doi.org/10.7224/1537-2073.2017-070

Borg Xuereb, C., Shaw, R.L. & Lane, D.A. (2016). Patients' and physicians' experiences of atrial fibrillation consultations and anticoagulation decision-making: A multi-perspective IPA design. *Psychology & Health*, 31(4), 436-455. https://doi.org/10.1080/08870446.2015.1116534

Brännström, M., Forssell, A., & Pettersson, B. (2011). Physicians' Experiences of Palliative Care for Heart Failure Patients. *European Journal of Cardiovascular Nursing, 10*(1), 64-69. https://doi.org/10.1016/j.ejcnurse.2010.04.005

Brocki, J. M., & Wearden, A. J. (2006). A critical evaluation of the use of interpretative phenomenological analysis (IPA) in health psychology. *Psychology and Health*, *21*(1), 87-108. https://doi.org/10.1080/14768320500230185

Brogan, P., Hasson, F., & McIlfatrick, S. (2017). Shared decision-making at the end of life: a focus group study exploring the perceptions and experiences of multi-disciplinary healthcare professionals working in the home setting. *Palliative Medicine, NA*, 1-10. https://doi.org/10.1177/0269216317734434

Bruce, J. M., Bruce, A. S., Catley, D., Lynch, S., Goggin, K., Reed, D., Lim, S., Strober, L., Glusman, M., Ness, A. R., & Jarmolowicz, D. P. (2016). Being Kind to Your Future Self: Probability Discounting of Health Decision-Making. *Annals of Behaviour Medicine, 50,* 297-309. https://doi.org/10.1007/s12160-015-9754-8

Bruce, J. M., Bruce, A. S., Lynch, S., Thelen, J., Lim, S., Smith, J., Catley, D., Reed, D. D., & Jarmolowicz, D. P. (2018). Probability Discounting of Treatment Decisions in Multiple

Sclerosis: Associations with Disease Knowledge, Neuropsychiatric Status and Adherence. *Psychopharmacology, 235*(11), 3303-3313. https://doi.org/10.1007/s00213-018-5037-y

Cameron, E., Rog, D., McDonnell, G., Overell, J., Pearson, O., & French, D. P. (2019). Factors influencing multiple sclerosis disease-modifying treatment prescribing decisions in the United Kingdom: A qualitative interview study. *Multiple Sclerosis and Related Disorders, 27*, 378-382. https://doi.org/10.1016/j.msard.2018.11.023

Charles, C., Gafni, A., & Whelan, T. (1999). Decision-making in the physician–patient encounter: revisiting the shared treatment decision-making model. *Social Science & Medicine*, *49*(5), 651-661. https://doi.org/10.1016/s0277-9536(99)00145-8

Childers, J., & Arnold, B. (2019). The Inner Lives of Doctors: Physician Emotion in the Care of the Seriously Ill. *The American Journal of Bioethics, 19*(12), 29-34. https://doi.org/10.1080/15265161.2019.1674409

Compston, A., & Coles, A. (2008). Multiple Sclerosis. *The Lancet, 372*(9648), 1502-1517. https://doi.org/10.1016/S0140-6736(08)61620-7

Costello, K., Kennedy, P., & Scanzillo, J. (2008). Recognising nonadherence in patients with multiple sclerosis and maintaining treatment adherence long term. *Medscape Journal of Medicine, 10*(9), 225.

Dennison, L., McCloy Smith, E., Bradbury, K., & Galea, I. (2016). How do people with Multiple Sclerosis experience Prognostic Uncertainty and Prognosis Communication? A Qualitative Study. *PLoS ONE, 11*(7), 1-14. https://doi.org/10.1371/journal.pone.0158982

Dong, S. T., Butow, P. N., Agar, M., Lovell, M. R., Boyle, F., Stockler, M., ... & Tong, A. (2016). Clinicians' perspectives on managing symptom clusters in advanced cancer: a semistructured interview study. *Journal of pain and symptom management*, *51*(4), 706-717. https://doi.org/10.1016/j.jpainsymman.2015.11.021

Edwards, M., Davies, M., & Edwards, A. (2009). What are the external influences on information exchange and shared decision-making in healthcare consultations: a

meta-synthesis of the literature. *Patient education and counseling*, *75*(1), 37-52. https://doi.org/10.1016/j.pec.2008.09.025

Elwyn, G., Durand, M. A., Song, J., Aarts, J., Barr, P. J., Berger, Z., … & Han, P. K. (2017). A three-talk model for shared decision-making: multistage consultation process. *BMJ, 359,* 1-7. https://doi.org/10.1136/bmj.j4891

Fenu, G., Lorefice, L., Frau, F., Coghe, G. C., Marrosu, M. G., & Cocco, E. (2015). Induction and escalation therapies in multiple sclerosis. *Anti-Inflammatory & Anti-Allergy Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Inflammatory and Anti-Allergy Agents)*, *14*(1), 26-34. https://doi.org/10.2174/1871523014666150504122220

Freedman, M. S., Selchen, D., Arnold, D. L., & Prat, A. (2013). Treatment Optimization in MS: Canadian MS Working Group Updated Recommendations. *Canadian Journal of Neurological Sciences, 40*(3), 307-323. https://doi.org/10.1017/S0317167100014244

Frerichs, W., Hahlweg, P., Müller, E., Adis, C., & Scholl, I. (2016). Shared Decision-Making in Oncology – A Qualitative Analysis of Healthcare Providers' Views on Current Practice. *PloS ONE, 11*(3), e0149789. https://doi.org/10.1371/journal.pone.0149789

GBD 2016 Multiple Sclerosis Collaborators (2019). Global, regional and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology, 18*, 269-285. https://doi.org/10.1016/S1474-4422(18)30443-5

Greb, E. (2012). Half of Patients with MS do not take Disease-Modifying Medication. *Neurology Reviews, 20*(2), 5.

Groot Koerkamp, B., Weinstein, M. C., Stijnen, T., Heijenbrok-Kal, M. H., & Hunink, M. M. (2010). Uncertainty and patient heterogeneity in medical decision models. *Medical Decision Making*, *30*(2), 194-205. https://doi.org/10.1177/0272989X09342277

Ha-Vinh, P., Nauleau, S., Clementz, M., Régnard, P., Sauze, L., & Clavaud, H. (2019). Health care services and disease modifying therapies use in community-based multiple sclerosis patients: Evolution from 2013 to 2015 and demographic

characteristics. *La Presse Médicale*, *48*(1), 1-19. http://doi.org/10.1016/j.lpm.2018.04.013

Heesen, C., Kleiter, I., Meuth, S. G., Krämer, J., Kasper, J., Köpke, S., & Gaissmaier, W. (2017). Benefit-risk perception of natalizumab therapy in neurologists and a large cohort of multiple sclerosis patients. *Journal of Neurological Sciences, 376,* 181-190. https://doi.org/10.1016/j.jns.2017.03.001

Heesen, C., Kleiter, I., Nguyen, F., Schäffler, N., Kasper, J., Köpke, S., & Gaissmaier, W. (2010). Risk perception in natalizumab-treated multiple sclerosis patients and their neurologists. *Multiple Sclerosis Journal*, *16*(12), 1507-1512. https://doi.org/10.1177/1352458510379819

Heesen, C., Köpke, S., Solari, A., Geiger, F., & Kasper, J. (2013). Patient autonomy in multiple sclerosis – Possible goals and assessment strategies. *Journal of the Neurological Sciences, 331*(1-2), 2-9. https://doi.org/10.1016/j.jns.2013.02.018

Imo, U. O. (2017). Burnout and psychiatric morbidity among doctors in the UK: a systematic literature review of prevalence and associated factors. *British Journal Psychiatry Bulletin, 41*, 197-204. https://doi.org/10.1192/pb.bp.116.054247

Inojosa, H., Schriefer, D., & Ziemssen, T. (2020). Clinical outcome measures in multiple sclerosis: A review. *Autoimmunity Reviews, 19*(5), 102512. https://doi.org/10.1016/j.autrev.2020.102512

Jarmolowicz, D. P., Reed, D. D., Bruce, A. S., Lynch, S., Smith, J., & Bruce, J. M. (2018). Modeling effects of side-effect probability, side-effect severity, and medication efficacy on patients with multiple sclerosis medication choice. *Experimental and Clinical Psychopharmacology, 26*(6), 599-607. https://doi.org/10.1037/pha0000220

Joseph-Williams, N., Elwyn, G., & Edwards, A. (2014). Knowledge is not power for patients: a systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. *Patient Education and Counseling*, *94*(3), 291-309. https://doi.org/10.1016/j.pec.2013.10.031

Jongen, P. J. (2017). Health-Related Quality of Life in Patients with Multiple Sclerosis: Impact of Disease-Modifying Drugs. *CNS Drugs, 31*, 585-602. https://doi.org/10.1007/s40263-017-0444-x

Joosten, E. A. G., DeFuentes-Merillas, L., de Weert, G. H., Sensky, T., vaan der Staak, C. P. F., & de Jong, C. A. J. (2008). Systematic Review of the Effects of Shared Decision-Making on Patient Satisfaction, Treatment Adherence and Health Status. *Psychotherapy and Psychosomatics, 77*(4), 219-226. https://doi.org/10.1159/000126073

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica. 47*(2), 263-291.

Kamm, C. P., Uitdehaag, B. M., & Polman, C. H. (2014). Multiple Sclerosis: Current Knowledge and Future Outlook. *European Neurology, 72*, 132-141. https://doi.org/10.1159/000360528

Kerasidou, A., & Horn, R. (2016). Making space for empathy: supporting doctors in the emotional labour of clinical care. *BMC Medical Ethics*, *17*(1), 1-5. https://doi.org/10.1186/s12910-016-0091-7

Kingwell, E., Leray, E., Zhu, F., Petkau, J., Edan, G., Oger, J., & Tremlett, H. (2019). Multiple sclerosis: effect of beta interferon treatment on survival. *Brain, 142*(5), 1324-1333. https://doi.org/10.1093/brain/awz055

Kozlowski, D., Hutchinson, M., Hurley, J., Rowley, J., & Sutherland, J. (2017). The role of emotion in clinical decision making: an integrative literature review. *BMC Medical Education, 17*, 255. https://doi.org/10.1186/s12909-017-10897

Kroencke, D. C., Denney, D. R., & Lynch, S. G. (2001). Depression during exacerbations in multiple sclerosis: the importance of uncertainty. *Multiple Sclerosis Journal*, *7*(4), 237-242. https://doi.org/10.1177/135245850100700405

Légaré, F., & Witteman, H. O. (2013). Shared Decision-Making: Examining Key Elements and Barriers to Adoption into Routine Clinical Practice. *Health Affairs, 32*(2), 276-283. https://doi.org/10.1377/hlthaff.2012.1078

Lamahewa, K., Mathew, R., Iliffe, S., Wilcock, J., Manthorpe, J., Sampson, E. L., & Davies, N. (2017). A qualitative study exploring the difficulties influencing decision making at the end of life for people with dementia. *Health Expectations, 21,* 118-127. https://doi.org/10.1111/hex.12593

Li, P., Ladage, V. P., Berger, J., Chahin, S., Jhaveri, M., Geremakis, C., Doshi, J. A. (2020). Disease-Modifying Therapy adherence and associated factors in a national sample of Medicare patients with Multiple-Sclerosis. *Value in Health, 23*(3), 328-334. https://doi.org/10.1016/j.jval.2019.10.011

Lindgren, Å., Bååthe, F., & Dellve, L. (2013). Why risk professional fulfilment: a grounded theory of physician engagement in healthcare development. *The International Journal of Health Planning and Management*, *28*(2), e138-e157. **https://doi.org/10.1002/hpm.2142**

Makoul, G., & Clayman, M. L. (2006). An integrative model of shared decision making in medical encounters. *Patient Education and Counseling*, *60*(3), 301-312. https://doi.org/10.1016/j.pec.2005.06.010

Matthews, J. (n.d). '*The lived experience of shared medical decision making in people with MS.'* (Unpublished doctoral thesis) University of Birmingham, Birmingham.

Messina, M. J., Dalla Costa, G., Rodegher, M., Moiola, L., Colombo, B., Comi, G., & Martinelli, V. (2015). The communication of multiple sclerosis diagnosis: The patients' perspective. *Multiple Sclerosis International*, *2015*, 1-7. https://doi.org/10.1155/2015/353828

NHS, (2019, January 31). *Shared decision making: summary guide.* Retrieved from https://www.england.nhs.uk/publication/shared-decision-making-summary-guide/

National Institute for Health and Care Excellence (2004). *Improving supportive and palliative care for adults with cancer: Cancer service guideline [CSG4].* Retrieved from https://www.nice.org.uk/guidance/csg4

National Institute for Health and Care Excellence (2019). *End of life care for adults: service delivery. NICE guideline [NG142].* Retrieved from https://www.nice.org.uk/guidance/ng142

National Institute for Health and Care Excellence (2019). *Shared Decision Making (Key therapeutic topic 23)*. Retrieved from https://www.nice.org.uk/advice/ktt23/resources/shared-decision-making-pdf-58758011521477

Noble, H., Brazil, K., Burns, A., Hallahan, S., Normand, C., Roderick, P., Thompson, C., Maxwell, P., & Yaqoob, M. (2017). Clinician vies of patient decisional conflict when deciding between dialysis and conservative management: Qualitative findings deom the PAlliative Care in chronic Kidney diSease (PACKS) study. *Palliative Medicine, 31*(10), 921-931. https://doi.org/10.1177/026921637704625

Remington, G., Rodriguez, Y., Logan, D., Williamson, C., & Treadaway, K. (2013). Facilitating Medication Adherence in Patients with Multiple Sclerosis. *International Journal of MS Care, 15,* 36-45. https://doi.org/10.7224/1537-2073.2011-038

Samalin, L., Genty, J., Boyer, L., Lopez-Castroman, J., Abbar, M., & Llorca, P. (2018) Shared Decision-Making: a Systematic Review focusing on Mood Disorders. *Current Psychiatry Reports, 20*(23), 1-11. https://doi.org/10.1007/s11920-018-0892-0

Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making, 16*(138)*,* 1-14. https://doi.org/10.1186/s12911-016-0377-1

Shafir, A., & Rosenthal, J. (2012). *Shared Decision-Making: Advancing Patient-Centred Care through State and Federal Implementation (pp 1-38).* Washington, DC: National Academy for State Health Policy

Shaw, R. (2010). Embedding reflexivity within experiential qualitative psychology. *Qualitative Research in Psychology, 7*(3), 233-243. https://doi.org/10.1080.14780880802699092

Smith, J. A., Flowers, P., & Larkin, M. (2009). Interpretative Phenomenological Analysis: Theory, Method and Research. Thousand Oaks, CA: Sage Publications.

Solari, A., Acquarone, N., Pucci, E., Martinelli, V., Marrosu, M. G., Trojano, M., Borreani, C., Messmer Uccelli, M. (2007). Communicating the diagnosis of multiple sclerosis – a qualitative study. *Multiple Sclerosis, 13*, 763-769. https://doi.org/10.1177/1352458506074689

The King's Fund (2011). *Making shared decision-making a reality: No decision about me, without me.* Retrieved from https://www.kingsfund.org.uk/publications/making-shared-decision-making-reality

Torkildsen, Ø, Myhr, K. M., & Bø, L. (2015). Disease-modifying treatments for multiple sclerosis – a review of approved medications. *European Journal of Neurology, 23*(S1), 18-27. https://doi.org/10.1111/ene.12883

Treadaway, K., Cutter, G., Salter, A., Lynch, S., Simsarian, J., Corboy. J., …, Frohman, E. M. (2009). Factors that influence adherence with disease-modifying therapy in MS. *Journal of Neurology, 256*(4), 568-576. https://doi.org/10.1007/s00415-009-0096-y

Truglio-Londrigan, M., Slyer, J. T., Singleton, J. K., & Worral, P. S. (2014). A qualitative systematic review of internal and external influences on shared decision-making in all health care settings. *JBI Database of Systematic Reviews and Implementation Reports, 10*(58), 4633-4646. https://doi.org/10.11124/jbisrir-2012-432

van Capelle, A. D. C., van der Meide, H., Vosman, F. J., & Visser, L. H. (2017). A qualitative study assessing patient perspectives in the process of decision-making on disease modifying therapies (DMT's) in multiple sclerosis (MS). *PloS one*, *12*(8), e0182806. https://doi.org/10.1371/journal.pone.0182806

Vargas, D. L., & Tyor, W. R. (2017). Update on Disease-Modifying Therapies for Multiple Sclerosis. *Journal of Investigative Medicine, 65*(5), 883-891. https://doi.org/10.1136/jim-2016-000339

Visser, L. H., & Van Der Zande, A. (2011). Reasons patients give to use or not to use immunomodulating agents for multiple sclerosis. *European Journal of Neurology*, *18*(11), 1343-1349. https://doi.org/10.1111/j.1468-1331.2011.03411.x

Wennberg, D. E., Marr, A., Lang, L., O'Malley, S., & Bennett, G. (2010). A Randomised Trial of a Telephone Care Management Strategy. *The New England Journal of Medicine, 363,* 1245-1255. https://doi.org/10.1056/NEJMsa0902321

Wilson, L., Loucks, A., Bui, C., Gipson, G., Zhong, L., Schwartzburg, A., Crabtree, E., Goodin, D., Waubant, E., & McCulloch, C. (2014). Patient centered decision making: Use of conjoint analysis to determine risk-benefit trade-offs for preference sensitive

treatment choices. *Journal of the Neurological Sciences, 344*(1-2), 80-87. https://doi.org/10.1016/j.jns.2014.06.030

Yardley,L. (2000). Dilemmas in qualitative health research. *Psychology and Health, 15(*2), 215-228. https://doi.org/10.1080/08870440008400302

Literature Review Press Release

Spotting those we ask to exaggerate

The Word Memory Test (WMT) can accurately detect people who are told to deliberately perform poorlyduring cognitive assessment, a recent meta-analysis from the University of Birmingham shows.

For many the idea of deliberately underperforming in our day to day lives or medical examinations (such as our eye test) is surprising. However, it is widely acknowledged that some people may seek to exaggerate difficulties for a wide variety of reasons. In some sectors, this may lead to the mistaken awarding of financial compensation or potential avoidance of criminal charges. Within healthcare, this leaves professionals with a problem as incorrect diagnosis can lead to unnecessary investigation and treatment. Psychologists who work with people with language, memory and planning difficulties (often following head injury or infections) routinely use additional tests to check that people are trying their best during psychology assessments. Recent guidance suggests that as many as three of these so called “effort tests” are used during assessments (Sherman et al., 2020). The WMT is one of the most widely used measures of effort and has been described as accurate in detecting those who are offering minimal effort across individual published studies.

When trying to explore how accurate a test is, researchers typically compare the test in question with a “gold standard” alternative. Unfortunately, no such “gold standard” effort test has been found to compare to. For this reason, researchers design studies which artificially assign people to different groups to find out if the test can tell between them.

Researchers at the University of Birmingham analysed data from a total of 18 studies in published research papers, which used the WMT to spot people who were asked to pretend to have a brain injury (simulators) alongside those who were asked to try their best (some of whom were already diagnosed with genuine impairment or injury). Their findings

suggested that not only was the WMT able to correctly identify up to 95% of simulators, but that this was broadly resistant to attempts to coach participants how to ‘beat the test’ and the offer of financial reward.

Lead researcher Rebecca Gordon says: “Whilst the available evidence appears strong, the conclusions are limited by the quality of the data which is reported in published studies. The WMT can be a highly effective way of detecting poor effort in people asked to ‘fake’ impairments in brain functions such as memory and language. As with any test, professionals should remain cautious to ensure that they use the WMT in line with the published instructions”.

Empirical Paper Press Release

“Just get on with it” – experiences of doctors as they navigate Shared Decision-Making

Clinicians working with patients with Multiple Sclerosis (MS) may need additional support to maintain their wellbeing as they attempt to offer collaborative and supportive care to those making decisions about treatment say researchers from the University of Birmingham.

MS is chronic health condition which affects approximately 2.2 million people worldwide (Wallin & GBD 2016 Multiple Sclerosis Collaborators, 2019). For those living with MS, many will experience episodes of so-called acute symptoms which may last for periods of weeks or months before making at least a partial recovery. However, the type and frequency of these episodes are recognised to have significant effects on people’s lives including their ability to engage with work and socialise. Although there is no cure for MS, there is a group of treatments known as ‘Disease-Modifying Therapies’ (DMTs) which aim to lower the number and intensity of acute episodes (relapses). A number of DMTs are available however as with any medications, these come with their own side-effects.

Changing attitudes within healthcare settings now recommend much more strongly that doctors and nurses ‘share’ decisions about medications with their patients rather than imposing them. However, there are several differences in the way this ‘sharing’ takes place in hospitals. A series of four interviews were carried out with doctors who regularly take part in decisions about whether to start or change MS treatment to find out what this process was like for them.

Doctors described that sharing decisions about medication relied upon having a good professional relationship with their patient. This meant that they needed to consider not only their patients’ physical health, but also the wider story of their patients’ lives – including

family responsibilities, working life. Sometimes this involved letting their patients make decisions that they may not agree with – but sometimes this needed them to be more assertive in order to protect their patients.

Doctors spoke of the challenges they faced because of the unpredictability of MS symptoms and the side-effects associated with the medications they prescribe. They described that whilst the sharing of information with their patients was important, they were not always able to control the quality of information that patients find by themselves – such as online. This information could hold the potential to be harmful either to the patient or their relationship with their doctor.

Crucially, doctors identified a hope to help their patients and shared how difficult this can be for them when they aren't able to do this, either due to large workloads or restrictions in the treatments they can offer. However, these emotions were rarely acknowledged and doctors described "just getting on with it". Findings from this research suggested that the offer of support to manage these emotions in a safe way for doctors may be beneficial.

Lead researcher, Rebecca Gordon stated: "These findings have important implications locally for the wellbeing of doctors and wider professionals. We expect them to be able to support patients and their families with difficult emotions and decisions and yet there is little acknowledgement or support for their own. If we are to maintain a workforce which can continue caring, especially as we recover from the recent Coronavirus pandemic, there is a responsibility for services to offer them support too".

References

GBD 2016 Multiple Sclerosis Collaborators (2019). Global, regional and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology, 18*, 269-285. https://doi.org/10.1016/S1474-4422(18)30443-5

Appendices

Appendix 1: Search Terms by Database

	Diagnostic Accuracy	Tests	Word Memory Test
PsycInfo	Sensitivity [OR] specificity [OR] test sensitivity (exp) [OR] test specificity (exp) [OR] valid* [OR] test validity (exp) [OR] psychometrics (exp) [OR] statistical validity (exp) [OR] predictive value* [OR] PPV [OR] NPV [OR] likelihood ratio* [OR] likelihood functions [OR] reference value* [OR] area under curve [OR] AUC [OR] ROC [OR] diagnos* [OR] false positive [OR] false negative [OR] psychodiagnostic measures (exp) [OR] (diagnos* adj3 (performance* OR accurac* OR utilit* OR value* OR effectiveness OR efficien*))	Symptom validity [OR] performance validity [OR] effort test*	Word Memory Test [OR] WMT
Medline	Sensitivity [OR] specificity [OR] sensitivity and specificity (exp) [OR] test sensitivity [OR] test specificity [OR] valid* [OR] reproducibility of results (exp) [OR] psychometrics (exp) [OR] false positive [OR] false positive reactions (exp) [OR] false negative [OR] false negative reactions [OR] predictive value* [OR] predictive value of tests (exp) [OR] PPV [OR] NPV [OR] reference value [OR] reference values (exp) [OR] likelihood ratio* [OR] likelihood functions [OR] ROC [OR] ROC curve (exp) [OR] AUC [OR] area under curve (exp) [OR] accuracy [OR] dimensional measurement accuracy (exp) [OR] data accuracy (exp)	Symptom validity [OR] performance validity [OR] effort test*	Word Memory Test [OR] WMT
Embase	Sensitivity [OR] specificity [OR] sensitivity and specificity (exp) [OR] test sensitivity [OR] test specificity [OR] valid* [OR] test validity [OR] validity (exp) [OR] psychometry (exp) [OR] predictive value* [OR] predictive value* (exp) [OR] PPV [OR] likelihood ratio* [OR] likelihood functions [OR] reference value [OR] reference value (exp) [OR] ROC	Symptom validity [OR] performance validity [OR] effort test*	Word Memory Test [OR] WMT

	Diagnostic Accuracy	Tests	Word Memory Test
	[OR] receiver operating characteristic (exp) [OR] area under the curve (exp) [OR] roc curve (exp) [OR] diagnos* [OR] diagnosis (exp) [OR] false positive result (exp) [OR] false positive [OR] false negative [OR] false negative result (exp) [OR] (diagnos* adj3 (performance* OR accurac* OR utilit* OR value* OR effectiveness OR efficien*))		
Web of Science	Sensitivity [OR] specificity [OR] "sensitivity and specificity" [OR] "test sensitivity" [OR] "test specificity" [OR] valid* [OR] "test validity] [OR] psychometric* [OR] "predictive value*" [OR] PPV [OR] NPV [OR] "likelihood ratio*" [OR] "likelihood functions" [OR] "reference value*" [OR] "receiver operating characteristic" [OR] "area under the curve" [OR] "ROC curve" [OR] diagnos* [OR] "false positive" [OR] "false negative"	Symptom validity [OR] performance validity [OR] effort test*	Word Memory Test [OR] WMT

Appendix 2: Data Extracted from Primary Studies

	Data Extracted
Descriptive Study Information	Year of publication
	Index according to which failure classified
	Simulator and control populations reported
	Sophistication of simulators / Level of instruction
	Presence/absence of incentive
	Cut-off used to determine failure
	Average age of simulator and control populations
	Average number of years education for simulator and control populations
	Mode of administration
	Partial or full administration of WMT
	Study design/Assignment to conditions
	Language
Data for 2x2 contingency Table	Number of simulators identified as 'failing' WMT – Disease Positive
	Number of simulators (Disease n)
	Number of control participants identified as 'failing' WMT – Control Positive
	Number of control participants (Control n)
Criteria for risk of bias assessment	Selection bias
	Selection applicability concerns
	Index test bias
	Index test applicability concerns
	Reference test bias
	Reference test applicability concerns
	Flow and timing

Appendix 3: Criteria defining the level of sophistication/coaching of simulating participants

	Criteria
Sophisticated simulators	Recruited/defined as a clinical population >50% of group have experience of the condition being asked to simulate. Received instructions and information about anticipated features of simulated condition – quiz administered to check understanding of instructions prior to WMT administration

Appendix 4: Univariate Analyses for Immediate Recall, Delayed Recall and Consistency

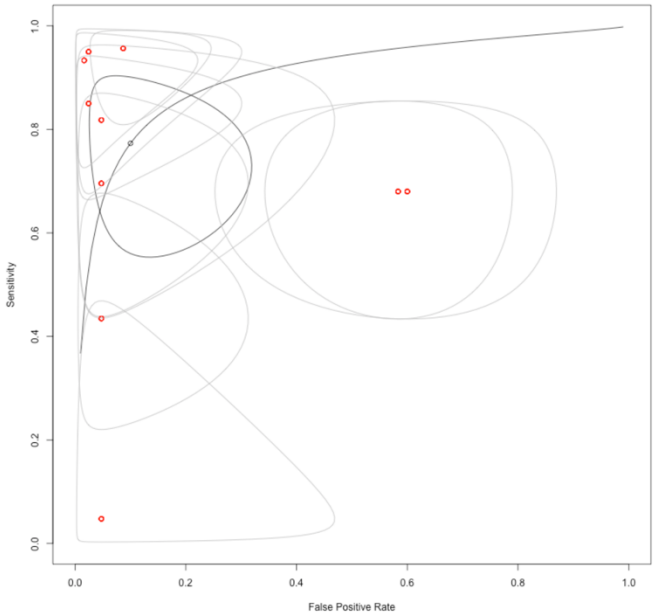
	Immediate Recall				
Study	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR-	DOR
Manual cut-off - <82.5%					
Edmundson, Berry, Combs et al.. (2017) - *Naïve Simulator*	0.696 (0.491 – 0.844)	0.952 (0.773 – 0.992)	14.609 (2.117 – 100.822)	0.320 (0.171 – 0.597)	45.714 (5.086 – 410.924)
Edmundson, Berry, Combs et al.. (2017) - *Soph. Simulator*	0.435 (0.256 – 0.632)	0.952 (0.773 – 0.992)	9.130 (1.275 – 65.384)	0.593 (0.410 – 0.860)	15.385 (1.755 – 134.876)
Elbaum, Golan, Lupu, Wagner & Braw (2019)	0.933 (0.787 – 0.982)	0.984 (0.861 – 0.998)	56.933 (3.637 – 891.173)	0.068 (0.018 – 0.259)	840.000 (36.304 – 19435.974)
Lau, Basso, Estevis, Miller, Whiteside, Coombs & Artensen (2017) – *Naïve Simulator*	0.950 (0.764 – 0.991)	0.976 (0.804 – 0.997)	38.950 (2.516 – 603.002)	0.051 (0.008 – 0.347)	760.000 (24.070 – 23996.314)
Lau, Basso, Estevis, Miller, Whiteside, Coombs & Artensen (2017) – *Sop. Simulator*	0.850 (0.709 – 0.929)	0.976 (0.804 – 0.997)	34.850 (2.248 – 540.202)	0.154 (0.073 – 0.323)	226.667 (12.014 – 4276.385)
Lupu, Elbaum, Wagner & Braw (2018)	0.957 (0.855 – 0.988)	0.913 (0.797 – 0.966)	11.000 (4.304 – 28.116)	0.048 (0.012 – 0.185)	231.000 (40.172 – 1328.312)
Merten, Henry & Hilsabeck (2004)	0.048 (0.005 – 0.333)	0.952 (0.667 – 0.995)	1.000 (0.022 – 45.854)	1.000 (0.826 – 1.211)	1.000 (0.018 – 55.519)
Shandera, Berry, Clark, Schipper, Graue & Harp (2010) – *Healthy Control*	0.680 (0.484 – 0.828)	0.400 (0.168 – 0.687)	1.133 (0.639 – 2.010)	0.800 (0.309 – 2.069)	1.417 (0.310 – 6.470)
Shandera, Berry, Clark, Schipper, Graue & Harp (2010) – *LD Control*	0.680 (0.484 – 0.828)	0.417 (0.245 – 0.612)	1.166 (0.757 – 1.796)	0.768 (0.366 – 1.613)	1.518 (0.472 – 4.882)
Tydecks, Merten & Gubbay (2006a)	0.100 (0.018 – 0.404)	0.952 (0.667 – 0.995)	2.100 (0.079 – 55.945)	0.945 (0.738 – 1.210)	2.222 (0.066 – 74.491)

	Delayed Recall				
Study	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR-	DOR
Manual cut-off - <82.5%					
Edmundson, Berry, Combs et al.. (2017) - *Naïve Simulator*	0.565 (0.368 – 0.744)	0.952 (0.773 – 0.992)	11.870 (1.695 – 83.096)	0.457 (0.284 – 0.735)	26.000 (2.966 – 227.940)
Edmundson, Berry, Combs et al.. (2017) - *Soph. Simulator*	0.304 (0.156 – 0.509)	0.952 (0.773 – 0.992)	6.391 (0.856 – 47.703)	0.730 (0.548 – 0.973)	8.750 (0.973 – 78.653)
Lau, Basso, Estevis, Miller, Whiteside, Coombs & Artensen (2017) – *Naïve Simulator*	0.950 (0.764 – 0.991)	0.976 (0.804 – 0.997)	38.950 (2.516 – 603.002)	0.051 (0.008 – 0.347)	760.000 (24.070 – 23996.314)
Lau, Basso, Estevis, Miller, Whiteside, Coombs & Artensen (2017) – *Soph. Simulator*	0.675 (0.520 – 0.799)	0.976 (0.804 – 0.997)	27.675 (1.776 – 431.280)	0.333 (0.212 – 0.523)	83.077 (4.649 – 1484.678)
Merten, Henry & Hilsabeck (2004)	0.048 (0.005 – 0.333)	0.952 (0.667 – 0.995)	1.000 (0.022 – 45.854)	1.000 (0.826 – 1.211)	1.000 (0.018 – 55.519)
Shandera, Berry, Clark, Schipper, Graue & Harp (2010) – *Healthy Control*	0.600 (0.407 – 0.766)	0.400 (0.168 – 0.687)	1.000 (0.549 – 1.820)	1.000 (0.407 – 2.455)	1.000 (0.224 – 4.468)
Shandera, Berry, Clark, Schipper, Graue & Harp (2010) – *LD Control*	0.600 (0.407 – 0.766)	0.417 (0.245 – 0.612)	1.029 (0.646 – 1.638)	0.960 (0.489 – 1.884)	1.071 (0.343 – 3.349)
Tydecks, Merten & Gubbay (2006b)	0.100 (0.018 – 0.404)	0.952 (0.667 – 0.995)	2.100 (0.079 – 55.945)	0.945 (0.738 – 1.210)	2.222 (0.066 – 74.491)

	Consistency				
Study	Sensitivity (95% CI)	Specificity (95% CI)	LR+	LR-	DOR
Manual cut-off - <82.5%					
Edmundson, Berry, Combs et al.. (2017) - *Naïve Simulator*	0.739 (0.535 – 0.875)	0.857 (0.654 – 0.950)	5.174 (1.765 – 15.166)	0.304 (0.150 – 0.619)	17.000 (3.658 – 79.002)
Edmundson, Berry, Combs et al.. (2017) - *Soph. Simulator*	0.522 (0.330 – 0.708)	0.857 (0.654 – 0.950)	3.652 (1.194 – 11.175)	0.558 (0.352 – 0.885)	6.545 (1.504 – 28.490)
Lau, Basso, Estevis, Miller, Whiteside, Coombs & Artensen (2017e)	0.95 (0.764 – 0.991)	0.976 (0.804 – 0.997)	38.950 (2.516 – 603.002)	0.051 (0.008 – 0.347)	760.000 (24.070 – 23996.314)
Lau, Basso, Estevis, Miller, Whiteside, Coombs & Artensen (2017f)	0.90 (0.769 – 0.960)	0.976 (0.804 – 0.997)	36.900 (2.383 – 571.324)	0.102 (0.040 – 0.260)	360.000 (18.098 – 7161.195)
Merten, Henry & Hilsabeck (2004)	0.048 (0.005 – 0.333)	0.952 (0.667 – 0.995)	1.000 (0.022 – 45.854)	1.000 (0.826 – 1.211)	1.000 (0.018 – 55.519)
Shandera, Berry, Clark, Schipper, Graue & Harp (2010c)	0.80 (0.609 – 0.911)	0.200 (0.057 – 0.510)	1.000 (0.693 – 1.443)	1.000 (0.231 – 4.335)	1.000 (0.160 – 6.255)
Shandera, Berry, Clark, Schipper, Graue & Harp (2010f)	0.80 (0.609 – 0.911)	0.250 (0.120 – 0.449)	1.067 (0.788 – 1.444)	0.800 (0.281 – 2.278)	1.333 (0.347 – 5.127)
Tydecks, Merten & Gubbay (2006c)	0.10 (0.018 – 0.404)	0.952 (0.667 – 0.995)	2.100 (0.079 – 55.945)	0.945 (0.738 – 1.210)	2.222 (0.066 – 74.491)

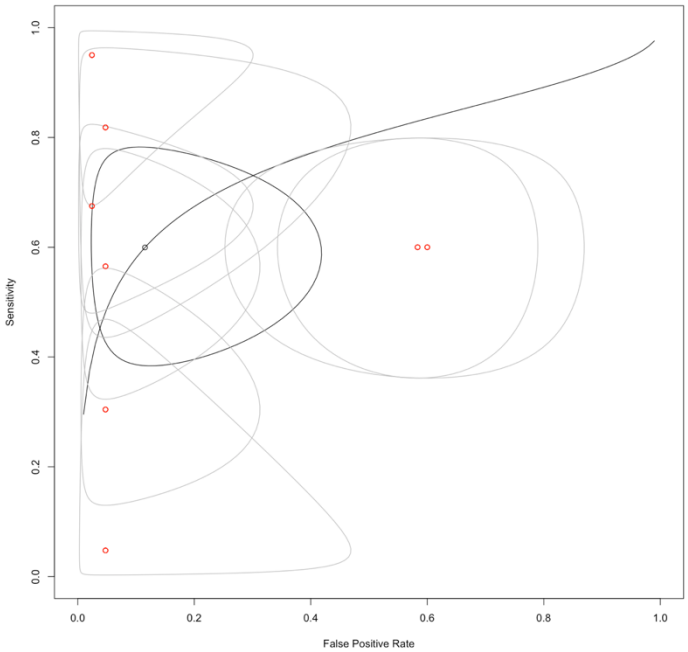
Appendix 5: sROC Curves for Immediate Recall, Delayed Recall and Consistency

Immediate Recall



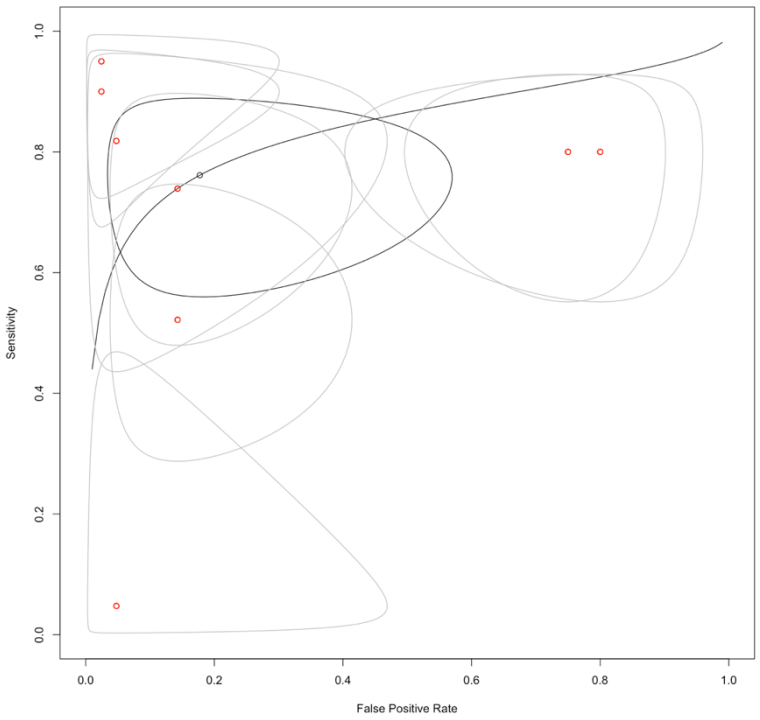
AUC=0.9

Delayed Recall



AUC=0.769

Consistency



AUC=0.834

Appendix 6: University of Birmingham Sponsorship Letter

UNIVERSITY OF BIRMINGHAM

FINANCE OFFICE

Miss Rebecca Gordon
School of Psychology
University of Birmingham

27 March 2019

Dear Miss Gordon

Project Title: **Exploring clinician experiences of shared decision-making in the treatment of Multiple Sclerosis**
IRAS ID: **263792**
Sponsor Reference: **RG_18-292**
UoB Ethics Reference: **ERN_18-2086**

Under the requirements of UK Policy Framework for Health and Social Care Research, the University of Birmingham agrees to act as Sponsor for this project. Sponsorship is subject to you obtaining a favourable ethical opinion, HRA approval and NHS R&D management approval where appropriate.

As Chief Investigator, you must ensure that local study recruitment does not commence until all applicable approvals have been obtained. Where a study is or becomes multi-site you are responsible for ensuring that recruitment at external sites does not commence until local approvals have been obtained.

Following receipt of all relevant approvals, you should ensure that any subsequent amendments are notified to the Sponsor, REC, HRA and relevant NHS R&D Office(s), and that an annual progress report is submitted to the Sponsor, REC and NHS R&D departments where requested.

Please ensure you are familiar with the University of Birmingham Code of Practice for Research (http://www.birmingham.ac.uk/Documents/university/legal/research.pdf) and any appropriate College or School guidelines.

Finally please contact researchgovernance@contacts.bham.ac.uk should you have any queries.

You may show this letter to external organisations.

Yours sincerely

Dr Birgit Whitman
Head of Research Governance and Integrity

cc: Dr Chris Jones; Dr Ruth Howard

University of Birmingham Edgbaston Birmingham B15 2TT United Kingdom
w: www.finance.bham.ac.uk

Appendix 7: Health Research Authority Approval Letter

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Miss Rebecca Gordon
Department of Clinical Psychology, School of Psychology,
University of Birmingham, 52 Pritchatts Road
Birmingham
B15 2TT

Email: hra.approval@nhs.net

29 May 2019

Dear Miss Gordon

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Exploring clinician experiences of shared decision-making in the treatment of Multiple Sclerosis: An Interpretative Phenomenological Analysis**
IRAS project ID:	**263792**
Protocol number:	**RG_18-292**
Sponsor	**University of Birmingham**

I am pleased to confirm that HRA and Health and Care Research Wales (HCRW) Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The attached document *"After HRA Approval – guidance for sponsors and investigators"* gives detailed guidance on reporting expectations for studies with HRA and HCRW Approval, including:

- Registration of Research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **263792**. Please quote this on all correspondence.

Yours sincerely,
Juliana Araujo

Approvals Specialist

Email: hra.approval@nhs.net

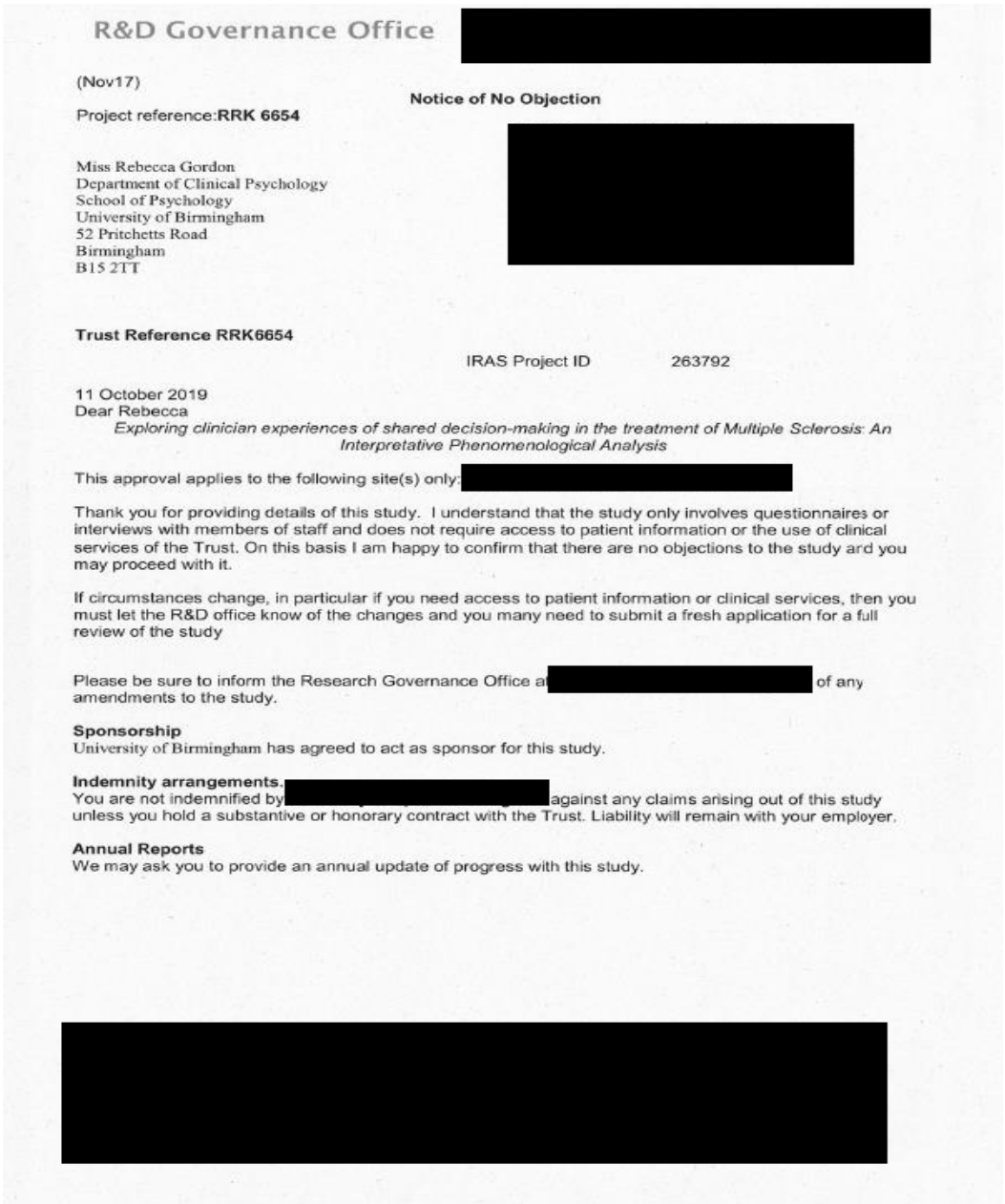
Copy to: *Sponsor Representative: Dr Birgit Whitman, University of Birmingham*

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Copies of advertisement materials for research participants [Email Advertisement]	1	15 July 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Insurance Letter]		
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [2018.2019 - Employers Liability Insurance letter]		02 July 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [2018.2019 - Generic Insurance Letter]		16 July 2018
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [2018.2019 - Public & Products Liability Letter]		03 July 2018
Interview schedules or topic guides for participants [Interview Schedule]	1	15 July 2018
IRAS Application Form [IRAS_Form_28032019]		28 March 2019
Letter from funder [Letter from Sponsor]		
Participant consent form [Participant Consent Form]	1	15 July 2018
Participant information sheet (PIS) [Participant Information Sheet]	1	15 July 2018
Research protocol or project proposal [Project Protocol]	1.3	15 March 2019
Summary CV for Chief Investigator (CI) [Rebecca Gordon CV]	1.0	25 January 2019
Summary CV for supervisor (student research) [Chris Jones - Academic CV]		
Summary CV for supervisor (student research) [Ruth Howard - Academic CV]	1	15 August 2018
263792, 19/HRA/2251, SE13 Application valid under consideration - REC review - non-commercial.eml		29 March 2019
263792, 19/HRA/2251, SE15 Application valid - Non-REC application.eml		29 March 2019

Appendix 8: Research and Development No Objection Letter



R&D Governance Office [redacted]

(Nov17)

Notice of No Objection

Project reference:**RRK 6654**

[redacted]

Miss Rebecca Gordon
Department of Clinical Psychology
School of Psychology
University of Birmingham
52 Pritchetts Road
Birmingham
B15 2TT

Trust Reference RRK6654

IRAS Project ID 263792

11 October 2019
Dear Rebecca

Exploring clinician experiences of shared decision-making in the treatment of Multiple Sclerosis: An Interpretative Phenomenological Analysis

This approval applies to the following site(s) only: [redacted]

Thank you for providing details of this study. I understand that the study only involves questionnaires or interviews with members of staff and does not require access to patient information or the use of clinical services of the Trust. On this basis I am happy to confirm that there are no objections to the study and you may proceed with it.

If circumstances change, in particular if you need access to patient information or clinical services, then you must let the R&D office know of the changes and you many need to submit a fresh application for a full review of the study

Please be sure to inform the Research Governance Office at [redacted] of any amendments to the study.

Sponsorship
University of Birmingham has agreed to act as sponsor for this study.

Indemnity arrangements.
You are not indemnified by [redacted] against any claims arising out of this study unless you hold a substantive or honorary contract with the Trust. Liability will remain with your employer.

Annual Reports
We may ask you to provide an annual update of progress with this study.

[redacted]

R&D Governance Office [redacted]

Yours sincerely,

[redacted]

Clark Crawford
Head of Governance

Copies to:

Service Departments
Division Manager.

Appendix 9: Participant Information Sheet

UNIVERSITY OF BIRMINGHAM

Participant Information Sheet

Project Title: Exploring Clinicians' Experiences of Shared Decision-Making in the Treatment of Multiple Sclerosis

Research Team: Rebecca Gordon (Trainee Clinical Psychologist), Dr Chris Jones, Dr Ruth Howard & [redacted]

We would like to invite you to take part in a research project being undertaken by Rebecca Gordon (Trainee Clinical Psychologist) as part of her Doctorate in Clinical Psychology. This project is interested in the ways in which clinicians understand and make sense of their experiences of making treatment decisions with patients with MS. Before you decide whether or not you wish to participate, it is important that you understand why this research is being conducted and what it will involve if you chose to take part. Please read the following information carefully and ask any questions which you may have.

What is the purpose of this research?
The purpose of the research is to gain a greater understanding of how clinicians experience the decision-making process when deciding on treatment options for Multiple Sclerosis.

Why have I been invited to take part?
You have been invited to take part because you are a medical professional and as part of your role you make decisions with patients about their treatment (for example, choosing to take medication, choosing to change medication or choosing to discontinue medication).

What will happen if I agree to take part?
If you decide to take part, you will be invited to attend an interview with Rebecca which will last approximately 45 minutes to 1 hour. This interview will be audio recorded. Rebecca will ask you about what factors you took into consideration when helping to support a patient to decide about their treatment. We are also interested in understanding what helped you in this process and whether there were any difficulties you faced. Direct quotations from your interview may be used when this study is written up as part of Rebecca's academic qualification; Rebecca will discuss this with you at the time and we will ensure that no identifying details are attached to the quote(s).

What will happen if I do not want to carry on with the study?
Participation in the study is entirely voluntary and you can withdraw from the study at any point before or during the interview, and up to two weeks after the interview. Your working role will not be affected if you do not choose to take part or if you withdraw.

What are the benefits of taking part?
While there are no direct benefits to you by taking part in this study, the findings will help doctors working in MS to support patients in making their treatment decisions. The findings will also be an important addition to the literature on shared decision-making in MS.

Version 1 – July 2018 IRAS ID: 263792

What will happen to the results of the research study?
The results of the study will be written as part of Rebecca's thesis for her Doctorate in Clinical Psychology. The results may be published in an academic journal, and will be shared with those who have an interest in this area of healthcare, including researchers, doctors and patients

Will my taking part be confidential?
Yes. Your interview will be audio recorded and transcribed. The recording of the interview and the transcription will be kept confidential and stored in line with the University of Birmingham's data governance policies.

What will happen to my data?
The University of Birmingham is the sponsor for this study based in the United Kingdom. We will be using information from you in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The University of Birmingham will keep identifiable information about you until May 2020.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we may keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

The University of Birmingham will keep your name and contact details (email address) confidential and will not pass this information to any other organisation. The University of Birmingham will use this information as needed, to contact you about the research study, and make sure that relevant information about the study is recorded to oversee the quality of the study. Certain individuals from regulatory organisations may look at these research records to check the accuracy of the research study. The people who analyse the information will not be able to identify you and will not be able to find out your name or contact information.

What happens if I have any concerns or would like to make a complaint?
If you have any concerns or wish to make a complaint about the research, then you can contact Rebecca's supervisors as follows:

Dr Chris Jones's email: [redacted]
Dr Ruth Howard's email: [redacted]

What happens if I have any questions?
If you would like to discuss any aspect of this research with Rebecca, please contact her by email: [redacted]

Appendix 10: Participant Consent Form

UNIVERSITY OF BIRMINGHAM

Consent Form

Project Title: Exploring Clinicians' Experiences of Shared Decision-Making in the Treatment of Multiple Sclerosis – An Interpretative Phenomenological Analysis

Research Team: Rebecca Gordon (Trainee Clinical Psychologist), Dr Chris Jones, Dr Ruth Howard & [redacted]

Please initial box

1. I confirm that I have read the information sheet dated July 2018 (version 1) for the above study. I have had the opportunity to consider the information, ask questions and had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and I am free to withdraw at any time during the research interview (and up to two weeks afterwards) without giving any reason. My working role will not be affected by any such decision. ☐
3. I understand that the research interview will be audio recorded ☐
4. I understand that the researcher will contact me after a two-week reflection period following my interview. During this conversation, I will be given the opportunity to withdraw my interview entirely or in part, without giving any reason or my working role being affected. ☐
5. I understand that my research interview data will be looked at by the researcher and relevant others from the University of Birmingham to ensure that the analysis is an accurate and fair representation of the data. ☐
6. I understand that direct quotes from my interview may be published in the write-up of the data but that my name and any identifying information such as age, gender or experience will not be given to such quotes. I will not be identifiable by my comments. ☐
7. I agree to take part in the above study. ☐

.....	.....	.....
Name of participant	Date	Signature
.....	.....	.....
Name of researcher	Date	Signature

Appendix 11: Interview Schedule

UNIVERSITY OF BIRMINGHAM

Draft Interview Schedule

Medical Professionals who work with patients affected by MS

Can you tell me about a case where you have recently supported someone with their medication choice?

- What was involved?
- How do you approach these consultations?
- Was anyone else involved?
    - Did you talk to anyone else?
    - What did they say?
    - What did that add?
    - When did you get a chance to talk to them?
- Was this a typical case for you? If not, could you describe a typical case?

Can you tell me about a case which has been challenging, particularly with reference to making a decision about medication?

- What was it that made it more challenging?
- What things did you do to help you manage this?
- What thoughts did you have about the individual making this decision?
- How were you feeling when that happened?
- What was most challenging for you?

What do you think a patient will consider when they are making decisions about their medication?

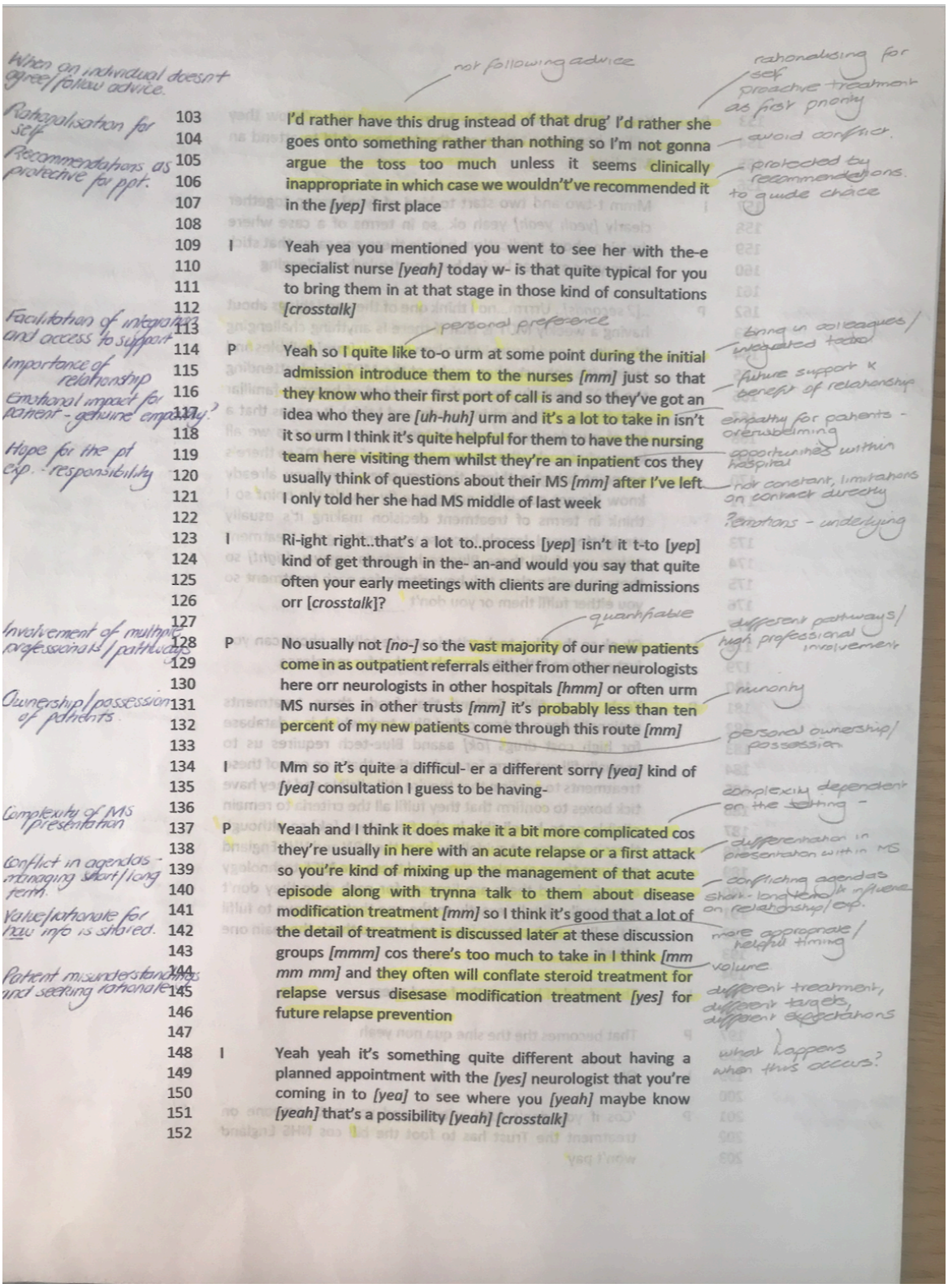
- What is important to them?
- How different is this to your own experience?
- What are the things that feel most important for you and your patient to consider when making these decisions?

What is the most difficult part of supporting a person to make decisions about their medication?

- How do you think about your role in this?
- Is it the same for all patients?
- What may be different?

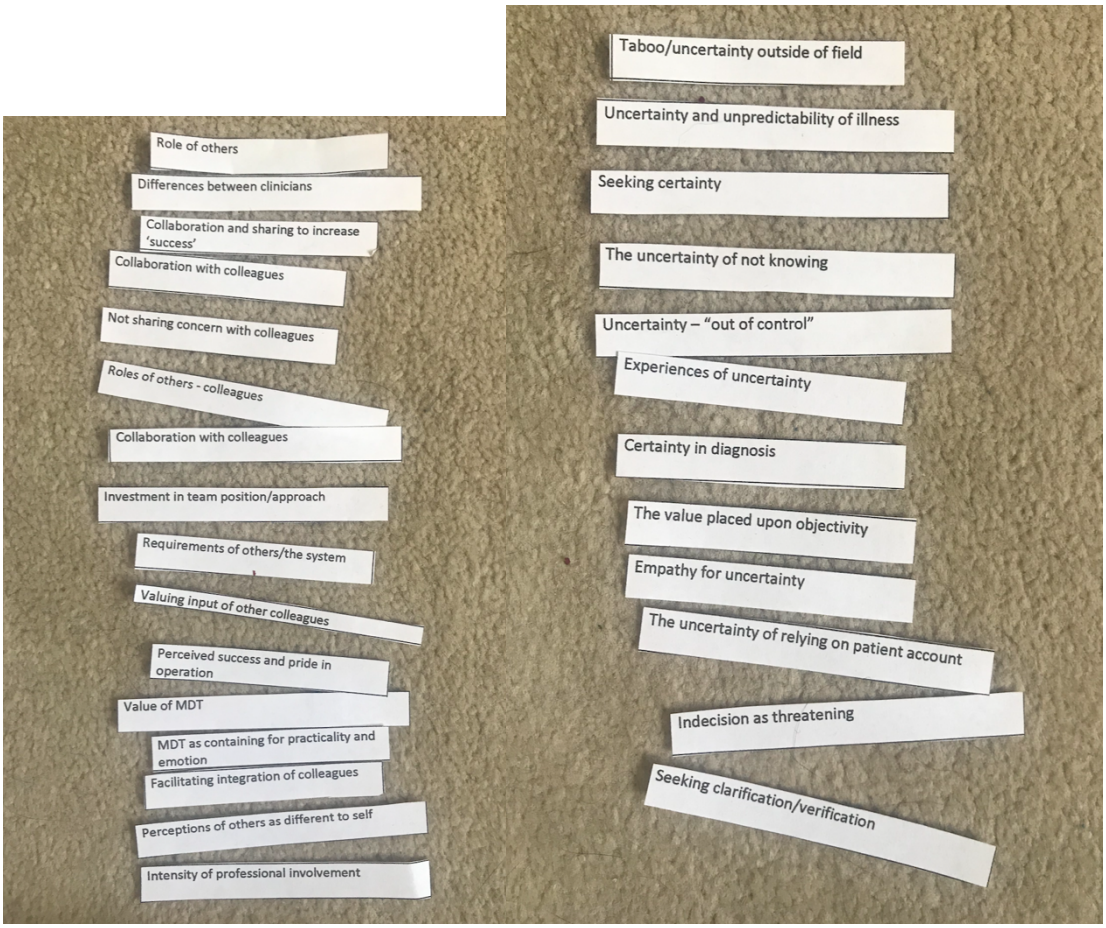
Version 1 – July 2018 IRAS ID: 263792

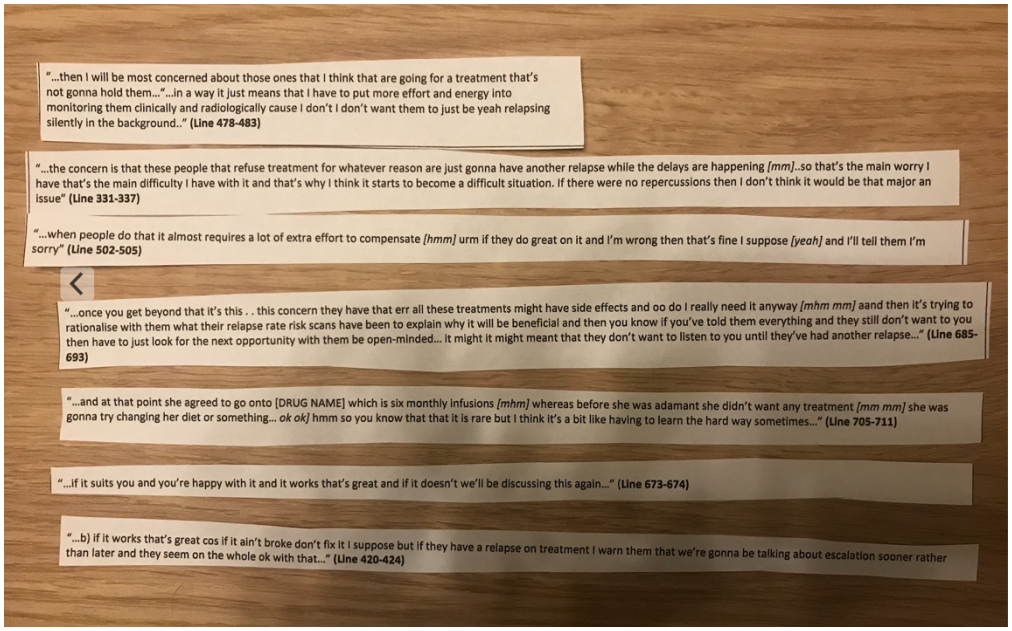
Appendix 12: Step 2 of analysis (Initial coding)



	103	I'd rather have this drug instead of that drug' I'd rather she
	104	goes onto something rather than nothing so I'm not gonna
	105	argue the toss too much unless it seems clinically
	106	inappropriate in which case we wouldn't've recommended it
	107	in the *[yep]* first place
	108	
I	109	Yeah yea you mentioned you went to see her with the-e
	110	specialist nurse *[yeah]* today w- is that quite typical for you
	111	to bring them in at that stage in those kind of consultations
	112	*[crosstalk]*
	113	
P	114	Yeah so I quite like to-o urm at some point during the initial
	115	admission introduce them to the nurses *[mm]* just so that
	116	they know who their first port of call is and so they've got an
	117	idea who they are *[uh-huh]* urm and it's a lot to take in isn't
	118	it so urm I think it's quite helpful for them to have the nursing
	119	team here visiting them whilst they're an inpatient cos they
	120	usually think of questions about their MS *[mm]* after I've left
	121	I only told her she had MS middle of last week
	122	
I	123	Ri-ight right..that's a lot to..process *[yep]* isn't it t-to *[yep]*
	124	kind of get through in the- an-and would you say that quite
	125	often your early meetings with clients are during admissions
	126	orr *[crosstalk]*?
	127	
P	128	No usually not *[no-]* so the vast majority of our new patients
	129	come in as outpatient referrals either from other neurologists
	130	here orr neurologists in other hospitals *[hmm]* or often urm
	131	MS nurses in other trusts *[mm]* it's probably less than ten
	132	percent of my new patients come through this route *[mm]*
	133	
I	134	Mm so it's quite a difficul- er a different sorry *[yea]* kind of
	135	*[yea]* consultation I guess to be having-
	136	
P	137	Yeaah and I think it does make it a bit more complicated cos
	138	they're usually in here with an acute relapse or a first attack
	139	so you're kind of mixing up the management of that acute
	140	episode along with trynna talk to them about disease
	141	modification treatment *[mm]* so I think it's good that a lot of
	142	the detail of treatment is discussed later at these discussion
	143	groups *[mmm]* cos there's too much to take in I think *[mm*
	144	*mm mm]* and they often will conflate steroid treatment for
	145	relapse versus disesase modification treatment *[yes]* for
	146	future relapse prevention
	147	
I	148	Yeah yeah it's something quite different about having a
	149	planned appointment with the *[yes]* neurologist that you're
	150	coming in to *[yea]* to see where you *[yeah]* maybe know
	151	*[yeah]* that's a possibility *[yeah] [crosstalk]*
	152	

Appendix 13: Step 3 of analysis (Emergent themes)

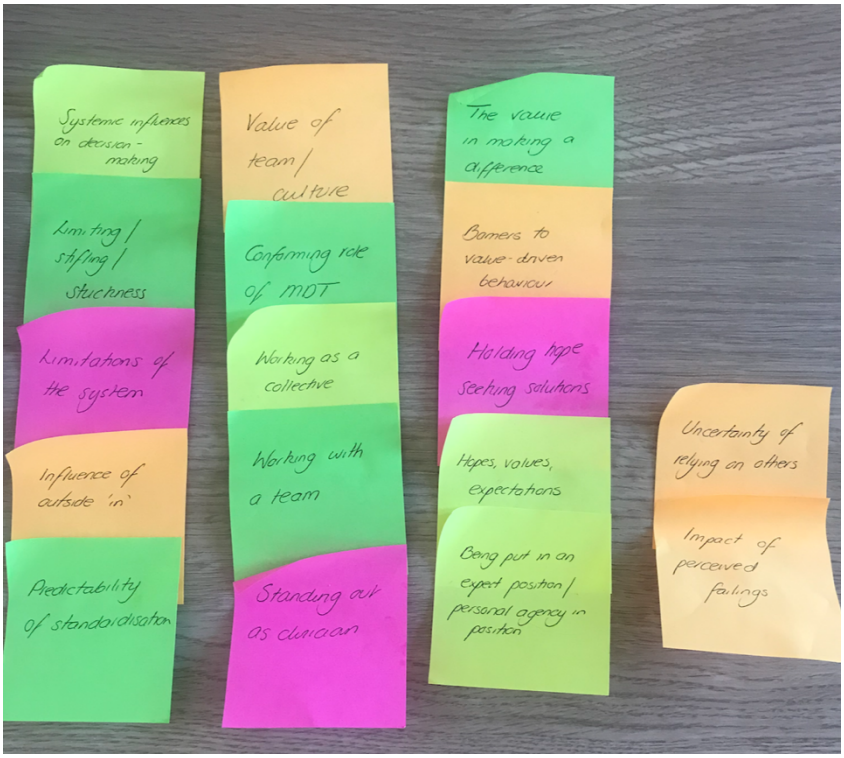


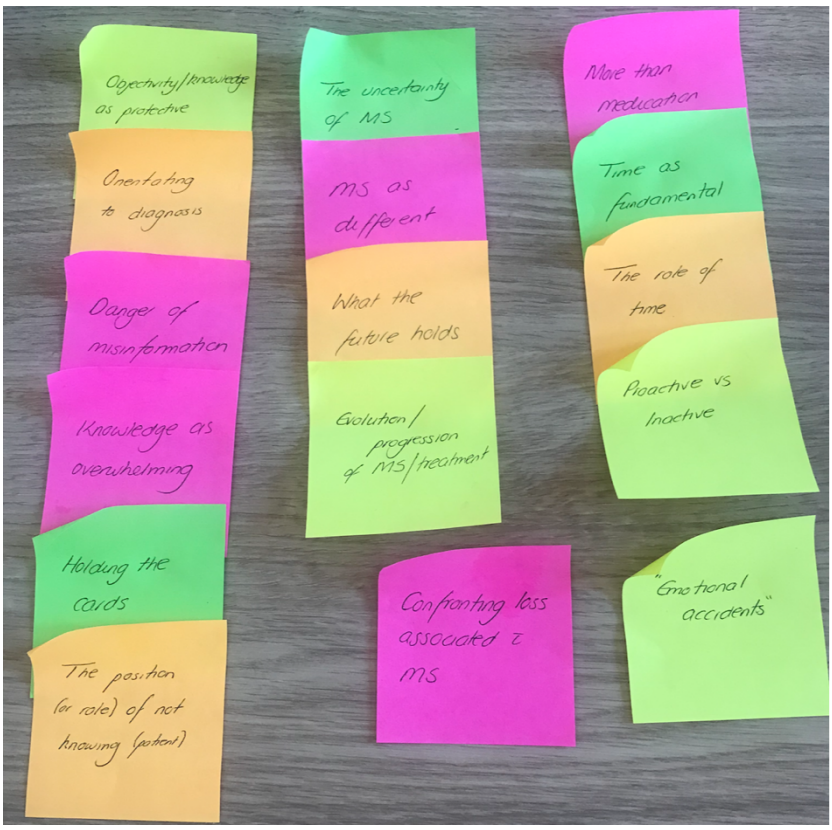


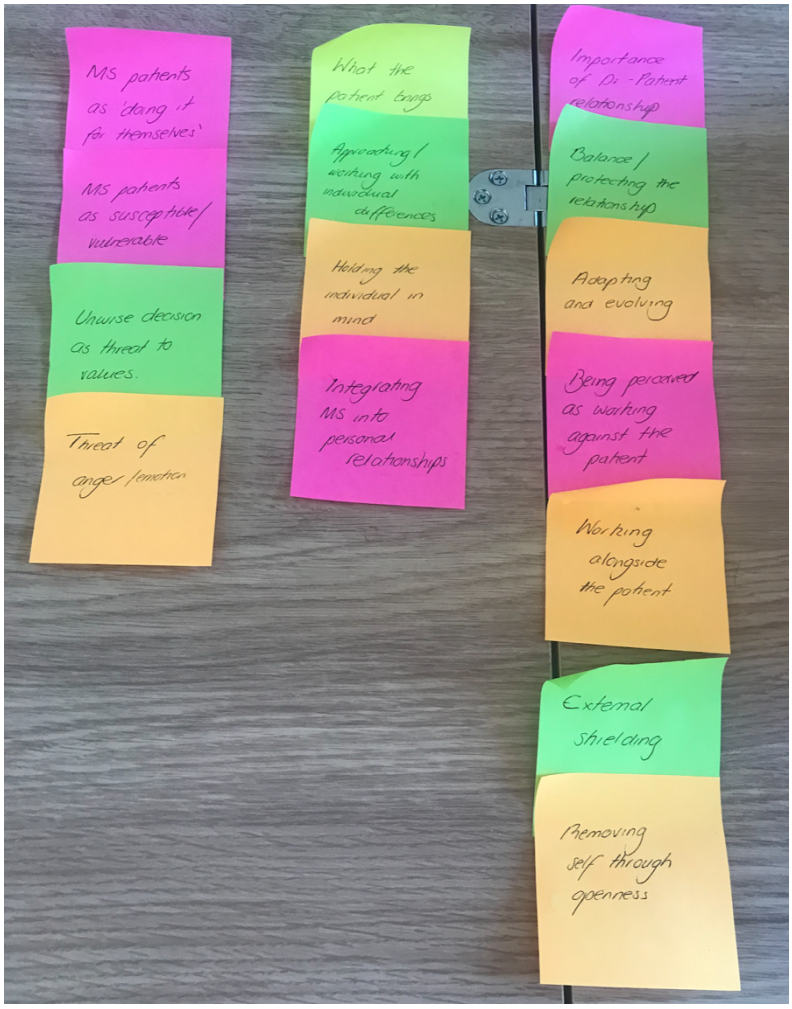
Appendix 14: Step 5 of analysis (Connections between participants)

NAVIGATING THE UNCERTAINTY	- The uncertainty of MS - Holding the cards - Individual differences
WORKING IN LINE WITH VALUES	- The value of making a difference - Time as fundamental - Balancing the therapeutic relationship - Threat of unwise decisions
THE CONFLICT OF CONTAINMENT	- Working with a team - Predictability of standardisation - Conforming role of MDT - Feeling stifled/stuck
NAVIGATING RELATIONSHIPS THROUGH UNCERTAINTY	- Integrating MS into personal relationships - MS patients 'doing it for themselves' - MS patients as vulnerable - Being perceived as working against the patient - Importance of the Dr-Patient relationship
HOLDING HOPE FOR LIFE	- Confronting the loss associated with MS - Holding hope - More than just medications
THE COMPLEXITY OF HOLDING THE KNOWLEDGE	- The danger of misinformation - Knowledge as overwhelming - Limitations of the system
DANCE WITH EMOTIONS	- Objectivity and knowledge as protective - Hopes, values and expectations - "Emotional accidents"
TRUSTING AND BEING TRUSTED	- Personal agency associated with position - Working as a collective - Systemic influences on decision making - What the patient brings
"IT'S A PROCESS..."	- Proactive vs. Inactive - Evolution and progression of MS
FEELING POWERLESS TO PROTECT	- Barriers to value-driven behaviour - Threat of emotions - Impact of others' perceived failings
NAVIGATING RELATIONSHIPS	- Adapting and evolving - Working alongside - Systemic barriers - Distancing self through openness
HOLDING THE UNCERTAINTY OF MS	- The position (or role) of not knowing - What does the future hold? - Orientating to diagnosis - The role of time - Holding the individual in mind
THE ROLE OF I IN TEAM	- The value of being part of a culture/team - The uncertainty of relying on others




Systemic influences on decision-making
Limiting / stifling / stuckness
Limitations of the system
Influence of outside 'in'
Predictability of standardisation
Value of team / culture
Confirming role of MDT
Working as a collective
Working with a team
Standing out as clinician
The value in making a difference
Barriers to value-driven behaviour
Holding hope Seeking solutions
Hopes, values, expectations
Being put in an expert position / personal agency in position
Uncertainty of relying on others
Impact of perceived failings


Objectivity/knowledge as protective
Orientating to diagnosis
Danger of misinformation
Knowledge as overwhelming
Holding the cards
The position (or role) of not knowing (patient)
The uncertainty of MS
MS as different
What the future holds
Evolution / progression of MS / treatment
Confronting loss associated w MS
More than medication
Time as fundamental
The role of time
Proactive vs Inactive
"Emotional accidents"


MS patients as 'doing it for themselves'
MS patients as susceptible/ vulnerable
Unwise decision as threat to values.
Threat of anger/emotion
What the patient brings
Appreciating/ working with individual differences
Holding the individual in mind
Integrating MS into personal relationships
Importance of Pt-Patient relationship
Balance/ protecting the relationship
Adapting and evolving
Being perceived as working against the patient
Working alongside the patient
External shielding
Removing self through openness