WHAT HELPS AND WHAT HINDERS IN COGNITIVE BEHAVIOURAL THERAPY INTERACTIONS?
THE DEVELOPMENT AND VALIDATION OF THE CONSULTATION INTERACTIONS CODING SCHEME.

SAMUEL MALINS, DClinPsy

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

April 2019

ABSTRACT

Background. Motivation is an important factor affecting outcomes of Cognitive Behavioural Therapy (CBT) for anxiety disorders. Recent research has defined motivation in terms of “patient activation”: the degree to which a person feels confident and able to actively manage their health. Patient activation has had widespread influence on healthcare, because it predicts a range of health outcomes. However, assessment of patient activation suffers from similar problems evident in other questionnaire-based measures of motivation: Patients’ often describe themselves as highly motivated even if they are not, for fear of negative consequences. Therefore, this thesis reports the development and validation of the Consultation Interactions Coding Scheme (CICS), which assesses patient activation from observations of turns-of-speech in CBT sessions.

Method. The CICS was developed in a total sample of 61 CBT recipients who experienced severe health anxiety alongside high rates of multimorbidity and high service use. Firstly, a systematic literature review addressed the question: Does observer-rated motivation predict clinical outcome in CBT for anxiety disorders? Next a mixed methods design was used to develop the CICS from a thematic analysis of CBT session interactions. Then the predictive validity of first and second session CICS scores was assessed with long- and short-term outcomes using multilevel modelling. Lastly, the relationship between short and long-term outcomes was explored with similar methods.

Results. The systematic review identified that observed motivation showed promise as a predictor of outcomes in CBT for anxiety disorders, but high risk of bias restricted generalisability. The CICS was developed to address the shortcomings identified. Good-to-excellent inter-rater reliability was achieved in all studies. Two naïve CICS raters were trained to use the measure reliably. The CICS achieved moderate-to-strong convergent validity with observer-rated therapeutic alliance and therapist competence. Specific interaction themes demonstrated predictive validity across outcomes, measurement methods and assessment time points. One consistent finding was that more time spent describing problems in early sessions predicted poorer outcomes both during treatment and up to 12-months later. Problem description interactions also significantly decreased directly prior to sudden outcome gains. Conversely, when patients were more engaged in discussion of positive changes they had noticed early in CBT, better outcomes ensued in physical and mental health both during therapy and several months afterwards.

Conclusion. A reliable measure has been developed, which has the potential for long- and short-term outcome prediction. The predictive validity is heavily weighted towards specific interaction types, which can be instructive for CBT therapists working with patients who have health anxiety and other comorbidities.

RELATED PAPERS AND PRESENTATIONS

[1]Malins, S., Kai, J., Atha, C., Avery, A., Guo, B., James, M. . . Morriss, R. (2016). Cognitive behaviour therapy for long-term frequent attenders in primary care: a feasibility case series and treatment development study. *British Journal of General Practice, 66*(651), e729-736.

[2]Malins, S. (2016). "Can we predict outcomes from session transcripts?" *Institute of Mental Health Research Conference.* University of Nottingham, 10th May.

Malins, S. (2017). "Intuition: The effective helper's wings or a dead weight?" *Motivational Interviewing Network of Trainers International Forum.* Dublin, 2nd – 6th October.

Malins, S. (2018). "Towards outcome feedback on psychological therapy interactions: Predicting outcome from first session talk." *British Psychological Society Annual Conference.* University of Nottingham, 2nd – 4th May.

Malins, S., Moghaddam, N., Morriss, R., Schroder, T., Cope, N. & Brown, P. (2018). Consultation Interaction Coding Scheme 1.6. Retrieved from https://figshare.com/articles/Consultation_Interaction_Coding_Scheme_CICS_/7302386

Malins, S., Moghaddam, N., Morriss, R., Schroder, T. (under review). Extending the Use of Routine Outcome Monitoring: Predicting Long-term Outcomes in Cognitive Behavioral Therapy for Severe Health Anxiety. *Psychotherapy Research.*

Morriss, R., Patel, S., Malins, S., Guo, B., Higton, F., James, M . . . Kaylor-Hughes, C. (2019). Clinical and economic outcomes of remotely delivered cognitive behaviour therapy versus treatment as usual for repeat unscheduled care users with severe health anxiety: a multicentre randomised controlled trial. *BMC medicine, 17*(1), 16.

Norwood, C., Moghaddam, N., Malins, S., & Sabin-Farrell, R. (2018). Working alliance and outcome effectiveness in videoconferencing psychotherapy: A systematic review and non-inferiority meta-analysis. *Clinical Psychology & Psychotherapy, 25*(6), 797-808.

Patel, S., Kai, J., Atha, C., Avery, A., Guo, B., Malins, S . . .Morriss, R. (2015). Clinical characteristics of persistent frequent attenders in primary care: case–control study. *BMC Family Practice, 32*(6), 624-630.

Patel, S., Malins, S., Guo, B., James, M., Kai, J., Kaylor-Hughes, C. . .Morriss, R. (2016). Protocol investigating the clinical outcomes and cost-effectiveness of cognitive–behavioural therapy delivered remotely for unscheduled care users with health anxiety: randomised controlled trial. *British Journal of Psychiatry Open, 2*(1), 81-87.

[1] Winner of the Nottinghamshire Healthcare NHS Trust Publication of the Year Award 2016

[2] Winner of the Institute of Mental Health Best Oral Presentation Award 2016

RELATED DATA

In line with recent guidelines emphasising the importance of transparency in psychological research (Miguel et al., 2014; Nosek et al., 2015), all the data used in this thesis have been made publically available for scrutiny, replication and collaboration. Each SPSS 24 dataset is accompanied with a related syntax file, which includes guidance notes and code to run the analysis carried out.

Chapters 4 and 5

https://doi.org/10.6084/m9.figshare.7306478.v3

Chapter 6

https://doi.org/10.6084/m9.figshare.7588511

Chapter 7

https://doi.org/10.6084/m9.figshare.7924337

Chapter 8

https://doi.org/10.6084/m9.figshare.7588550

ACKNOWLEDGEMENTS

If it takes a village to raise a child, it seems it has taken at least a medium-sized town for me to do a PhD. I can say with sincerity that this has been a joy (most of the time), when I know it could have been a horror or never even happened without the people below.

Professors Richard Morriss and Thomas Schröder. Miles Davis said “Sometimes you have to play a long time to be able to play like yourself.” I am grateful that you trusted me in the many turns of this project and that you gave me the freedom to start to sound like myself. It has been a privilege to learn from the breadth and length of your research knowledge and experience. The perspectives you brought to this project have strengthened it no end.

Dr Nima Moghaddam. I do not think either of us would have believed that a chat about interesting ideas would turn into a few chats about interesting questions that would eventually turn into hours of poring over therapy session interactions! Your quiet wisdom has shaped all the big decisions in this project and your friendship allayed my anxieties more times than I’d care to mention. Thank you and here’s to future chats about interesting ideas.

Paula Brown and Dr Naomi Boycott. Thank you for being patient with me! I’ve appreciated how many times you chose to think better of me than you probably ought to have done. Your contribution has been so significant.

I would like to thank the regular attenders study and urgent care study teams for their support throughout both projects. In particular, I am thankful to Shireen Patel. There would be no data without you! You have been a brilliant researcher to work with and a great friend too. Thanks for your honesty, feedback, willingness to go the extra mile and many laughs. Dr Boliang Guo,

thank you for sticking with me until I understood multilevel modelling and having a critical eye on the ways I have used it.

I am fortunate to have been mentored by Dr Jon Crossley. Thank you for helping me not just to complete a PhD but to see how it fits with what I value in life. You've helped me better understand how I might be a loving husband, father, son, brother, friend, colleague, supervisor and bandmate, alongside being a PhD student.

Dr Felicity Baker. You were well within your rights to just say "no" about my involvement in research. There was nothing in it for you except helping me. Thank you for persistently fighting my corner, I still owe you.

Dr Jo Tedstone. In the fullness of time you have become more than the service lead that you were to me since I qualified. You're a friend, inspiration and the best example of why I love being a clinical psychologist. Thank you for seeing qualities in me that I couldn't see in myself.

Dr Jo Levene and her Supreme Dream Team. I have been very lucky in life, but getting to work with and for you has been the best stroke of luck going for me. Thank you for constantly picking up the pieces from the mess I've left behind whilst doing this PhD. I have continually benefitted from your dolphin-like good will. Thank you for valuing this PhD as something important to me potentially beneficial for patients. It helped me focus on why I am doing this.

Mrs L Malins. We both know this is as much yours as it is mine! Thank you for making it possible for me to complete a PhD and barely compromise on anything that I value in life. You have just made it work, with grace, compassion, patience and most of all love whilst looking really pretty all the

while. You have been and are everything I could have hoped for in a wife and many things I didn't even know I should've hoped for.

Daughter no.1. I've really appreciated that you have been genuinely interested in the work that I do. In fact even if you were pretending you pretended really well! So thank you for keeping me interested ☺

Daughter no. 2 your spirit of fun helped me see the fun in this PhD. You kept me excited about it and how it could help people. Thank you ☺

Daughter no. 3. Thanks for not listening to Mummy when she said not to disturb me. Every time you came I needed a kiss and a cuddle anyway ☺

Mum, Percy, John and Dorothy. I have valued what I've learnt from you more than ever whilst writing this: That things are never as bad as they seem before you've slept on it; that it is almost always ok to laugh and be playful with difficulties; that hard work can feel good when you're doing it *and* when you've finished, plus many other lessons I have in my heart without even knowing it. Thank you.

Dad. There's a kind of poetry in me, the son of a man who cared deeply about patient relationships, writing a thesis about them. I imagine you would have found this interesting. I'd love to have had the opportunity to talk about it.

Shaws-in-law: You've been awesome! Thanks for everything you've done to keep the Malins boat afloat whilst I have been totally preoccupied.

Above all, my Heavenly Dad. As I thank these people I thank you for them and for every amazing breath of life you've afforded me. Time and chance comes to everyone, but you've given me more time and more chances than any right minded person would think I deserved. Thank you for never giving me what I deserve.

TABLE OF CONTENTS

ABSTRACT 2
RELATED PAPERS AND PRESENTATIONS 3
RELATED DATA 4
ACKNOWLEDGEMENTS 5
TABLE OF CONTENTS 8
LIST OF TABLES 11
LIST OF FIGURES 15
LIST OF APPENDICES 16
LIST OF ACRONYMS 17
LIST OF ACRONYMS RELATED TO THE CICS 19
1. BACKGROUND 20
1.1 Variability in psychological therapy outcomes 21
1.2 Therapist contribution to outcome 22
1.3 Difficulties explaining the benefits of psychological therapy 24
1.4 Limitations to process-outcome research 25
1.5 Observational process-outcome methods 28
1.6 Patient activation 39
1.7 The multimorbidity context 41
1.8 Routine Outcome Monitoring (ROM) 45
1.9 Summary 52
2. SYSTEMATIC REVIEW OF OBSERVER-RATED MOTIVATION 56
2.1 Introduction 56
2.2 Method 59
2.3 Results 65
2.4 Discussion 81
3. AIMS AND HYPOTHESES 89
3.1 Thesis plan 89
3.2 Aims 89
3.3 Objectives and hypotheses 91
4. DEVELOPMENT OF THE CICS: THEMATIC ANALYSIS 95
4.1 Introduction 95

4.2 Method 102
4.3 Results 116

5. DEVELOPMENT OF THE CICS: RELIABILITY AND VALIDITY 128

5.1 Introduction 128
5.2 Method 134
5.3 Results 144
5.4 Summary 153

6. THE PREDICTIVE VALIDITY OF THE CICS 158

6.1 Introduction 158
6.2 Method 168
6.3 Results 189
6.4 Summary ..... **Error! Bookmark not defined.**

7. THE CICS' SENSITIVITY TO PROXIMAL OUTCOMES 214

7.1 Introduction 214
7.2 Study 1: Using early CICS scores to predict session-by-session outcomes 215
7.3 Study 2: Changes in patient activation prior to sudden gains 236
7.4 Overall summary 252

8. ESTABLISHING THE VALUE OF PROXIMAL OUTCOMES 256

8.1 Introduction 256
8.2 Unaddressed issues in ROM 257
8.3 Comorbidity and ROM 258
8.4 Health anxiety and comorbidities 258
8.5 Sudden gains and ROM 259
8.6 Method 260
8.7 Results 269
8.8 Summary ..... **Error! Bookmark not defined.**

9. DISCUSSION 291

9.1 Introduction 291
9.2 Summary of findings 291
9.3 Relationship to existing literature 295
9.4 Strengths 302
9.5 Limitations 305

9.6 Implications for future research 310
9.7 Implications for clinical practice 314
9.8 Conclusion 316

10. REFERENCES 318
11. APPENDICES..... 343

LIST OF TABLES

Table 1.1: Sample items from each stage identified on the PAM 40

Table 2.1. Review inclusion and exclusion criteria 62

Table 2.2. Summary of included studies' risk of bias and applicability 71

Table 2.3. Summary of studies investigating both positive and negative elements of observed motivation that were applicable to the review..... 78

Table 2.4. Summary of studies investigating negative elements of observed motivation that were applicable to the review..... 79

Table 2.5. Summary of studies investigating negative elements of observed motivation, but not applicable to the review..... 80

Table 4.1. Summary of perspectives contributing to the qualitative research approach (adapted from Parker, 2005) 97

Table 4.2. Participant primary care consultation rates and health problems. 106

Table 4.3 Phases of Thematic Analysis (V. Braun & Clarke, 2006) and quality assurance methods used at each stage (adapted from Elliott et al., 1999; Nowell et al., 2017)..... 110

Table 4.4. Example audit trail for the Structuring and Task Focus theme development..... 112

Table 4.5. Summary of the overall theme template 113

Table 4.6. CICS coding scores and equivalent, mapped patient activation levels (adapted from Deeny, Thorlby & Steventon, 2018) 120

Table 4.7. Problem analysis and understanding coding level descriptors and exemplar quotes..... 121

Table 4.8: Exemplar quotes with corresponding CICS themes and level coded. 122

Table 5.1. Intra-class coefficients for transformed scores and turn-by-turn ratings. 145

Table 5.2. Spearman's rho correlation between therapeutic alliance, therapist competence and CICS themes. 147

Table 5.3. Spearman's rho correlations between early session structuring score and change in outcome score at six month follow-up. 149

Table 6.1: Essentials elements of cognitive behaviour therapy for health anxiety (adapted from P. Tyrer, Cooper, Tyrer, et al., 2011) 163

Table 6.2. Patient activation level descriptors and exemplars for the Noticing Change or Otherwise (NCO) CICS theme..... 179

Table 6.3. Steps in the development of the MLM model assessing first session structuring interactions as a predictor of health anxiety severity. 187

Table 6.4. Demographics and clinical characteristics of included and excluded participants..... 191

Table 6.5. Consultation Interactions Coding Scheme IRR..... 192

Table 6.6. Observer-rated working alliance inventory IRR 192

Table 6.7. Session 1 CICS and WAI-O correlations with 12-month follow-up outcomes..... 194

Table 6.8. Session 2 CICS and WAI-O correlations with 12-month follow-up outcomes..... 195

Table 6.9. Multilevel model of noticing change interactions at session two predicting all health outcomes across time..... 199

Table 6.10. Multilevel model of problem and context description interactions at session two predicting follow-up general health and generalised anxiety. 206

Table 6.11. Multilevel model of facilitative and general interactions at session two predicting follow-up depression and generalised anxiety..... 208

Table 6.12. Multilevel model for structuring and task focused interactions at session one as a predictor of health anxiety 209

Table 7.1. Steps in the development of the MLM model assessing the predictive ability of session 1 self-evaluative interactions as a predictor of session-by-session ORS-rated wellbeing..... 223

Table 7.2. Multilevel model where session 2 interactions about noticed change (or the absence of change) predict session-by-session wellbeing (ORS). 225

Table 7.3. Multilevel model where the proportion of session 1 interactions taken up with problem description predicts session-by-session wellbeing (ORS) 225

Table 7.4. Multilevel model where the proportion of session 2 interactions taken up with problem description predicts session-by-session wellbeing (ORS) 226

Table 7.5. Multilevel model where evaluations of self or therapy at session 1 predict session-by-session wellbeing (ORS) 228

Table 7.6. Multilevel model where evaluations of self or therapy at session 2 predict session-by-session wellbeing (ORS) 229

Table 7.7. Multilevel model where structuring and task focus interactions at session 2 predict session-by-session wellbeing (ORS) 229

Table 7.8. Reduction in unexplained outcome variance between patients over time by CICS theme 232

Table 7.9. Reduction in unexplained rate of change variance by CICS theme 232

Table 7.10. Patient activation level descriptors and exemplars for the Information Discussion CICS theme..... 243

Table 7.11. Sudden and non-sudden gainer sample characteristics. 245

Table 8.1. Illustration of MLM development for last session ORS predicting health anxiety over time. 265

Table 8.2. Demographics and clinical characteristics of included and excluded participants 270

Table 8.3. Pearson’s correlations between baseline outcome scores and wellbeing at the first treatment session. 272

Table 8.4. Spearman’s correlations between 6-month follow-up outcome scores and wellbeing at the last treatment session. 273

Table 8.5. Predictive validity of patient wellbeing at the last treatment session on follow-up outcomes over 3, 6, 9 and 12-month follow-ups, controlling for baseline severity and changes over time. 275

Table 8.6. Pre-post treatment effect sizes and predictive effect sizes for post-treatment wellbeing and health anxiety. 280

LIST OF FIGURES

Figure 1.1. Theory of change in MI. 34

Figure 1.2. Percentage of patients not-on-track for therapeutic improvement at point of feedback who eventually met criteria for clinically significant change at termination (adapted from Lambert, 2015) 46

Figure 1.3: Feedback clinical support tools decision tree (adapted from Lambert, 2015) 49

Figure 2.2. Risk of bias across studies 72

Figure 2.3. Model applicability across studies 73

Figure 4.1. The coding process. 115

Figure 4.2. Process model from thematic analysis. 118

Figure 6.1. The vicious cycle of reassurance seeking. 161

Figure 6.2: Morriss et al.'s (2019) CONSORT diagram: Participant flow into the RCT 175

Figure 6.3. Participant Flowchart. 176

Figure 6.4. Pictorial illustration of the hierarchical structure of data in the MLM analysis 184

Figure 6.5. The print version of the ORS 220

Figure 7.1. Comparative scores for CICS interaction themes and the WAI-O at critical sessions (N) and control sessions (N – 1) 247

Figure 7.2. Comparative scores for CICS interaction themes and the WAI-O at critical sessions for sudden gainers and control sessions for matched non-sudden gainers 248

Figure 8.1. The average sudden gain. The averaged ORS scores are shown for the sessions around the pre-gain session (N). 282

Figure 8.2. Percentage greater outcome improvement achieved by treatment responders with sudden gains versus non-sudden gaining treatment responders. 283

LIST OF APPENDICES

APPENDIX A: Systematic review search strategy 344

APPENDIX B: Example completed PROBAST form 348

APPENDIX C: Developmental phase consent form 355

APPENDIX D: Theme development examples 356

APPENDIX E: Examples of coder reflections 359

APPENDIX F: Consultation Interactions Coding Scheme manual 361

APPENDIX G: Included baseline assessments for chapters 6, 7 and 8 394

APPENDIX H: Participant information sheet for chapters 6, 7 and 8 403

APPENDIX I: CBT information sheet 408

APPENDIX J: Patient-participant consent form 412

APPENDIX K: Cause pie chart example 413

APPENDIX L: Transcriber confidentiality agreement 414

APPENDIX M: Notes from coder discussion 415

APPENDIX N: Chapter 6 bootstrapped CICS MLM results 417

APPENDIX O: Chapter 8 bootstrapped ORS MLM results 421

LIST OF ACRONYMS

AAQ II	Acceptance and Action Questionnaire II
CBT	Cognitive Behavioural Therapy
CCT	Counter-Change Talk
CHARMS	Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies
CICS	Consultation Interactions Coding Scheme
CT	Change Talk
CTS-R	Cognitive Therapy rating Scale - Revised
DSM-5	Diagnostic and Statistical Manual of mental disorders - 5th Edition
DSM-IV	Diagnostic and Statistical Manual of mental disorders - 4th Edition
EQ5D	Euroqol - 5 dimensions
FIS	Facilitative Interpersonal Skills
GAD-7	Generalised Anxiety Disorder scale - 7 items
IAPT	Improving Access to Psychological Therapies
ICC	Intra-class correlation coefficient
IPA	Interpretative Phenomenological Analysis
IRR	Inter-rater reliability
MI	Motivational Interviewing
MI-CBT	Motivational Interviewing as a precursor to Cognitive Behavioural Therapy
MLM	Multilevel Modelling
NHS	National Health Service
ORS	Outcome Rating Scale
PAM	Patient Activation Measure

PHQ-15	Patient Health Questionnaire - 15 items
PHQ-9	Patient Health Questionnaire - 9 items
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST	Prediction model Risk Of Bias Assessment Tool
PSWQ	Penn State Worry Questionnaire
QUIPS	Quality in Prognosis Studies
RCT	Randomised Controlled Trial
ROM	Routine Outcome Monitoring
RRS	Ruminative Responses Scale
SCID	Structured Clinical Interview for the Diagnostic and statistical manual of mental disorders IV
SF-36-MCS	Short-Form 36 items - Mental Component Scale
SF-36-PCS	Short-Form 36 items - Physical Component Scale
SFBT	Solution Focused Brief Therapy
SHAI	Short form Health Anxiety Inventory
TAU	Treatment As Usual
VAS	Visual Analogue Scale
WAI-O	Working Alliance Inventory - Observer rated
WSAS	Work and Social Adjustment Scale

LIST OF ACRONYMS RELATED TO THE CICS

API	Action Planning and Idea generation
EST	Evaluations of Self or Therapy
IDI	Information Discussion
NCO	Noticing Change or Otherwise
PAU	Problem analysis and understanding
PCD	Problem or Context Description
STF	Structuring and Task Focus
OTH	Other interaction type
IOT	Integration of Themes
COF	Collaborative Flow

1. BACKGROUND

Despite strong evidence of effectiveness, psychological therapies are unlikely to have met their full potential to improve health (Lambert, 2013). One means of addressing this situation is to develop and disseminate replicable interventions for specific problems using experimental evidence. Another way to improve overall effectiveness is to focus on improving the outcomes achieved by individual psychological therapists themselves. The latter is perhaps more likely to impact on practicing psychological therapists, given that mixed problems are more typically seen in routine practice and overall improvements may be applicable across problem-types (Goldberg, Babins-Wagner, et al., 2016). Several methods have been developed recently to achieve this goal, most of which either directly or indirectly aim to increase therapists' responsiveness to specific patient[3] presentations (Stiles, Honos-Webb, & Surko, 1998). In order to achieve this type of responsiveness greater attention to interaction-style and type is required. Furthermore, there is a growing understanding that practice of relevant therapeutic skills is required to achieve better outcomes (Goodyear & Rousmaniere, 2017). This chapter reviews the quality of tools to help therapists improve their effectiveness. The focus of this thesis is to design a tool that addresses shortcomings identified.

[3]The terms "patient", "client" and "service-user" have been applied interchangeably for people attending psychological therapy. There is debate over preferred terms and whether stigma is attached to certain terms more than others (McLaughlin, 2009). Although there is little evidence that any terms is strongly preferable to others, a recent meta-analysis indicated that the term "patient" was narrowly preferred over "client" in the UK ("service-user" was actively disliked), so "patient" is used throughout this thesis (Dickens & Picchioni, 2012).

1.1 Variability in psychological therapy outcomes

A number of psychological therapies have been evidenced as effective for a range of health difficulties by comparison to no treatment, waiting list control or non-specific supportive therapy control (e.g. Lambert, 2013; Lipsey & Wilson, 1993). This includes the treatment of mental health problems, such as depression and anxiety disorders (Wampold & Imel, 2015), but also includes treatment of long-term physical health problems such as chronic pain (Williams, Eccleston, & Morley, 2012). Cognitive Behavioural Therapy (CBT) has been shown to be as effective or more effective than comparator psychological therapies in the majority of trials where random allocation, direct comparisons took place (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). When researcher allegiance is accounted for the superiority of CBT reduced but remained to a small extent and the clinical value of any superiority has been hotly debated (Luborsky et al., 1999; Luborsky et al., 2002).

Despite the overall effectiveness of evidence-based psychological therapies, variability in outcomes remains a central concern. The majority of people completing a course of psychological therapy report recovery or reliable improvement in mental wellbeing, but 5-10% report deterioration in spite of therapy (Lambert, 2010). Approximately 20% of patients attending psychological therapy also dropout prematurely (Swift & Greenberg, 2012); treatment ending in this way is much less likely to lead to improvement (Clark et al., 2009). Furthermore, improvement and recovery rates achieved in trials are often not replicated in clinical practice (e.g. N. B. Hansen, Lambert, & Forman, 2002). In summary, psychological therapies are not effective for a

large proportion of those who seek them, and much research has identified ways in which the effectiveness of psychological therapies could be improved.

1.2 Therapist contribution to outcome

Part of the reason for the limits to overall effectiveness in psychological therapies is the potential for wide variations in effectiveness between individual therapists. Although studies of therapist effects indicate that the majority of therapists are broadly similar in effectiveness, the bottom 15-20% of therapists are consistently and distinctively less effective than the majority; conversely, the top 15-20% of therapists are consistently and characteristically more effective than others (Barkham, Lutz, Lambert, & Saxon, 2017). A key example of this effect, at a service level, is found within the UK's national Improving Access to Psychological Therapies (IAPT) program. It involves more than a million referrals and several thousands of psychological therapists annually. Rates of recovery from both anxiety and depression symptoms varied from 85% at the most effective service - far exceeding recovery rates typically achieved in clinical trials - to 7% - lower than spontaneous recovery rates (NHS Digital, 2018). The extent of variability between therapists on an individual level has been demonstrated in studies where the most and least effective therapists are compared. For example, Okiishi, Lambert, Nielsen, and Ogles (2003) found that the most efficient and effective therapists had an improvement rate ten times the average. This meant that the most effective therapists could have successfully treated ten patients in the time it took the average therapist to treat one. Conversely, patients of the least effective therapists showed an average *deterioration* in functioning. This illustrates that differences in therapist effectiveness are not just a clinical issue but a service

delivery and cost effectiveness issue. Importantly, the differences between therapists become more pronounced with patients who have more severe or complex difficulties (Johns, Barkham, Kellett, & Saxon, 2018). Therefore, therapist skill becomes more important to overall effectiveness where patients have complex or severe difficulties. In practice, this is likely to include patients with multiple comorbid problems or multimorbidity.

Even when the best available controls are imposed upon therapist variability – such as identical training, standardised therapy protocols, monitoring for adherence and competence, routine supervision of practice, seeing patients with the same diagnosis and no comorbidities – differences between therapists play an important role in outcome variance (Crits-Christoph & Gallop, 2006; Krupnick et al., 1996). Typically, therapist variability accounts for 5-9% of outcome variance, but the effect can range from 5-55% (Baldwin & Imel, 2013; Barkham et al., 2017).

However, few studies used to establish therapist effects were specifically designed for this purpose and several were under-powered to do so according to established guidelines (Schiefele et al., 2017). Because of this less planned approach, few variables that may be important moderators of therapist effects have been assessed in relevant studies. A further problem caused by the more opportunistic study design of therapist effects research (as opposed to a priori designed) is that some types of therapy, outcome, patient and therapist are over-represented, whilst others are almost entirely missing. In particular, studies focused on psychological treatments for comorbid physical and mental health problems are almost entirely missing from the evidence base (with two notable exceptions in chronic fatigue: Goldsmith,

Dunn, Bentall, Lewis, & Wearden, 2015; Wiborg, Knoop, Wensing, & Bleijenberg, 2012). There has also been a paucity of research incorporating quality of life and general health assessment (as opposed to mental health assessment) within therapist effects (again with a notable exception in one study: Goldsmith et al., 2015). Notwithstanding this critique, therapist effects are consistently present across study, population, therapy and therapist types. Overall, this indicates that a significant difference in therapist effectiveness is important to overall effectiveness of psychological therapies.

It is noteworthy that therapist differences are not typically explained by therapeutic orientation, sex, training type or differences in competence (Branson, Shafran, & Myles, 2015; Okiishi et al., 2003; Webb, DeRubeis, & Barber, 2010). This is important because these factors have sometimes been identified as the cause of therapist differences (e.g. H. Tyrer et al., 2015), when the preponderance of research suggests the causes lie elsewhere. Some key areas have been identified as potential causes of therapist difference and methods have been developed to address this issue. However, these methods are dependent upon an understanding of what causes change in psychological therapy.

1.3 Difficulties explaining the benefits of psychological therapy

Despite the breadth of successful applications for psychological therapies, controversies remain about the factors causing the benefits experienced by the majority who receive treatment. Part of the reason for this debate is the discrepancy between the purported causes of change presented by a given psychotherapeutic theory and the small role this is often found to play in outcome variance from large-scale meta-analytic studies (e.g.

Wampold & Imel, 2015). When trials of different therapies and disorders are grouped in meta-analyses, that do not differentiate on the basis of diagnosis, it is often found that only a small proportion of outcome variance is explained by the experimentally and theoretically derived causes proposed in psychological therapy models (S. Braun, Gregor, & Tran, 2013; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Lambert, 2013). Factors such as the therapeutic relationship are often identified as causing at least as much of the outcome variance, with both components accounting for the minority of outcome variance (Flückiger, Del Re, Wampold, & Horvath, 2018; Norcross & Lambert, 2018).

On the other side of the coin, a range of largely cognitive and behavioural theorists argue that the meta-analyses on which these hypotheses are based are too inclusive (DeRubeis, Brotman, & Gibbons, 2006). They incorporate multiple study designs and studies of varying quality, which is seen to inevitably lead to more general, prosaic findings. Furthermore, many studies of specific psychological therapies for identified disorders have evidenced theoretically important contributions made by distinct (typically cognitive) techniques (e.g. A. T. Beck & Dozois, 2011). It is also common for there to be differences found in the effectiveness of given psychological therapies for different disorders, which undermines the idea that the effectiveness can be grouped meaningfully across diagnoses (Cuijpers et al., 2013).

1.4 Limitations to process-outcome research

Research on both sides of this debate has often failed to address a number of key complications in assessing the cause and effect relationship in psychological therapy. Crucially, almost all studies have relied heavily on self-

report questionnaires for assessment of both process and outcome variables. This means that a patient typically rates their therapeutic relationship on a series of Likert scaled questions and then rate their psychological distress using a similar method. Such methods are shown to inflate correlations between the two measures (Campbell & Fiske, 1959). Process-outcome associations are further inflated by having the same rater assessing both the process and the outcome. For example, correlations between therapeutic alliance and outcome were significantly inflated (r = .30) when the patient rated both alliance and outcome compared to when the therapist or an observer rated either assessment (Horvath, Del Re, Flückiger, & Symonds, 2011). This means that self-reported ratings of alliance may not be reliably independent of outcome measures when the same person completes both, as has been the case in the majority of studies.

Any and all self-reported questionnaires have reliability and validity problems in and of themselves, by the nature of the method. There are long-standing and well-documented issues of what are called “demand characteristics”. This means that those completing self-report measures often report what they feel might be more socially desirable or expected by those administering the measure (Stone, Bachrach, Jobe, Kurtzman, & Cain, 1999). There are also examples of respondents guessing at the hypotheses of research behind self-report measures and responding in a way that confirms or disconfirms these predicted hypotheses (Garcia & Gustavson, 1997). There are even examples where participants appear to have rated themselves as more distressed in an attempt to receive study-delivered psychological therapy (Heppner, Wampold, & Kivlighan, 2008).

The dynamic changes that take place within and between sessions of psychological therapy may mean it is unsuitable for self-reported assessments in some instances. Within-session changes that take place in psychological therapy can have a significant impact on outcome, but may not be picked up by assessments that rate the session as a whole (Safran, Muran, & Eubanks-Carter, 2011). Therapist responsiveness is an important factor in this area. Therapist and patient responsiveness refers to behavioural changes presented or affected by a given therapy context (Stiles et al., 1998). Therapist responsiveness can confound psychological therapy processes measured by self-report questionnaire. For example, a therapist may increase efforts to form a stronger alliance if they feel therapy is going badly or alternatively relax these efforts if it seems to be going well. Therefore, the presumption that therapeutic alliance is relatively stable over time does not necessarily hold, nor can it be seen as unidirectional.

Lastly, attempts to segment individual contributions to outcome from specified elements of therapy (such as separating therapist and patient influences) has served to under-estimate the interactive, inter-dependent nature of psychological therapy. For example, when verbal confrontations were given by therapists, patient resistance increased and when empathic reflections were used resistance decreased (Moyers, Miller, & Hendrickson, 2005). This suggests that individual patients are responsive to individual therapist interactions and vice-versa, which can be shaped to change the tone of the conversation. Taken together, it is difficult to maintain that each therapeutic factor is separate from the others, with clear evidence of interdependence. In order to gain a genuine understanding of these factors

and preserve their relationship to others, a naturalistic observation of combined therapy interactions would seem most fitting.

To address the criticisms of self-report-based process-outcome research and the importance of early interactions, much recent work has focused on the use of observational measures to identify important therapist and patient processes contributing to outcome.

1.5 Observational process-outcome methods

1.5.1 Therapist interpersonal skills

Observer-rated assessment of therapist interpersonal skills have repeatedly been shown as an important factor in therapy outcome. The Facilitative Interpersonal Skills (FIS) performance task has received the most research in this area. The FIS performance task uses standardised video segments of sessions with challenging patients to assess a range of therapist interactional skills and responsiveness (Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009). These include verbal fluency, emotional expression, persuasiveness, hopefulness, warmth, empathy, alliance-bond capacity and problem focus – all identified from existing clinical research (Norcross & Wampold, 2011). Therapists respond to two segments from four problematic therapy process types: an angry patient, a passive patient, a confused patient and a controlling patient. Therapist responses to each segment are then recorded and rated by trained observers.

Initial evidence showed that FIS performance predicted therapist level variability in outcome where demographic factors (e.g. age, therapeutic orientation) and self-reported social skills did not, when placed in the same predictive model (Anderson et al., 2009). This suggests that the most effective

therapists had proficient facilitative interpersonal skills, particularly in challenging therapeutic situations compared to those achieving poorer outcomes. It also supports the use of observer-rated assessment over self-report given that self-reported social skills were not an effective predictor but observed skills were. A more likely causal relationship was established in an RCT where non-therapists' possessing a high level of FIS achieved better psychological therapy outcomes with depressed patients than those with low FIS. Interestingly, there was no difference between clinical psychology doctoral students (i.e. those who had been trained in specific therapy techniques) and students of non-helping courses (e.g. biology or history – those who had received no such training) once FIS was accounted for (Anderson, Crowley, Himawan, Holmberg, & Uhlin, 2015).

An important nuance to understanding the impact of FIS was added by a study of clinical psychology trainees. Anderson, McClintock, Himawan, Song, and Patterson (2016) strengthened the evidence for therapist FIS as a predictor of outcome, by administering FIS assessments more than a year prior to trainees seeing the patients with whom outcomes were assessed. It also replicated the finding that self-reported inter-personal skills did not predict outcome. The study added that trainees with higher FIS had a particularly pronounced superiority in patient outcome over those with lower FIS in therapy of shorter duration. This suggests that interactional skills are particularly important in very early sessions – often the period of therapy where the majority of improvements take place (Lambert, 2010).

Further evidence for observer-rated therapist interpersonal skills was added by Schöttke, Flückiger, Goldberg, Eversmann, and Lange (2017), in a

longitudinal study using Therapist-Related Interpersonal Behaviours. The group assessment of Therapist-Related Interpersonal Behaviours used a provocative video to initiate a group discussion among rated therapists. Trained observers rated positive communication; empathy and attunement; respect and warmth; management of criticism, and willingness to co-operate from therapist contributions to the discussion. A second assessment method used experts to conduct structured interviews rating therapists' personal-strengths and therapists' motivation for interest in patients, self-reflection and specified therapeutic approaches. Results indicated that observer-rated Therapist-Related Interpersonal Behaviours predicted therapist-level variance in outcomes over five years of patient outcomes. However, the expert-led interview did not. Both models controlled for therapist gender, theoretical orientation and amount of supervision alongside patient order in therapist caseload, patient age, gender, comorbidities, problem severity and personality disorder diagnosis. These findings underline the value of interpersonal responsiveness and skill in therapist effectiveness over the long-term. It once again demonstrates the value of observer-rated assessment of therapist performance over direct, self-reported questions.

Overall, these studies suggest that relational and interactive skills are likely to make up an important part of the abilities possessed by the most effective therapists. The evidence also suggests that observer-rated interaction-style appears to be a fruitful source of data for predicting psychological therapy outcome. In particular, it was shown to be a better predictor than self-reported assessments in direct comparisons. This ideology has been the particular focus of Motivational Interviewing (MI): A psychological

therapy focused on increasing the likelihood of specified changes by minimising resistance and maximising a patient's intrinsic reason's for change (W. R. Miller & Rollnick, 2012).

1.5.2 Motivational Interviewing assessments

Motivational Interviewing has focused on observations of interpersonal skills and interaction-level, in-session ratings of motivation (Moyers, Martin, Manuel, Hendrickson, & Miller, 2005; Moyers, Miller, et al., 2005). From this research two key findings have emerged that are relevant to the current review: 1) The amount and strength of observer-rated patient motivation expressed in MI sessions is a replicable predictor of outcome (e.g. Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003) 2) Observer-rated therapist utterances predict patient motivation (e.g. Moyers, Martin, Houck, Christopher, & Tonigan, 2009). A body of research has focused on therapist utterances that are more likely to improve patient motivation and reduce resistance to therapist direction. These include open (as opposed to closed) questions, affirmations, reflections and summaries (Moyers, Martin, et al., 2005). Recently global ratings of therapists' MI skills have been added, which include: evocation and elicitation of patient's reasons for change; collaboration (as opposed to didactic maintenance of the expert role); emphasising patient autonomy, and maintenance of focus during discussion (Moyers, Rowell, Manuel, Ernst, & Houck, 2016). Overall, MI-consistent therapist behaviours aim to use less questioning and directing, greater use of reflections, summaries and affirmations to promote empathy, attunement, empowerment and autonomy within a focused conversation about change. Central to MI guidance on

therapist behaviours has been the importance and necessity of managing and reducing patient resistance in order to improve therapeutic outcomes.

Patient resistance can be viewed as any behaviour that opposes, obstructs, blocks or diverts therapists' direction (Chamberlain, Patterson, Reid, Kavanagh, & Forgatch, 1984). Several studies have used the Client Resistance Code to rate in-session patient resistance behaviours. For example, patient resistance at session 1 of CBT for generalised anxiety accounted for 23-32% of the variance of worry one-year post-treatment (Westra, 2011). This was particularly impressive given that resistance interactions took up less than 5% of in-session interactions. Again, self-reported measures were not significant predictors in models where observer-rated assessment was included. Similar findings have been replicated in several studies of different psychological therapies and patient populations (Ahmed, Westra, & Constantino, 2012; Aviram & Westra, 2011; Button, Westra, Hara, & Aviram, 2015; Safran et al., 2011; Westra, Aviram, Connors, Kertes, & Ahmed, 2012).

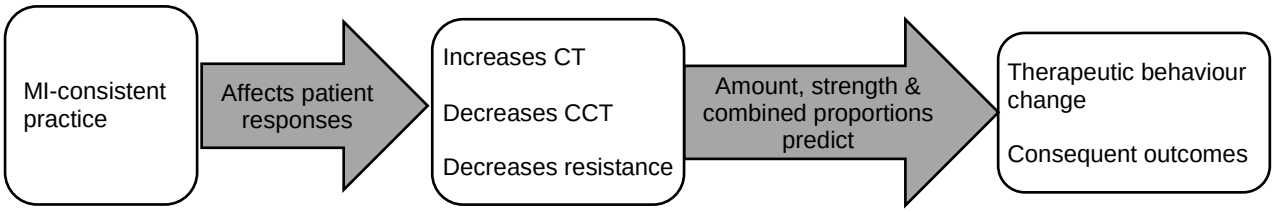
Importantly, interactions between patient resistance and moderating therapist responses have been identified. Beutler, Harwood, Michelson, Song, and Holman (2011) drew together evidence identifying that patients showing more resistance or reactance had poorer outcomes when paired with more directive therapists, but better outcomes when paired with non-directive therapists. Conversely, less resistant patients had better outcomes with more directive therapists. This highlights that factors identified as "belonging" to either patient or therapist can only be fully understood when the interaction between both parties is accounted for. The principle of matching therapist

responses to patient presentations has been applied to observer-rated resistance in a recent RCT (Westra, Constantino, & Antony, 2016). The RCT compared CBT for generalised anxiety with CBT synthesised with MI (MI-CBT) – using responsive therapist methods to manage patient resistance. Results indicated that patients in the MI-CBT group showed greater reduction in worry at one-year post-treatment follow-up than patients receiving CBT alone, despite receiving the same number of sessions. This suggests that the infusion of responsive MI-based management of resistance helped to reduce the negative impact that high resistance typically has on outcome. The specific importance of combining therapist responses with certain patient presentations at particular times and in particular contexts was underlined by detailed analysis of interactions where patients presented resistance. Aviram, Westra, Constantino, and Antony (2016) showed that patients whose therapists showed MI-consistent responses at times of disagreement had considerably better outcomes and subsequent resistance later in therapy. However, MI-consistent therapist responses at non-disagreement points had no effect on outcome. This underscores the point that it is the combination of important patient-related process markers and therapist responses that help to determine what differentiates the most and least effective therapists.

Motivational Interviewing posits that MI-consistent behaviours lead to increased patient motivation, which can be observed through an increase in “Change Talk” (CT): verbal expression of the desire, reasons, ability or commitments to make therapeutic changes (W. R. Miller & Rose, 2009). Specific MI-consistent behaviours are also hypothesised to reduce “counter-change” or “sustain talk” (CCT): patient utterances expressing ambivalence,

reluctance or rejection of the focal change. Lastly, MI theory proposes that increasing CT and reducing CCT increases the likelihood of therapeutic behaviour change, thus improving outcome (Figure 1.1). A series of recent meta-analyses have investigated the specific interaction-level, observer-rated processes deemed to cause change in MI, which shed further light on the interplay between therapist behaviours, patient behaviours and consequent outcome.

Figure 1.1. Theory of change in MI.



Three meta-analyses have partially validated this hypothesis; aggregation of studies assessing the relationships described has established that MI-consistent behaviour *does* increase CT, but it also increases CCT too. Behaviours inconsistent with MI (e.g. confronting, arguing, directing) were correlated with more CCT and not CT. When tested independently, CT did not relate to therapeutic behaviour change at follow-up (e.g. reduced alcohol intake), but CCT did have a strong negative association as predicted (Magill, Apodaca, et al., 2018). However, when proportions of CT and CCT were combined there was a composite predictive value to both. In a detailed analysis of CT and CCT subtypes, it was found that one CT subtype (“other” change talk) was actually associated with poorer outcomes, whilst the proportion of reason-based CT was correlated with improved outcome (Magill, Bernstein, et al., 2018). Counter-change talk was again more reliably and

generally associated with outcome independent of CT. A third review showed similar findings and confirmed that a composite score of CCT and CT was the best way of accounting for the role CT has on outcome – on its own CT was not associated with outcome (Magill et al., 2014). Significant to the current review, all significant findings relied on observer-rated interaction level analysis and none of the globally rated assessments (evocation, collaboration, empathy and focus – rated for the whole session) were associated with outcome or process variables. This suggests that detailed observations of specific patient-therapist exchanges may have a benefit over global, whole-session single ratings.

Taken together, research on resistance, CT, CCT and MI-consistent therapist behaviours suggests that:

1. Observational measures may be useful in addressing the shortcomings of self-report-based prediction of outcome.
2. Specific pairing of more helpful and less helpful patient-therapist interactions is possible through synthesis of patient and therapist behaviours using observational analysis.
3. Detailed interaction-level analysis may have advantages over global session-level assessment when it comes to motivational factors in psychological therapy.
4. Findings from this type of research can give very clear and practical guidance for therapists on how to improve their approach, in a way that global assessments cannot. For example, baseline problem severity is often found to be an important predictor of outcome. Although there are ways in which this can be managed differently, it is not as clear and

practical as guidance on in-session behaviours that are best suited to specific patient presentations.

There are also important criticisms of this body of research:

1. Data is largely derived from MI in drug and alcohol treatment, where early engagement is an identified issue. In addition, deterioration and drop out is almost double that of all psychological therapies (Lilienfeld, 2007; McHugh et al., 2013). Therefore, findings may not be wholly generalisable to all psychological therapies. That said, a body of work has been developing within anxiety disorders specifically (e.g. Aviram et al., 2016).
2. Although the measures used achieved adequate inter-rater reliability (IRR) for total numbers of therapist and patient behaviours (e.g. Intra-class correlation coefficient [ICC] = .89 for number of affirmations in a session; Moyers et al., 2016), none have established whether coders can differentiate behaviours at specific points in a session. This means that observers can agree on the overall amount of specific therapist and patient behaviours, but it is unclear whether they agree at the same particular point in sessions (Moyers, Martin, Catley, Harris, & Ahluwalia, 2003). This is an important limitation, because the matching of certain patient and therapist behaviours is central to the overall value of observer-rated motivational factors in the prediction of outcome.
3. Current evidence does not explain the apparent differential value of negative motivational factors as compared to positive factors. Specifically, resistance and CCT appear to be better predictors of outcome than CT, but it is unclear why or how to respond to this

situation. Furthermore, MI-consistent behaviours, such as fostering empathy, elicit more CCT and resistance as well as CT – contrary to what has been expected. There are no clear explanations for this or guidance as to how therapists might refine their practice to address the situation.

1.5.3 Observational assessments from CBT

Observer-rated assessments of in-session behaviours have also been used to understand causal change processes in CBT, most extensively in cognitive therapy for depression (Feeley, DeRubeis, & Gelfand, 1999; Strunk, Brotman, & DeRubeis, 2010; Strunk, Brotman, DeRubeis, & Hollon, 2010; Tang, DeRubeis, Beberman, & Pham, 2005; Webb et al., 2012). Associated studies have focused on how specific CBT therapist skills relate to outcome and how they interact with cognitive changes and CBT skills in patients. Interestingly, similar patterns emerged from an assessment of CBT therapist behaviours, patient responses and their impact on outcome (Strunk, Brotman, & DeRubeis, 2010). Two CBT adherence behaviours predicted outcome: concrete cognitive methods and negotiating/structuring session content. Patient facilitation or inhibition of therapist adherence also predicted subsequent change in depressive symptoms. Patient facilitation and inhibition of therapist adherence could be seen as close relatives of CT and resistance. Therapist negotiation behaviours could be seen as related to MI-consistent behaviours, such as collaboration.

Again, close attention has been paid to the temporal relations of patient-therapist interactions and outcome. For example, Tang et al. (2005) identified that there was a large increase in cognitive changes that were explicitly stated

in sessions by patients (e.g. reporting a change in belief or change in understanding of a belief) directly prior to large stable improvements in depressive symptoms. This again could be seen as a “reason” subtype of CT.

1.5.4 Observer-rated measures: Summary

This section has shown how observer-rated process measures are important in establishing the cause of the significant differences between individual psychological therapists. Observational methods like those reviewed may be more reliable and instructive for therapists than self-report measures. There are potential convergences across therapeutic modalities focused on responsive matching of therapist interactions to patient presentations in order to improve outcome. Much research has focused on different motivational factors, as expressed in observable psychological therapy interactions. However, there is a mismatch in the research reviewed, wherein the importance of matching particular therapist behaviours to important patient presentations has been emphasised. Yet, none of the observer-rated tools used have established reliability at the level of turn-of-speech, only at session or session-segment level. This means that current methods cannot reliably infer that identified and specified behaviours have been matched or otherwise. Furthermore, the developments described have occurred predominantly within mental health and addictions research when similar developments have been emerging in behavioural medicine more generally. In particular, the concept of patient activation has emerged as an important predictor of many health outcomes within behavioural medicine (Hibbard, Stockard, Mahoney, & Tusler, 2004), but the concept has been largely ignored in psychological therapies.

1.6 Patient activation

Patient activation is a significant and relatively new concept in health, which has become a well-used and researched approach to improving self-management (Hibbard & Greene, 2013; Hibbard, Mahoney, Stock, & Tusler, 2007). Patient activation is defined as the degree to which a person feels confident and able to be actively involved in managing their own health (Hibbard, Mahoney, Stockard, & Tusler, 2005). The Patient Activation Measure (PAM: see Table 1.1 for sample items; Hibbard et al., 2004) has been used to assess patient activation across multiple settings. Several studies have replicated the finding that higher patient activation levels predict improvement in a range of health outcomes and lower health service use, even when controlling for key socioeconomic factors (Barker, Steventon, Williamson, & Deeny, 2018; Greene, Hibbard, Sacks, Overton, & Parrotta, 2015; Hibbard & Gilburt, 2014). Furthermore, intervention studies suggest that patient activation can be seen as a health outcome in itself, because important health indicators improve when patient activation is targeted and increased (Greene & Hibbard, 2012). The PAM is used as an assessment and outcome measure to facilitate tailored, patient-centred care on the basis of patient activation level. Use is widespread, with a large-scale national pilot recently completed in the UK National Health Service (NHS England, 2018).

Table 1.1: Sample items from each stage identified on the PAM

Patient Activation Stage 1: Believes Active Role Important	Taking an active role in my own health care is the most important factor in determining my health and ability to function
Patient Activation Stage 2: Confidence and Knowledge to Take Action	I am confident that I can take actions that will prevent or minimize some symptoms or problems associated with my health condition.
Patient Activation Stage 3: Taking Action	I am confident I can figure out solutions when new situations or problems arise with my health condition.
Patient Activation Stage 4: Staying the Course Under Stress	I am confident that I can maintain lifestyle changes like diet and exercise even during times of stress

However, the PAM is the only assessment of patient activation. It measures retrospective, self-reported attitudes, knowledge and behaviour. This method of assessment comes with difficulties, which were highlighted in an implementation analysis of PAM-use from the UK NHS pilot (Armstrong et al., 2017). Several patients and clinicians noted biases in responses dependent on patients' perceptions of the PAM's purpose (e.g. responding more positively if patients believed access to care would otherwise be restricted). Some clinicians felt more benefit could be gained through an unstructured consultation, because the wording of the questions could come across as antagonising. This led to mediated PAM completion by healthcare professionals, bringing the validity of the data into question.

Perhaps the most important critique of the PAM was the difficulty clinicians experienced in translating PAM data into relevant changes to consultation behaviours. This meant that, due to their general nature, PAM scores were difficult to apply to individual consultations with specific patients, whose needs may be complex. This illustrated a clear disconnection between the PAM assessment method and interventions developed to improve patient

activation: Many intervention studies designed to improve self-management and increase patient activation focus on changing *interaction-style* within healthcare consultations (e.g. training patients to ask helpful questions during healthcare consultations; Deen, Lu, Rothstein, Santana, & Gold, 2011). However, the PAM does not directly assess interaction behaviours in healthcare consultation or give feedback of sufficient detail to inform these consultations, despite the evident importance of interaction-style to improving patient activation.

Patient activation is identified as a particularly important concept, in relation to health outcome and healthcare use, for patients who have multiple comorbid problems. Addressing multimorbidity of this kind is a key challenge for healthcare in general and is identified as a priority issue for UK NHS care in the coming years (NHS England, 2019). Furthermore, supporting patients to manage their own care will be a central strategy to achieving this aim and the PAM is the most likely means of measuring this development (NHS England, 2018).

1.7 The multimorbidity context

The effects of having multiple health conditions can be profound for both the patients in question and the health services treating them. Simply having multimorbidity contributes to reduced quality of life, increased risk of death from several causes and difficulties with daily activities (Kanesarajah, Waller, Whitty, & Mishra, 2018; Sharabiani, Aylin, & Bottle, 2012). Furthermore, the knock on effects of multimorbidity to working life augment the scale of the problem (Taskforce on Multiple Conditions, 2018). For example, reduced ability to work reduces earning potential when those with

multimorbidity are more likely to have a lower socioeconomic status at the outset. A recent Health Foundation briefing paper (Stafford et al., 2018) highlighted that the majority of NHS primary and secondary care costs are spent on those with multimorbidity, with three quarters of the medication budget spent on people with multiple chronic conditions. Those with four or more conditions had treble the number of outpatient and primary care appointments with nearly quadruple the number of medications (Stafford et al., 2018). This is a growing issue for health services to manage. For example, 1 in 10 emergency hospital admissions were for those with 5 or more conditions in 2006/7, whereas it was 1 in 3 in 2015/16 (Steventon, Deeny, Friebel, Gardner, & Thorlby, 2018). The cost of multimorbidity to the NHS is expected to increase by £4 billion in the next five years (Charlesworth, 2018).

Central to the problems surrounding multimorbidity is that healthcare services have historically struggled to effectively treat people with multiple problems (Guthrie, Payne, Alderson, McMurdo, & Mercer, 2012). This is partly because treatments have typically focused on individual conditions in isolation, rather than the combined effect of multiple conditions and multiple treatments (Mercer, Gunn, Bower, Wyke, & Guthrie, 2012). Therefore, interventions that minimise side-effects and are generally effective across conditions have been difficult to select and apply, leaving wide variations in the care provided (Barnett et al., 2012). Secondly, the majority of treatment and care for chronic conditions is undertaken by the patients themselves or their carers (Deeny, Thorlby, & Steventon, 2018). This has meant that helping people gain motivation and confidence to improve self-management is central to future healthcare strategy (NHS England, 2018, 2019). Thirdly, and perhaps most

importantly, multimorbidity that involves comorbid physical and mental health problems are likely to be the single biggest driver of health and social care costs use (Kasteridis et al., 2015; Naylor et al., 2012). Furthermore, the ability of patients in this group to self-manage their care is highly affected by clinician responsiveness and interaction-style (Derksen, Bensing, & Lagro-Janssen, 2013; Mercer et al., 2016). In addition, patient activation plays a more significant role in the prognosis of those with comorbid physical and mental health problems (Barker et al., 2018). Overall, this suggests harnessing interventions that improve self-management would play an important role in addressing multimorbidity and the interaction-style of clinicians will be important to this end.

As discussed, there is a clear overlap between multimorbidity and high service use. Long-term high service utilisers have an average of three long-term conditions, high rates of anxiety related to health and patterns of service use that indicate ongoing increased consultation rates with little amelioration of their difficulties (Patel et al., 2015). Clinician empathy and communication style have been identified as important routes for interventions with these types of complex need (Derksen et al., 2013). Furthermore, psychological therapies have been evidenced for a number of disorders associated with multimorbidity and long-term high service use. In particular, CBT for health anxiety has aimed to address the combined impact of physical and mental health problems alongside high service use (Salkovskis, Warwick, & Deale, 2003). For example, persistent bodily checking is an identified factor maintaining health anxiety, wherein fear of bodily symptoms (such as pain) leads to recurrent body checking. However, repeatedly checking the body exacerbates

symptoms such as pain; thus pain is worsened and anxiety is consequently maintained (Salkovskis & Warwick, 1986). This pattern often leads to increased healthcare consultation rates (Salkovskis, 1989). Therefore, treatment that leads to improvements in health anxiety could be expected to have concurrent improvement in physical health and reduction in healthcare use. However, engagement with mental health treatment is particularly problematic within the multimorbid patient population (Naylor et al., 2016). For example, patients with multiple difficulties may not wish to have their problems framed as having a psychological component and ineffectual communication from healthcare professionals can exacerbate this resistance (Ring, Dowrick, Humphris, Davies, & Salmon, 2005; Salmon, Humphris, Ring, Davies, & Dowrick, 2006). Therefore, early session interaction and communication skills are especially important.

In summary, this section has highlighted that patients with multimorbidity are likely to receive more variable care and have poorer outcomes. However, interventions that focus on improving motivation and confidence in self-management are likely to be of great value to patients and health services. Increasing clinician responsiveness and patient motivational factors (particularly patient activation) are identified as key cross-cutting treatment targets to enhance self-management and improve outcomes. Indeed specific psychological therapies maybe important for people with comorbid physical and mental health problems. When considered in the context of the earlier discussion about variability in therapist effectiveness, it is likely that the complexities presented by patients experiencing multimorbidity would emphasise differences between therapists in effectiveness. Several studies

have suggested that the more severe and complex a patient's difficulty the more valuable therapist skill becomes (DeRubeis et al., 2005; Hollon & Beck, 2013; Jacobson & Hollon, 1996). Therefore, a system for assessing the way that therapist-patient interactions link to outcome would be best assessed in a context where therapist skills are most tested. In the following section current well-evidenced methods for improving therapist effectiveness are discussed.

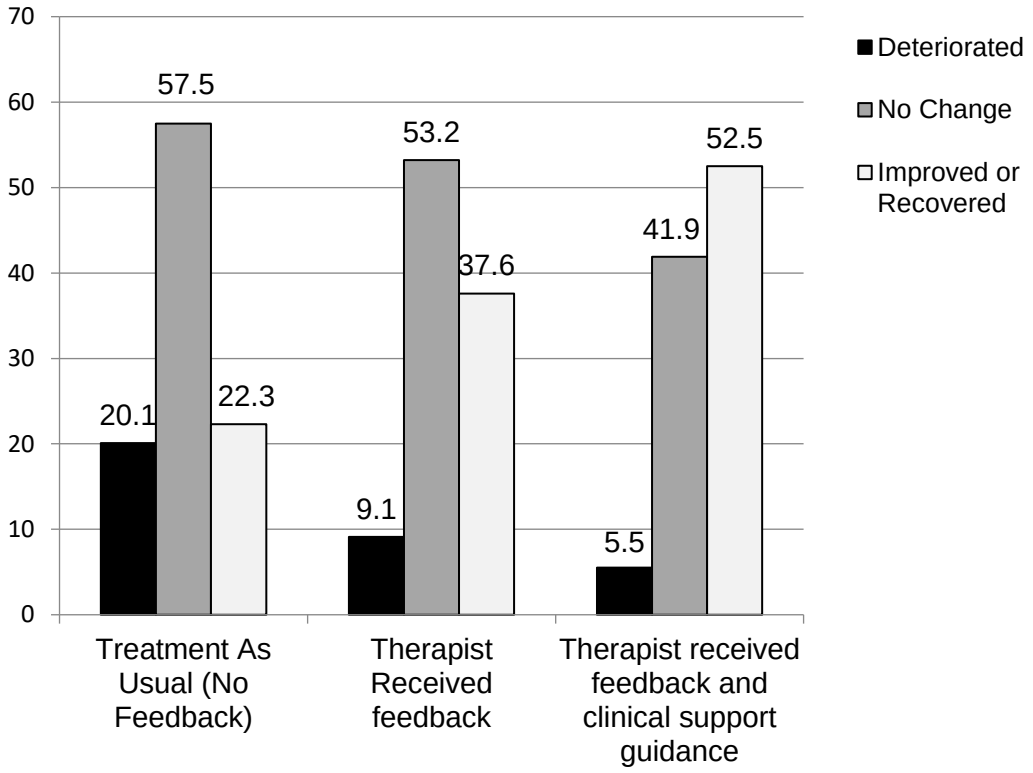
1.8 Routine Outcome Monitoring (ROM)

To date, one of the most effective means of reducing dropout and therapy failure as a whole, has been through the provision of standardised feedback on predicted outcome (Duncan & Reese, 2015; Lambert, 2010; Lambert, Whipple, & Kleinstäuber, 2018). Shimokawa, Lambert, and Smart (2010) showed that the proportion of patients improving and recovering increased stepwise as therapists were offered algorithm derived predictions of outcome, then increased to a greater extent when feedback was accompanied by training on how to use feedback to best effect. Both were superior as compared to those in therapy where their therapist was not offered feedback. Moreover, the proportion of patients experiencing no change or deterioration fell stepwise from those receiving no feedback to those receiving feedback and again for those receiving feedback, training and guidance (Figure 1.2). Importantly, the benefits were most pronounced for those who had experienced little change or had deteriorated at the point feedback was given. This indicates that provision of feedback can change the trajectory of those likely to have a poor outcome.

Part of the rationale for the effectiveness of ROM is that psychological therapists (similar to other healthcare professionals) present as overconfident

about their individual abilities, underestimate the likelihood of patient deterioration and over-estimate the likelihood of patient improvement (Hannan et al., 2005). For example, 91.6% of surveyed psychological therapists identified themselves as above the 75th centile of performance as compared to their peers (Walfish, McAlister, O'Donnell, & Lambert, 2012). Where such flawed self-assessments exist, standardised feedback can ameliorate the situation; as occurs in ROM (Dunning, Heath, & Suls, 2004).

Figure 1.2. Percentage of patients not-on-track for therapeutic improvement at point of feedback who eventually met criteria for clinically significant change at termination (adapted from Lambert, 2015)



However, the feedback presented and the guidance offered is necessarily general. For example, if a “Readiness to Change Intervention” is indicated from feedback, there may be great variability in how to “discuss readiness to change with the patient” (Figure 1.3). In this context, the specifics

of interaction type and style are particularly important, given that interpersonal style cuts across specified interventions (Anderson et al., 2015; Anderson et al., 2009). The exclusive use of self-report questionnaires in current ROM is also a concern, because it does not provide detailed enough data for ongoing practice improvement (Goldberg, Rousmaniere, et al., 2016).

1.8.1 ROM in the context of therapist development

The general feedback available through ROM may not help therapists develop their skills in the way more detailed feedback on interaction-style might. In a large longitudinal study of 170 psychological therapists who had treated 6,591 patients, Goldberg, Rousmaniere, et al. (2016) found that therapists outcomes got slightly worse over time even where ROM was being implemented – greater experience did not lead to better outcomes. In contrast, a case study centre demonstrated average improvement in outcomes with time and experience where deliberate practice was used to enhance detailed in-session skills (Goldberg, Babins-Wagner, et al., 2016). These studies call for more detailed feedback on ways of relating that can be linked to outcome, which could be developed through a focus on interaction-style.

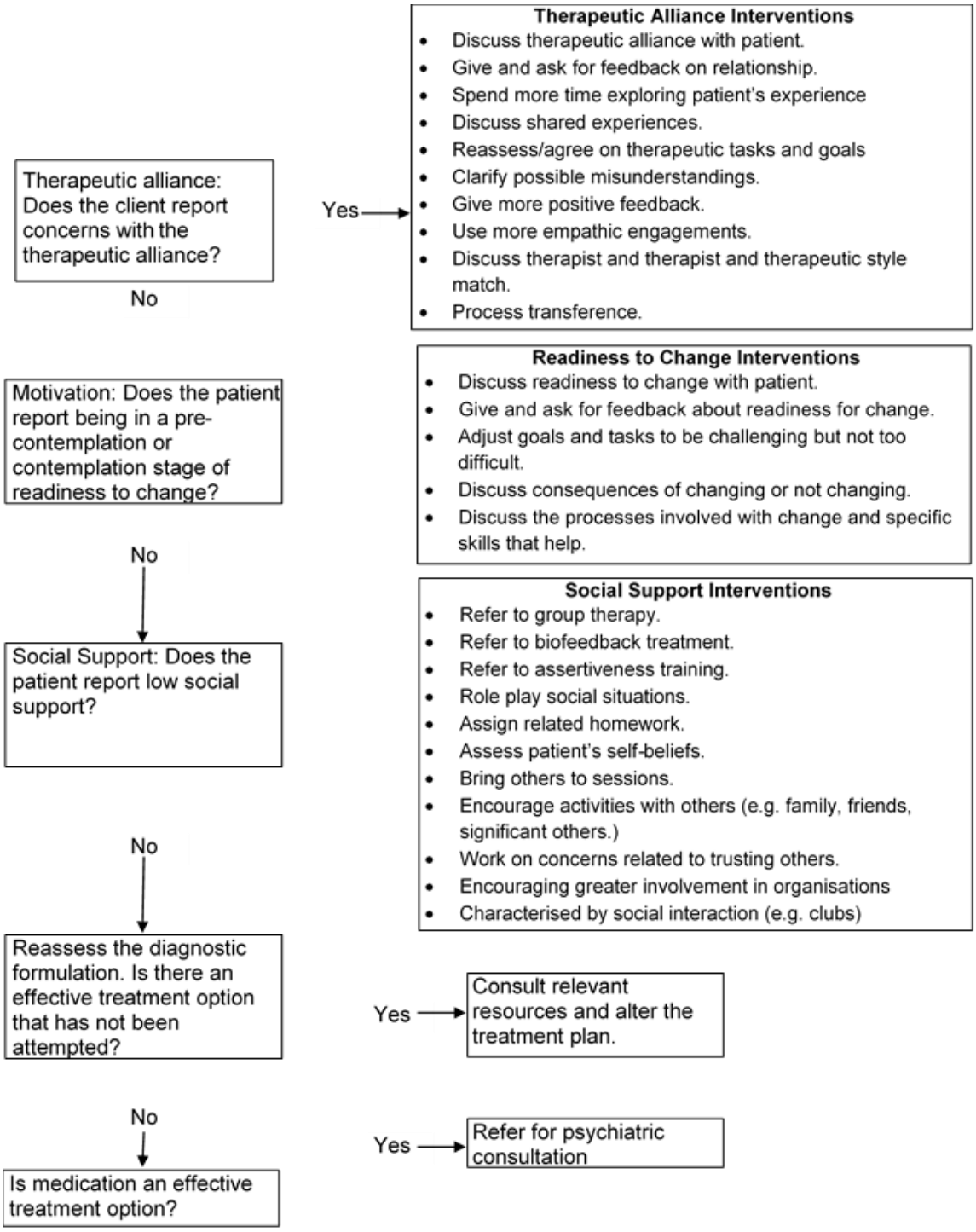
1.8.2 ROM in the context of decision-making

The type of feedback offered in ROM (Figure 1.3) also assumes that therapist in-session decision-making is primarily strategic and logical when this is unlikely to be the case. The preponderance of research on decision-making suggests that most decisions are built upon associative reasoning, long-standing habits of previous decisions and will be biased in numerous ways including a skew towards simpler methods or methods used before (Kahneman, 2011). This applies to interaction-styles as they are based on

long-standing, over-learned skills. Given this context it would be helpful if intuitive thinking could be trained to make interaction decisions that are the most therapeutic. This type of intuition may already be at work in the most effective therapists, as it is for outstanding members of other professions (e.g. expert intensive care nurses; Crandall & Getchell-Reiter, 1993).

Expert intuition is defined as highly complex pattern recognition: A cue in the environment prompts an expert's access to information that is helpful in the given situation, from their memory of related situations (Simon, 1992). Kahneman and Klein (2009) outlined the agreed conditions for the development of intuitive expertise. This included having a "high-validity environment" (p. 524), where there is sufficient regularity for intuitive expertise to be developed. That is, an environment where cues for specific knowledge are to some degree related to events relevant to the expertise. In an interaction based example of a high validity environment, there would need to be a reliable relationship between a patient *stating* that they commit to change a particular behaviour and the behaviour *actually* changing in order to develop intuitions on interaction-styles that could make change more likely; this type of relationship has been demonstrated (Amrhein et al., 2003). Other conditions necessary for intuitive expertise acquisition are: rapid and accurate feedback on performance and adequate practice (Kahneman & Klein, 2009). It is also identified as acceptable for there to be significant uncertainty as long as there remains some regularity to cues and their consequences (as in poker).

Figure 1.3: Feedback clinical support tools decision tree (adapted from Lambert, 2015)



Given these conditions, interaction feedback could help develop intuitive expertise in the type of facilitative interactional style which appears to contribute to the effectiveness of the best therapists and enable therapists to improve outcomes over time. It is noteworthy that psychological therapists will already be using intuitive thinking to inform their decisions (whether they know it or not) and are likely to be over-confident in their performance on this front. As such, designing and applying an interaction feedback system of this kind could not only improve outcomes but prevent the type of positive skew in self-assessments to which all therapists are liable (Dunning et al., 2004).

If specific patient-therapist interactions were predictive of outcome (as has been demonstrated: Magill, Apodaca, et al., 2018) then offering therapists feedback on interactions that are associated with improvement, given certain patient presentations, could have similar effects to ROM. This type of feedback would have the added benefit of giving more specific feedback on session contents and how to remediate problems. With this level of in-session cues, stronger intuitions and associated adaptations may be developed for improved outcomes. This type of feedback also does not rely on solely self-report measures with their inherent drawbacks.

1.8.3 An existing example of interaction feedback

The Cognitive Therapy Rating Scale – Revised (CTS-R; Blackburn et al., 2001) rates CBT competence on 12 six-point Likert scales, assessing key components of CBT. It has used in-session interaction types to assess and improve the competence of CBT therapists in training using feedback (Barnfield, Mathieson, & Beaumont, 2007; Milne, Baker, Blackburn, James, & Reichelt, 1999). There is evidence of an association between therapist

competence and patient outcome in some studies of depression and anxiety disorders (Ginzburg et al., 2012; Strunk, Brotman, DeRubeis, et al., 2010). However, the most recent meta-analysis suggests this competence-outcome association is very weak (Webb et al., 2010). Furthermore, a recent naturalistic study of CBT therapists trained in the UK IAPT programme found no association between therapist competency and outcome even though competency improved through training (Branson et al., 2015).

One potential reason for this lack of association may be that the CTS-R is primarily theory driven as opposed to outcome driven. For example, agenda setting (item 1) and homework setting (item 12) may have differential associations with outcome and could therefore be weighted differently in scoring, to ensure therapists who are more likely to get better outcomes get higher scores. This type of weighting or outcome focus has not be carried out. So the differential importance attributable to different types of therapeutic interaction have been missed. Another explanation for the lack of association is that “meta-competencies” held by exceptional therapists may cut across therapy disciplines and identified components (Heinonen, Lindfors, Laaksonen, & Knekt, 2012; Okiishi et al., 2006). As such, designing an interaction assessment tool more focused on what leads to improved outcome may provide the best resource for therapist training. This is particularly important when tools in current use do not have strong evidence that they are leading to the accreditation of therapists who can achieve outcomes any better than unqualified therapists.

1.9 Summary

Establishing factors that contribute to enhanced outcome in psychological therapy is a complex and difficult process, with which researchers have had some success and a number of failings. Most research has relied on self-reports rather than observational methods and few have preserved the naturalistic context where all factors are, in reality, interdependent. Moreover, much process-outcome research does not account for the way therapists and patients are responsive to the actions of the other; the therapeutic process is more dynamic than is likely to be picked up by between-session questionnaires.

Feedback informed practice through ROM has been an area of success. Feedback has led to reduced therapy failure of all kinds and greater therapeutic improvements. This has impacted training, practice and overall therapy guidance in a meaningful way (Boswell, Kraus, Miller, & Lambert, 2015). This type of ROM data has highlighted the importance of early sessions of psychological therapy to eventual outcome. However, processes in the earliest sessions have been under-investigated in psychological therapy research, despite evidence from ROM that they might be particularly important for eventual outcome.

Learning from exceptionally effective and efficient psychotherapists and their former patients also appears helpful (B. P. Hansen, Lambert, & Vlass, 2015). It indicates that a particular way of being in interactions with patients is important, through which specific techniques are delivered. However, ROM is limited to general overall feedback and may not help therapists build and improve their therapeutic skills over time. More detailed, interaction-level

feedback may be able to affect the intuitive judgements that make up the majority of a therapist's decision-making (Kahneman, 2011).

Overall, the psychological therapist's in-session thinking, decision-making and responding patterns are:

1. Focused around an interaction with a specific patient;
2. Directly related and adapted to particular responses of the individual patient;
3. Mostly automatic or based on overlearned skills;
4. Unlikely to be systematically and continuously improved long-term by changes to purely strategic decision-making processes;
5. In need of systematic, specific and accurate feedback on changes in interaction required to improve outcome.

There is evidence that this can be and indeed has been done (Blackburn et al., 2001; Moyers, Miller, et al., 2005), but the theory specific nature did not allow for weighting towards associated outcome. An approach more geared towards outcome prediction would better serve this purpose.

Patients with multimorbid physical and mental health problems, using health services at a high rate are a key population where care quality has been lacking. Improving the health of this population is of paramount importance to future developments in healthcare provision. Use of therapeutic interaction skills are likely to be of benefit to patients experiencing multimorbidity and could help improve self-management of long-term conditions. In particular, addressing and enhancing patient motivation or activation level is likely to lead to improved outcomes and reduced healthcare use. An observational assessment of helpful and hindering interactions in clinical consultations for

patients with multimorbidity could contribute to the improvement of interaction skills. Given that psychological therapies are focused on therapeutic interaction-styles, such an instrument would be best first-developed in psychological therapy interactions for those with multimorbidity. The complexity of difficulties presented and requirement for a high level of psychological therapist skill would emphasise differentiation of helpful and hindering interactions and their relations to outcomes. Given that those with multimorbidity have a broad range of difficulties, a range of validated outcome measures should be used to assess outcome prediction. Cognitive Behavioural Therapy for health anxiety aims to improve both physical and mental health and may also reduce healthcare costs. Therefore, patients with multimorbidity, using services at a high rate who are receiving CBT for health anxiety would be an appropriate population to develop such an interaction-rating assessment.

This chapter posits that current methods for helping individual psychological therapists improve their effectiveness are:

4. Too general;

1 Too specific to particular patient groups or disorders and therefore -

2 Have limited applicability to patients with multiple problems;

3 Not sufficiently related to important clinical outcomes or omit important outcomes, and/or

4 Inadequate to account for the combined role patients and therapists play in creating helpful interactions.

Therefore, an assessment tool is proposed in this thesis that aims to address these shortcomings.

In conclusion, there does not currently exist an interaction-level, observer-rated measure of motivational factors that incorporates the combined contribution of therapist and patient factors. In the long-term this type of measure could assess and give feedback on the type of interactions that would lead to better outcomes and help to improve therapist expertise. Therapeutic competency assessments are being used in this way without a strong link to outcome. People with multimorbid physical and psychological health needs could benefit most from this type of improvement. There is also potential that guidance from this type of measure could be beneficial to all healthcare interventions within behavioural medicine, given the importance of clinician-patient interactions in the field.

This thesis describes the development of a measure which aims to address the issues outlined. However, a systematic understanding of literature within this particular area is required. Therefore, the following section describes a systematic literature review of observer-rated, interaction assessments of motivational factors taken during CBT for anxiety disorders, where the results have been used to predict outcomes. This will give an understanding of the characteristic ways in which these types of interaction affect outcome.

2. SYSTEMATIC REVIEW OF OBSERVER-RATED MOTIVATION

2.1 Introduction

Anxiety disorders are a set of highly disabling conditions, affecting approximately 8.2 million people in the UK (Fineberg et al., 2013). Anxiety disorders cost €11.7 billion in the UK during 2010, including direct medical costs and indirect costs such as lost productivity, making them some of the most personally and financially costly of all disorders (Fineberg et al., 2013). There is a wealth of evidence supporting the use of CBT as a first line treatment for anxiety disorders (Cuijpers et al., 2014). However, many people receiving CBT for anxiety either do not benefit at all, only partly benefit or relapse soon after treatment completion (Hunot, Churchill, Teixeira, & de Lima, 2007). Therefore, although CBT for anxiety has a clinical benefit on average, many individuals do not experience the full extent of this potential gain.

Patient motivation has been identified as central to outcomes in CBT and associated therapies (Antony, Ledley, & Heimberg, 2005; Arkowitz, Westra, Miller, & Rollnick, 2008). This is because patients attending CBT are required to actively participate in a range of activities within and outside therapy sessions to gain the most benefit. Consequently, low patient motivation or active resistance to CBT activities can be an important barrier to clinical improvement. Low patient motivation (expressed through noncompliance or resistance to therapist direction) has been identified by CBT therapists as a key barrier to greater clinical effectiveness in treatment of anxiety (Szkodny, Newman, & Goldfried, 2014). Therapists have even suggested that the gap between the effectiveness of clinical research and clinical practice could be explained by the restrictions imposed by resistance

and low patient motivation in CBT for anxiety (McAleavey, Castonguay, & Goldfried, 2014).

Motivation can play such an important role in treatment of anxiety disorders because patients may feel ambivalent about letting go of long-standing strategies for managing the problem, even if there is a clear detrimental effect long-term (De Nadai & Karver, 2013). For example, people with generalised anxiety disorder may feel that there is a protective function to worry, such as helping them to be prepared for the worst (Borkovec & Roemer, 1995). In conjunction with CBT procedures that may require anxiety-sufferers to gradually reduce or fully abandon such behaviours, it is understandable that many patients with anxiety disorders are reluctant to engage with CBT (Dozois, Westra, Collins, Fung, & Garry, 2004). In this way, a patient may be both highly motivated to engage with CBT, but also highly motivated to maintain anxiety-related habits, leading to ambivalence: being motivated for more than one course of action (Westra, 2012). Several expressions of ambivalence have been identified in the literature. Ambivalence can include homework noncompliance; disagreeing or arguing with the therapist; avoidance of certain topics in session, and minimal verbal responses (C. F. Newman, 1994; Westra et al., 2012). Therefore, motivation and associated ambivalence can be expressed more broadly and subtly than just noncompliance. The significance of ambivalence expressed through disagreements has been highlighted by patient reports. Patients reported that points in therapy where they disagreed with therapists' direction were among the most important to their overall experience (Viklund, Holmqvist, & Zetterqvist Nelson, 2010). This suggests that points where therapist direction

clashes with patient motivation can have a high impact on the general view of therapy. Such accounts are consistent with findings from therapeutic relationship rupture and repair research indicating that more intense ruptures lead to poorer outcomes, particularly if they are not acknowledged and addressed (Muran et al., 2009).

In response to the identified importance of motivation in CBT for anxiety disorders and other conditions, self-report measures of patient motivation have been developed. However, significant validity concerns have been identified with this method due to ceiling effects and social desirability biases (W. R. Miller & Johnson, 2008; Westra, 2011). This has resulted in inconsistent associations between self-reported motivation and outcome, likely due to either a lack of variance in responses (i.e. most patients describing themselves as highly motivated) or patients not responding honestly (e.g. fearing possible reprisal if expressing low motivation) (Kampman, Keijsers, Hoogduin, & Hendriks, 2008; Pinto, Pinto, Neziroglu, & Yaryura-Tobias, 2007). Therefore, methods for measuring patient motivation from observations were adopted from the substance misuse field for application to CBT for anxiety. This method is closely linked to the therapeutic modality of MI, wherein therapists attend to patients' expressions of motivation, manage discord with empathic responses, and elicit patients' motivation for positive therapeutic change (W. R. Miller & Rollnick, 2012). Within this framework greater in-session talk about change (or CT) is associated with improved outcomes. Increased talk about staying the same (CCT) is associated with poorer outcomes (W. R. Miller & Rose, 2009). The MI skill code (Glynn & Moyers, 2009) was developed to measure these verbal dynamics observationally. The client resistance code (Chamberlain et

al., 1984) has been used as an observational measure of interpersonal resistance and disagreements specifically.

There have been promising results from using observational assessments of motivation in CBT for anxiety disorders to predict outcomes. This has included narrative attempts to draw the evidence together (Westra & Norouzian, 2018). However, there has not been any systematic review of the whole body of research. Recent meta-analyses suggest that negative elements of observed patient motivation (CCT) were a consistent predictor of behavioural outcomes, but observed positive motivation (CT) was a weak inconsistent predictor of outcome (Magill, Apodaca, et al., 2018; Magill, Bernstein, et al., 2018; Magill et al., 2014). Yet, the studies included were heavily weighted towards MI for substance and alcohol misuse, with none included that reported outcomes from CBT for common mental health problems such as anxiety.

Therefore, this chapter uses a systematic review methodology to ask: Does observer-rated patient motivation predict anxiety-related, post-treatment outcomes in individual CBT for adults with anxiety disorders?

2.2 Method

The protocol for this review was registered prior to completion on Prospero (http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42019124416). The review included studies where predictive models were used to measure the prognostic value of observer-rated patient motivation to clinical outcomes in CBT for anxiety disorders.

2.2.1 Selection criteria

2.2.1.1 Population

The target population was adults with a primary anxiety disorder at entry to the study, receiving CBT adapted to any anxiety disorder in any setting. The review included studies where any anxiety disorders (and other comorbid anxiety disorders) were treated with CBT as at least one of the evaluated treatments. Therefore, studies which included other treatments (e.g. CBT as an adjunct to medication) would also be included.

2.2.1.2 Predictor of interest

The predictor of interest was observer-rated patient motivation, assessed up to (and not beyond) halfway through CBT sessions. This included negative aspects of motivation (e.g. patient resistance) and positive aspects of motivation (e.g. CT). Earlier sessions were the focus of this review because the majority of clinical gains occur in earlier sessions of psychological therapy (Lambert, 2010). In addition, observed motivation has been identified as a significant predictor of outcome from the first session in the substance misuse field, so similar findings may transfer to anxiety disorders (Moyers et al., 2009).

2.2.1.3 Outcome

The main outcomes of interest were self-reported anxiety using standardised assessments. Outcomes were required to be assessed pre-treatment and at least post-treatment (not before), alongside follow-up assessments after treatment.

2.2.1.4 Included study designs

Studies were included where observer-rated motivation was used to predict post-treatment outcome. Prediction models including other additional predictors were included.

2.2.1.5 Exclusion criteria

Studies were excluded where participants had primary disorders which were not anxiety disorders, even if comorbid with an anxiety disorder. However, studies where patients had multiple anxiety disorders were included. Selected studies were restricted to individual CBT. Therefore, studies including group, couple or family treatments were excluded. Neuro-imaging, biomarkers or substrates were not defined as observational measures of motivation in this review and so were excluded (see Table 2.1 for inclusion and exclusion criteria).

Study selection was performed by the thesis author via an initial screening of papers by title and abstract. Studies included or queried at screening then had the full texts reviewed at a second stage.

Table 2.1. Review inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
• Primary research or reviews of primary research • Quantitative research • Treatment studies • Population is being treated primarily for anxiety disorders • Includes treatment of anxiety disorders with individual cognitive behavioural therapy • Includes an observer-rated assessment of in-session motivational factors • Includes outcome prediction using motivational assessment results • Articles written in English • Studies of adults (over 18 years old)	• Qualitative research • Population is primarily treated for a non-anxiety disorder • Studies where cognitive behavioural therapy is not offered as at least one of the treatments • Studies of group-, couples or family- based cognitive behavioural therapy • Studies using neuroimaging or biomarkers and substrates for outcome prediction • No observer-rated assessment of motivational factors is included in the study • Observer rated assessments are not used to predict outcome • Studies involving children

2.2.2 Search Strategy

The search strategy was informed by search terms used in published search strategies from the Cochrane Library, including systematic reviews of CBT and anxiety disorders. Published protocols on Prospero were searched for similar reviews, but no overlapping reviews were found. Three major medical and psychological literature databases were separately searched using the OVID interface: MEDLINE, EMBASE and PsycINFO. Searches were limited to English language publications. No limits were set on the publication period. Studies were identified by combining search terms associated with 1) anxiety disorders 2) CBT and 3) motivational processes (see Appendix A for the full search strategy). Boolean operators were used to identify adjunctive words and phrases that would otherwise be missed. Both free text search terms and subject headings (e.g. MeSH terms in MEDLINE) were used to

identify studies. Reference lists of all included publications were hand-searched for other relevant literature. Search results were imported into Endnote X8.

2.2.3 Quality assessment

There has been much international collaboration in the last five years to establish agreed quality criteria for systematic reviews of prediction models (Debray et al., 2017; Moons et al., 2014; Wolff et al., 2019). This culminated in this year's publication of the Prediction model Risk Of Bias Assessment Tool (PROBAST; Wolff et al., 2019), which has been internationally recognised and recommended (e.g. https://methods.cochrane.org/prognosis/our-publications). The PROBAST system was used to assess both risk of bias and applicability of included studies, because observed patient motivation can be viewed as a potential prognostic indicator (i.e. significant changes in observed motivation may predict significant changes in clinical outcome). The key study features assessed in PROBAST are partly a response to methodological reviews, which identified that reporting of prediction studies was very poor and the methods used equally poor (Bouwmeester et al., 2012; Collins, Omar, Shanyinde, & Yu, 2013; Mallett, Royston, Dutton, Waters, & Altman, 2010; Steyerberg et al., 2010). Furthermore, risk of bias was hardly ever assessed in previous research. Therefore, PROBAST assesses risk of bias in the participants included; the predictors used; the outcomes assessed and the analytic method. In PROBAST each predictor model assessed is also evaluated for applicability to the review question. This includes assessing the applicability of participants, predictors and outcomes. The PROBAST tool was used for

systematic and transparent evaluation of each included study to identify shortcomings likely to lead to distorted estimates of model prediction.

The Quality in Prognosis Studies (QUIPS; Hayden, Côté, & Bombardier, 2006) tool was not used, despite it being appropriate for the review. The rationale for this decision centred on the developments in quality assessment for prediction models in the 13 years since the QUIPS tool was developed. For example, QUIPS does not assess the applicability of each study, whereas PROBAST does. The PROBAST tool is a more strict and rigorous assessment, but there is broad agreement that higher standards should be required when assessing predictive studies in future. One element of PROBAST, which is unlikely to be relevant for the type of studies reviewed is the requirement of calibration or model evaluation to attain low risk of bias. This requirement was excluded from study evaluations, because psychological therapy process models would be highly unlikely to include individual-level outcome prognosis predictions.

2.2.4 Data extraction and synthesis

The CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS; Moons et al., 2014) was used to select recommended data that should be extracted from studies. A meta-analysis would only be performed if two or more studies were identified as validation studies, in accordance with recommendations for prediction model reviews (Debray et al., 2017). Validation studies are those that apply an already developed prediction model to a new sample, as opposed to developmental studies – models developed and tested in a single sample.

2.3 Results

Searches were conducted on 31st January 2019. A total of 4090 titles and abstracts were examined from PsycINFO (n = 1510); MEDLINE (n = 998); EMBASE (n = 1575), and hand searches (n = 7). Duplicate records (n = 1242) and screened exclusions (n = 2755) were removed before full-text assessment of the remaining 86 articles, of which 10 were included in the review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) flow diagram in Figure 2.1 summarises the selection of papers in the review.

2.3.1 Characteristics of included studies

Three included studies assessed positive *and* negative aspects of patient motivation, while seven studies only focused on negative aspects of patient motivation. All ten studies were secondary analyses of two RCTs in which CBT alone was compared to MI as a pre-cursor to CBT (MI-CBT) for generalised anxiety disorder (Westra, Arkowitz, & Dozois, 2009; Westra et al., 2016). Negative aspects of observed patient motivation were assessed in similar ways across studies. Most frequently the adapted Client Resistance Code (Chamberlain et al., 1984) was used to assess patients' reluctance to engage, from coded observations of interactional dynamics. The MI Skill Code (Glynn & Moyers, 2009) was also used to rate observed negative motivation, but through patients' CCT. That is, patients talking about *not* changing in several ways (e.g. lacking desire, showing unwillingness or expressing ambivalence about change). In one study CCT was separated into ambivalent CCT and resistant/oppositional CCT where the patient was directly opposing the therapist's actions. Positive aspects of observed motivation were assessed

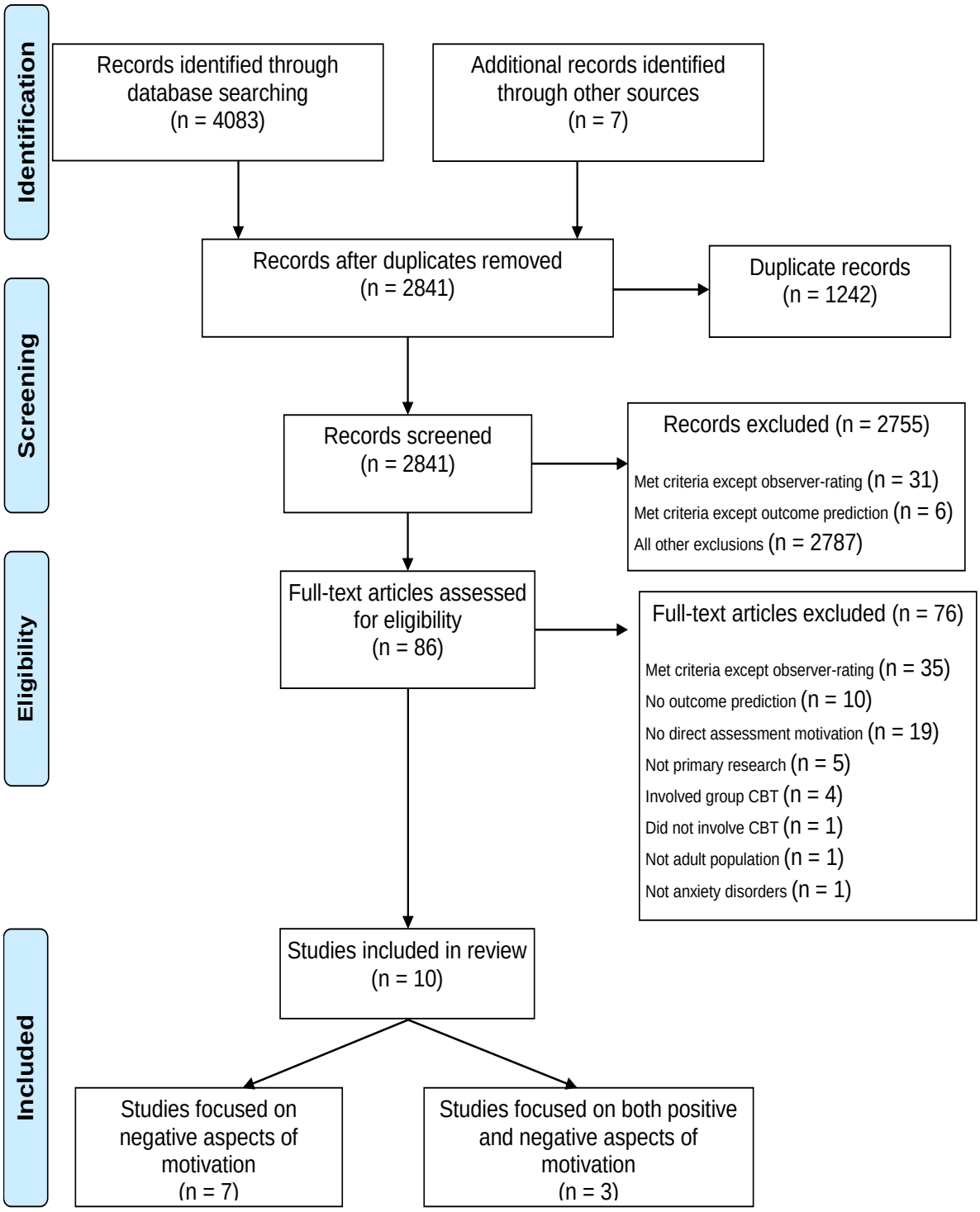
using CT on the MI skill code: Patients talking about desire, hopes, plans or commitments to change in clinically meaningful ways.

A combination of multiple linear regression, Multilevel Modelling (MLM) and path analysis was used across studies to assess predictive validity. Other predictors were modelled with observed motivation. These included:

- patient-rating of therapist empathy (Barrett-Leonard Relationship Inventory; Barrett-Lennard, 1962),
- patient-reported motivation (change questionnaire; W. R. Miller & Johnson, 2008) (client motivation for therapy scale; Pelletier, Tuson, & Haddad, 1997),
- therapeutic alliance (working alliance inventory; Horvath & Greenberg, 1989) and homework compliance (homework rating scale II; Kazantzis, Deane, & Ronan, 2005)
- therapist rating of patients' friendly submissiveness (impact message inventory; Kiesler & Schmidt, 1993)
- and therapist-rated homework compliance (homework compliance scale; Primakoff, Epstein, & Covi, 1989).

The outcome for all included studies was the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) at either post-treatment or one-year post-treatment follow-up. The PSWQ is a widely used 16-item measure of worry with high internal consistency, test-retest reliability and convergent validity (Brown, Antony, & Barlow, 1992; Meyer et al., 1990). Baseline PSWQ score was modelled as a predictor in all studies Therefore, all analyses controlled for baseline severity.

Figure 2.1. PRISMA flow diagram



2.3.2 Quality Assessment

Using the PROBAST criteria, if one assessed area was determined to show a high risk of bias then a rating of “high risk of bias” was given for the study’s overall rating (see Appendix B for an example of a completed PROBAST form for Lombardi et al., 2014). Using this metric, all of the included studies showed a high risk of bias (see Table 2.2 and Figures 2.2 and 2.3 for a cross-study summary). The rating was universally related to the methods of analysis used, but this, in turn, related to problems with either the sample size or the way participants were selected. In some studies sample size was below the recommended 10-20 participants per predictor rule of thumb for regression suitability (Harrell, 2015). The remainder were on the borderline for suitable sample size (all less than 15 participants per predictor). Therefore, all studies were at high risk of bias from the analyses presented and samples used.

The other common source of potential bias was unexplained sub-selection of included participants or incomplete reporting of results. This led to high risk of bias in the majority of studies (7 of 10). Studies presented a smaller sample of participants than had been included in the source trial, but often did not explain either how missing data were managed or why some participants were excluded. In such studies no comparison was presented between included and excluded participants on demographic or clinical characteristics. Therefore, it was unclear whether excluded participants were characteristically different from included participants. Hence, a high risk of bias was introduced. In two studies high risk of bias was identified that related to predictors used in models, because predictors were not assessed or selected

in the same way for all participants. However, the outcomes used were appropriate and relevant across all studies.

In terms of applicability, six of the ten studies were deemed applicable to the review question (see Table 2.2 and Figure 2.3). The other four studies were not focused on the direct prediction of outcome from observed motivation, but were mediation analyses dependent on different treatment groups. These studies focused on whether there was a differential effect of observed motivation on outcome dependent on whether participants received CBT alone or MI-CBT - four sessions of MI as a precursor to CBT. This did not directly answer the review question, but asked whether observed motivation was a more important mediator for MI-CBT versus CBT alone. Furthermore, the direct prediction of motivation on outcome was not reported in these studies.

Taken together, only one study (Westra, 2011) was rated as having high applicability and low risk of bias for participants, predictors and outcome. All studies had participant samples too small for genuinely dependable results, as they had too few participants per predictor for high reliability. As a result, all studies showed high risk of bias in analysis and, therefore, overall high risk of bias. The dependence of all studies on subsets of data from only two RCTs means that the generalisability is limited, because there is likely to be significant overlap in the study samples. Both source RCTs focused on treatment of generalised anxiety disorder, so results of this review may not be applicable to anxiety disorders in general. All studies only used one assessment of worry (PSWQ) as the clinical outcome of interest. Therefore, implications of their findings may not be generalisable to other important

clinical outcomes in anxiety disorders (e.g. severity and chronicity of physiological aspects of anxiety; A. T. Beck, Epstein, Brown, & Steer, 1988). Furthermore, a mixture of partial and semi-partial correlations were reported, so the unique predictive validity of observed motivation cannot be assessed. Lastly, one person (Henny Westra) was an author on all of the included papers and was lead author on both of the source RCTs. Therefore, this author's perspectives are likely to filter through the design, conduct and interpretation of all the results reported.

Overall, during quality assessment three categories of study were identified:

1. Those assessing both positive and negative elements of motivation that were highly applicable to the review question (n = 3)
2. Those assessing only negative elements of motivation that were highly applicable to the review question (n = 3)
3. Those assessing only negative elements of motivation that *not* directly applicable to the review question (n = 4)

Therefore, the narrative synthesis will be discussed in these groups.

Table 2.2. Summary of included studies' risk of bias and applicability

Study	Risk of bias (ROB)				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Studies of both positive and negative observed motivation									
Lombardi et al., 2014	-	+	+	-	+	+	+	-	+
Poulin et al., 2018	-	+	+	-	+	+	+	-	+
Button et al., 2015	-	+	+	-	+	+	+	-	+
Studies of negative observed motivation									
Constantino et al., 2017	+	+	+	-	+	-	+	-	-
Aviram et al., 2011	-	-	+	-	-	-	+	-	-
Aviram et al., 2016	-	-	+	-	-	-	+	-	-
Coyne et al., 2018	+	+	+	-	+	-	+	-	-
Hara et al., 2015	-	+	+	-	+	+	+	-	+
Sijercic et al., 2016	-	+	+	-	+	+	+	-	+
Westra, 2011	+	+	+	-	+	+	+	-	+

Note. " + " = low risk of bias / high applicability; " - " = high risk of bias / low applicability

Figure 2.2. Risk of bias across studies

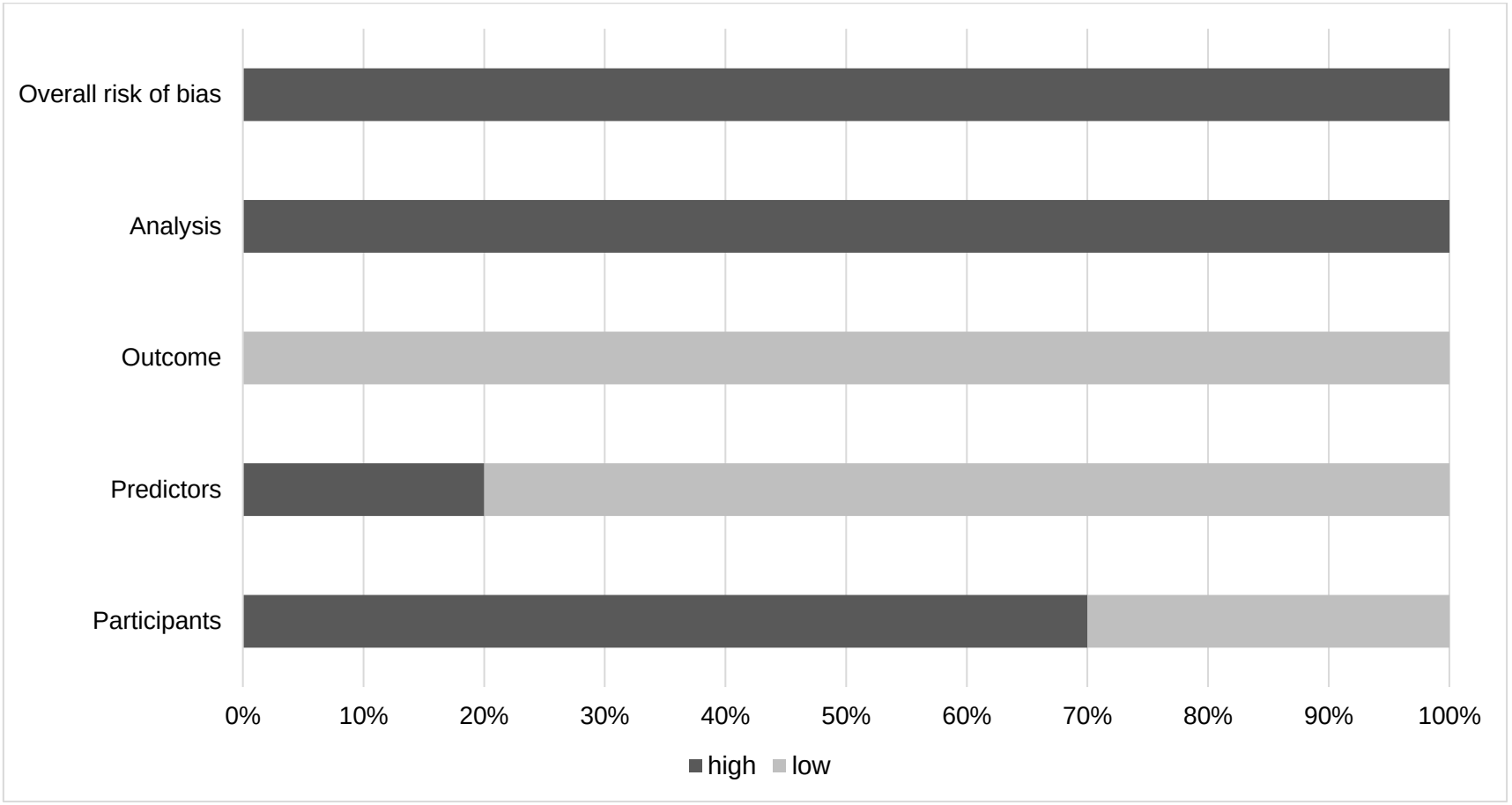
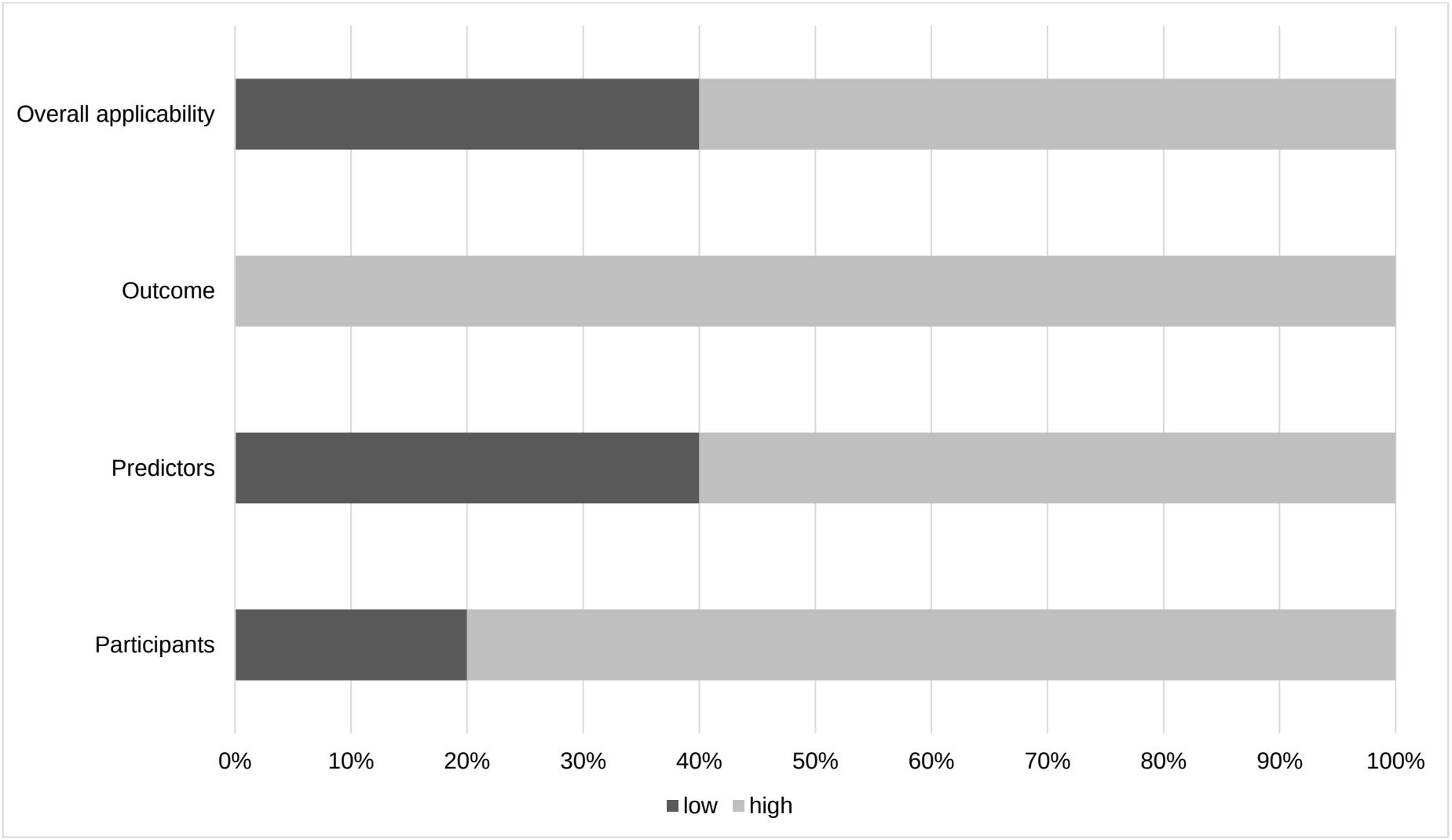


Figure 2.3. Model applicability across studies



2.3.3 Narrative Synthesis

All the included studies conducted developmental rather than validation studies. Therefore, no meta-analysis was performed as only validation studies are suitable (Debray et al., 2017). Furthermore, included studies inconsistently reported R^2 for partial correlations, semi-partial correlations and the whole final model. Therefore, direct comparisons across studies were not possible. Findings from each sub-group of studies will be discussed in turn, through narrative synthesis (see Tables 2.3, 2.4 and 2.5 for study summaries).

2.3.3.1 Applicable studies of positive and negative motivation

All three studies assessing positive and negative elements of observed patient motivation reported that negative aspects of motivation were a more consistent predictor of outcome than positive aspects of motivation (Button et al., 2015; Lombardi, Button, & Westra, 2014; Poulin, Button, Westra, Constantino, & Antony, 2018). Therefore, patient resistance to therapist actions and, to a lesser extent, patients' CCT (talking about maintaining current behaviour, beliefs and attitudes without change) negatively affected improvement in post-treatment worry. Negative aspects of observed motivation also outperformed standardised assessment of self-reported motivation. One measure of self-reported motivation (the Change Questionnaire) predicted post-treatment outcome in one study, but not another. The Client Motivation for Therapy Scale did not predict outcome in either study where it was assessed. Conversely, observed motivation consistently predicted outcome in both studies that compared self-reported motivation with observed motivation. Overall, these studies suggested that observed motivation is a more reliable

predictor of outcome than self-reported motivation and negative aspects of observed motivation are a more potent predictor than positive aspects.

2.3.3.2 Applicable studies of negative aspects of motivation

The three applicable studies investigating negative aspects of observed motivation added important nuances to the above reported findings. Firstly, observed resistance (negative motivation) was found to be superior to therapist-rated resistance (Hara et al., 2015; Westra, 2011). The authors inferred that this meant therapists were unlikely to be adequately sensitive to patient resistance for meaningful in-session adaptation to occur. One study (Sijercic, Button, Westra, & Hara, 2016) combined CCT with resistance to sub-categorise CCT into talk which expressed *ambivalence* about change and CCT that expressed *opposition* to therapist direction, using the interactional context of the surrounding turns-of-speech. This study suggested that *oppositional* CCT predicted post-treatment and one year follow-up outcomes, but *ambivalent* CCT did not when they were modelled together. In summary, these studies suggest that interactionally oppositional CCT is the most potent sub-category of negative patient motivation with regard to outcome prediction. In addition, observed motivation is superior to both self-rated and therapist-rated motivation.

2.3.3.3 Inapplicable studies investigating negative aspects of motivation

Although the studies described below do not have direct relevance to the review question, they contain peripherally relevant findings. One study (Aviram & Westra, 2011) identified that first-session resistance outperformed therapist-rated homework compliance across therapy as a mediator of treatment condition and post-treatment worry (MI-CBT more than CBT alone).

Mid-treatment resistance also outperformed patient-rated homework compliance, therapeutic alliance and empathy as a mediator of treatment condition and 12 month post-treatment worry (MI-CBT more than CBT alone) (Constantino, Romano, Coyne, Westra, & Antony, 2017). The different treatment-related features of MI-CBT versus CBT were highlighted by the finding that resistance-reduction was a more important outcome mediator than "friendly-submissiveness" (or compliance) in MI-CBT than CBT alone (Coyne, Constantino, Westra, & Antony, 2018). Whereas the opposite was found for CBT alone: fostering compliance was a more important predictor of outcome than reducing-resistance. The findings from this study suggest that the mediation analyses above may not be applicable to CBT alone and so are not directly relevant to the review question at hand. Nonetheless, Aviram et al. (2016) highlighted an important distinction in responsiveness to resistance in CBT alone. They reported that where CBT therapists responded with empathy at times of interpersonal resistance or disagreement better outcomes ensued by comparison to where therapists did not respond empathically. However, general expressions of empathy by therapists (i.e. at randomly selected time points where disagreements were not occurring) did not predict outcome. This suggests that identifying and specifically responding empathically to patient expressions of negative motivation can have an impact on outcome, but being empathic in general is less important. Although this study may say something about how to respond to expressions of negative motivation, it can say little about the overall predictive value of negative aspects of motivation for outcome.

In summary, some means of addressing the problems posed by negative patient motivation are identified through these mediation analyses. However, they can say little about the specific review question. They were still generally supportive of the notion that observed negative motivation predicts post-treatment outcome.

Table 2.3. Summary of studies investigating both positive and negative elements of observed motivation that were applicable to the review

Study	Participants	Outcomes predicted	Candidate predictors	Missing data	Model	Results	Interpretation and discussion
Lombardi et al., 2014	n = 37. CBT only participants from Westra et al's (2009) RCT of CBT vs MI-CBT for GAD	Post-treatment PSWQ	Observed 1st session positive and negative motivation (CT and CCT; MISC 1.1); self-reported motivation (CMOTS and CQ); baseline worry (PSWQ)	Where session 1 recordings were missing, session 2 data used instead	Multiple linear regression.	Observed negative motivation (CCT) and one self-reported motivation measure (CQ not CMOTS) predicted post-CBT worry. Observed positive motivation did not (CT)	Early observed negative motivation a better predictor of GAD outcome than observed positive motivation
Button et al., 2015	n = 37. CBT only participants from Westra et al's (2009) RCT of CBT vs MI-CBT for GAD	Post-treatment PSWQ	Observed 1st session positive and negative motivation (resistance: CRC; CT and CCT: MISC 1.1); baseline worry (PSWQ)	Where session 1 recordings were missing, session 2 data used instead	Bootstrapped mediation analysis	First-session resistance (negative motivation) predicted post-treatment worry, but first session CT (positive motivation) and CCT (ambivalence) did not	Early observed negative motivation a better predictor of GAD outcome than observed positive motivation
Poulin et al., 2018	n = 67. from Westra et al's (2016) RCT of CBT vs MI-CBT for GAD	Post-treatment and 12-month PSWQ	Observed 1st session positive and negative motivation (CT, CCT and resistance; MISC 1.1 and CRC); self-reported motivation (CMOTS and CQ); baseline worry (PSWQ)	Complete case analysis only used	Multilevel modelling: clients nested in therapists	All elements of observed first session motivation (CT, CCT-A and CCT-R) predicted worry at 12 months post-treatment, though CT was not significant post-treatment. Self-reported motivation and baseline worry did not	Early observed negative motivation is a more consistent predictor of CBT for GAD outcome than observed positive motivation. Self-reported motivation is not a predictor of GAD outcome

Note. Abbreviations: CBT: cognitive behavioural therapy; MI-CBT: motivational interviewing as a precursor to CBT; GAD: generalised anxiety disorder; PSWQ: Penn state worry questionnaire; CT: change talk; CCT: counter-change talk; CMOTS: client motivation for therapy scale; CQ: change questionnaire; CRC: adapted client resistance code; MISC: motivational interviewing skill code; CCT-A: ambivalence-related counter-change talk; CCT-R: resistant/oppositional counter-change talk.

Table 2.4. Summary of studies investigating negative elements of observed motivation that were applicable to the review

Study	Participants	Outcomes predicted	Candidate predictors	Missing data	Model	Results	Interpretation and discussion
Hara et al., 2015	n = 40 receiving CBT alone from Westra et al's (2016) RCT of CBT vs MI-CBT for GAD.	Post-treatment PSWQ	Observed early-session (2-7) negative motivation (resistance; CRC) average therapist rated resistance; baseline worry (PSWQ).	Where session 1 recordings were missing, session 2 data used instead	Multiple linear regression.	Early observed resistance predicted post-treatment worry, whereas therapist rated resistance did not	Observed resistance is superior to therapist rated resistance as an outcome predictor in CBT for GAD
Sijercic et al., 2016	n = 37. CBT only participants from Westra et al's (2009) RCT of CBT vs MI-CBT for GAD.	Post-treatment and 12-month PSWQ	Observed first session negative motivation (resistance: CRC; ambivalent and oppositional CCT: MISC 1.1); baseline worry (PSWQ).	Where session 1 recordings were missing, session 2 data used instead	Multiple linear regression.	Observed oppositional CCT predicted post-treatment and 1-year follow-up worry, but ambivalent CCT does not when modelled together	Observed oppositional interactions are a better predictor of outcome than ambivalent interactions in CBT for GAD
Westra, 2011	n = 38. CBT only participants from Westra et al's (2009) RCT of CBT vs MI-CBT for GAD.	Post-treatment and 12-month PSWQ	Observed first session negative motivation (resistance: CRC); self-reported motivation (CQ and CMOTS); baseline worry (PSWQ).	Where session 1 recordings were missing, session 2 data used instead	Multiple linear regression.	Observed first-session resistance predicted both post-treatment and 1-year follow-up worry, whereas self-reported motivation did not when modelled together	Observed negative motivation is superior to self-reported motivation as an outcome predictor in CBT for GAD

Note. Abbreviations: CBT: cognitive behavioural therapy; MI-CBT: motivational interviewing as a precursor to CBT; GAD: generalised anxiety disorder; PSWQ: Penn state worry questionnaire; CT: change talk; CCT: counter-change talk; CMOTS: client motivation for therapy scale; CQ: change questionnaire; CRC: adapted client resistance code; MISC: motivational interviewing skill code.

Table 2.5. Summary of studies investigating negative elements of observed motivation, but not applicable to the review

Study	Participants	Outcomes predicted	Candidate predictors	Missing data	Model	Results	Interpretation and discussion
Aviram et al., 2011	n = 35. High worry severity participants from Westra et al's (2009) RCT of CBT vs MI-CBT for GAD.	Post-treatment PSWQ	Observed negative motivation at 1st session (resistance; CRC); therapist-rated homework compliance (HCS); baseline worry (PSWQ).	Where session 1 recordings were missing, session 2 data used instead	Path analysis.	First-session resistance had an indirect effect on post-treatment worry. It outperformed homework compliance as a mediator of post-treatment worry	Observed negative motivation at the first and second session predicts outcome in CBT and MI-CBT. It may be a more potent predictor than therapist-rated homework compliance
Aviram et al., 2016	n = 30 receiving CBT alone from Westra et al's (2016) RCT of CBT vs MI-CBT for GAD.	Post-treatment PSWQ	Observed negative motivation pre-disagreement and post-disagreement (resistance; CRC); general and disagreement-specific MI-behaviours; mid-treatment CBT competence; baseline worry (PSWQ).	No information, presumed excluded	Correlation and multiple linear regression.	Pre-disagreement resistance did not correlate with post-treatment worry, but post-disagreement resistance did. Resistance-related regression results were not reported	MI-type behaviours at times of disagreement are predictive of outcome, but MI-type behaviours in general are not.
Constantino et al., 2017	n = 85 from Westra et al's (2016) RCT of CBT vs MI-CBT for GAD.	12-month PSWQ	Observed mid-treatment negative motivation (resistance; CRC); patient rating of therapist empathy (BLRI), working alliance (WAI-C) and homework compliance (HRS-II); baseline worry (PSWQ).	modelled using multilevel modelling	Bootstrapped mediation analysis	Mid-treatment resistance had an indirect effect on 12-month worry and outperformed empathy, homework compliance and alliance as mediator of 12-month post treatment worry	Observed negative motivation at mid-treatment predicts outcome in CBT and MI-CBT. It may be a more potent predictor than self-reported empathy, working alliance or homework compliance.
Coyne et al., 2018	n = 85 from Westra et al's (2016) RCT of CBT vs MI-CBT for GAD.	Post-treatment PSWQ	Observed mid-treatment negative motivation (resistance; CRC); therapist-rated friendly submissiveness through therapy (IMI); baseline worry (PSWQ).	Last observed carried forward	Bootstrapped mediation analysis	Mid-treatment resistance had an indirect effect on post-treatment worry for MI-CBT patients; friendly submissiveness had an indirect effect on outcome for CBT patients	Reduced client resistance at mid-treatment predicts improved outcome in MI-CBT. Increased friendly submissiveness in CBT improves worry reduction

Note. Abbreviations: CBT: cognitive behavioural therapy; MI-CBT: motivational interviewing as a precursor to CBT; GAD: generalised anxiety disorder; PSWQ: Penn state worry questionnaire; CRC: adapted client resistance code; HCS: homework compliance scale; BLRI: Barrett-Lennard Relationship Inventory; WAI-C: working alliance inventory – client-rated; HRS-II: homework rating scale; IMI: impact message inventory.

2.4 Discussion

2.4.1 Summary of evidence

This review indicates that observed patient motivation is a significant predictor of outcome in CBT for anxiety disorders across six applicable predictor studies from two relevant RCTs. Results also suggest that observed *negative* motivation is a more consistent and potent outcome predictor than observed *positive* motivation. Specifically, negative motivation that accounts for the interactional context (e.g. turns of speech leading to expressions of negative motivation - as in observed resistance) performed better than measures of negative motivation abstracted from interpersonal context (e.g. patient speech about not being ready, willing or able to change in isolation from therapist interactions). Observed motivation also consistently outperformed other measures of patient motivation including two forms of patient-reported motivation and therapist-rated resistance. There were indications that observed motivation might outperform empathy, therapeutic alliance and homework compliance, but associated studies were not designed to directly answer the review question. Notably, most of the studies in this review identified the predictive validity of observed motivation at the first session of CBT. This suggests that there are important prognostic markers of motivation in the very first interactions of CBT for anxiety disorders.

These findings fit with previous literature identifying motivation as an important prognostic factor in CBT for anxiety disorders (Antony et al., 2005). This review partially supports the proposal that patient ambivalence predicts outcome in anxiety disorders (Westra & Arkowitz, 2010). Where ambivalence assessment took account of the interpersonal context, as in observed

resistance, consistent outcome prediction was observed up to one year post-treatment. However, more inconsistent results were found from assessment of ambivalent speech, measured in isolation from the turns-of-speech surrounding the assessed utterance. Inconsistent results were also observed for positive aspects of motivation (talking about desire, willingness and commitment to change), with most studies suggesting a much weaker relationship to outcome than negative motivation. This fits with recent findings from the measurement of observed motivation in MI (Magill, Apodaca, et al., 2018; Magill, Bernstein, et al., 2018; Magill et al., 2014). Yet, this may be inconsistent with the theoretical importance placed on eliciting talk about change in MI (W. R. Miller & Rollnick, 2012). Current MI theory appears to place equal weight on positive and negative aspects of observed patient motivation, but this review suggests that negative aspects of motivation predict outcome more reliably – in anxiety disorders at least.

2.4.2 Limitations

The findings reported may only be applicable to the outcome of worry in generalised anxiety disorder, because no other outcomes or disorders were assessed in included studies. Generalised anxiety disorder could be viewed as having features relevant to all anxiety disorders, but little can be said about the specifics of any other disorder. This limitation also means that little can be inferred about non-worry related outcomes, such as the frequency and severity of bodily anxiety symptoms.

With respect to study quality, all of the included studies had a high risk of bias associated with the findings reported. Sample size and analytic methods applied to prediction models were both problematic, thus, restricting

the dependability of the results. Specifically, the incomplete reporting of results or absence of account for missing data affected all applicable studies. All studies were also small in scale and scope with none reaching the top-end of minimum participant-to-predictor ratios (Harrell, 2015). Furthermore, only data from two source RCTs were used across all studies and both had the same first author, who also authored all the included studies. This suggests a limit to the range of perspectives, measurement and analytic techniques which might reduce the generalisability of findings. It is notable that more sophisticated prediction methods, such as MLM were not used to account for repeated outcome measurement in any study. This was the case despite session-by-session outcome measurements being taken and multiple follow-up assessments in both source trials. Accounting for repeated measurement would help to provide more accurate estimates of predictive effects, by accounting for change over time rather than at a single outcome point.

In spite of these limitations, it is important to note that similar criticisms are endemic in the whole field of predictive modelling (Bouwmeester et al., 2012). Furthermore, the high standard of analysis required in PROBAST aimed to establish a benchmark for future prediction modelling research, in acknowledgement that previous studies have been seriously lacking (Wolff et al., 2019). Given this context, it was always likely that a systematic review of previous predictive modelling using PROBAST would find problems with historic methodology and reporting. The predictive models reviewed are also not typical for larger-scale, health research predictive modelling. Large-scale models often use a combination of risk (e.g. smoking) or prognostic variables (e.g. exercise) to predict specific health outcomes (e.g. cardiovascular

disease) across much larger patient populations. Therefore, some recommended methodologies may not effectively map onto process research within psychological therapy, but key methods were excluded from quality assessment (e.g. calibration). It is therefore, possible that the QUIPS quality assessment and its more generous rating criteria could have been more appropriate for small studies of predictive factors. Nevertheless, the high risk of bias reported could also be indicative of the gap in methodological standards between large-scale epidemiological research versus the comparatively small-scale models in psychological therapy research.

2.4.3 Implications

This review suggests that observed motivation is an important prognostic process marker from the very start of CBT for anxiety disorders. In particular, negative aspects of motivation, such as opposition to therapist direction should be addressed by therapists, perhaps using MI skills to respond (Westra & Norouzian, 2018). This review also suggests that observed motivational assessment should account for interpersonal context and emphasise negative aspects of motivation more than positive elements. These assessments should be used over and above patient- or therapist-rated motivation measures, which do not appear to predict outcomes as effectively. Given that therapists seem insensitive to expressions of motivation that *can* be detected by observers, greater emphasis should be placed on supporting CBT therapists to recognise motivational language use and how to respond most effectively. Current evidence suggests that integration of MI techniques at points of therapist-patient disagreement may support improved effectiveness of CBT for generalised anxiety disorder (Aviram et al., 2016; Westra et al.,

2016). Yet, there may be other related means of responsive intervention if therapists are alert to the way motivation can present through patients' in-session language.

The major methodological limitations of all reported studies should be addressed in future research. If included studies are typical for the field, it is unlikely that sufficient statistical power will be achieved from individual investigations. Therefore, patient-level aggregation of data may be needed to combine data across appropriate studies and obtain more dependable results. Ideally, this would include studies not originally set up to address motivation in the CBT intervention, perhaps where appropriate consents to use session recordings in secondary analyses have been obtained. Then observed motivation can be assessed by unrelated coders and outcome predictions made by independent researchers. This would help to reduce the potential for bias, spread the range of researchers involved in associated work, as well as extending the sample size with consistent analytic methods. Larger samples would make it possible to use more robust methods of evaluating prediction models, account for missing data and generate more reliable regression models (Wolff et al., 2019). In particular, it would be helpful to include studies other than those informed by MI. Despite the important and significant contribution made by Westra and colleagues, all of the current research on observed motivation in CBT for anxiety disorders appears to have been carried by a similar research group. This introduces potential bias, because their perspective is clearly informed by MI theory and technique. The MI approach underpinned the conduct of their studies. It is therefore possible that unintended bias could have been introduced even to CBT without MI in

associated trials, due to the researchers' a priori perspective. Consequently, the CBT offered in the trials used for this review might be more focused on addressing patient motivation (and hence, observed motivation becomes a more important predictor of outcome) than more typical CBT for anxiety disorders.

Although this review indicates the potential value of negative elements of observed motivation, further research is required to better conceptualise positive aspects. The imbalance in predictive validity between negative and positive aspects of observed motivation may constitute an empirical difference: negative aspects of motivation are, in fact, better predictors of outcome than positive aspects. Alternatively, the imbalance could relate to problems with the way positive motivation has been conceptualised. For example, negative aspects of motivation were most consistently and potently predictive of outcome when the *interpersonal context* of negative motivation was incorporated: measuring both patient language and the interactional context of surrounding therapist speech. However, there was/is no contextualised comparison measure for positive aspects of motivation. Only abstract positive patient motivational language was used to predict outcome. Therefore, future research should aim to develop measures of observed positive motivation that account for interpersonal context surrounding patient motivational talk. Similarly, some of the most persuasive results were produced when combining language based-assessment of negative motivation (the MI skill code) and interaction-based observations of negative motivation (the client resistance code). Therefore, future research might benefit from a measure that combines

the linguistic and interactional elements of observed motivation for both positive and negative aspects of motivation in one assessment.

2.4.4 Conclusions

This chapter presents a systematic review that supports the use of observed patient motivation assessment in CBT for anxiety disorders to predict post-treatment outcomes. The review suggests that this method may be superior to other means of rating motivation completed by therapists or patients. Furthermore, outcomes can be predicted using observed motivation from the very start of treatment. This suggests that there are prognostic indicators in patient-therapist interactions from the very first exchanges in therapy. Future research should address the methodological limitations present in all the included studies. In particular, use of an adequate sample size, acceptable handling of missing data, use of repeated measurement and validation of predictive findings should be carried out. Future studies should also go beyond generalised anxiety disorder as a target condition and worry as a target outcome. A broader range of researchers should also be included in future research to avoid potential bias. Research on observed motivation in CBT would benefit from an integrated assessment tool which incorporates both linguistic and interactional elements of positive and negative motivation.

The current thesis sets out to address the shortcomings of current research in this area, by developing an integrated assessment of positive and negative aspects of motivation where ratings combine both language used and interactional processes observed. The development and validation of the tool will aim to address concerns about sample size, handling of missing data and analytic methods used. The thesis will also add a new researcher perspective

to observational assessment in CBT for anxiety disorders. In addition, it will assess health anxiety and other conditions to broaden the perspective beyond generalised anxiety. Greater breadth will also be added in terms of outcomes predicted, with multiple anxiety-specific measures included, but also important outcomes used less frequently in psychological therapy process research: Quality of life; general health, and functional impairment. In summary, the research presented in this thesis aims to address some of the key criticisms of outcome-prediction research published thus far in observed motivation during CBT for anxiety disorders. This thesis also aims to make a unique contribution to the way that observed motivation is assessed. It aims to make ratings more accessible across health conditions and meaningful in terms of both speech and interactional context.

3. AIMS AND HYPOTHESES

3.1 Thesis plan

This thesis describes a mixed methods approach to the development of an assessment called the Consultation Interactions Coding Scheme (CICS). Observational assessors use the CICS to rate each patient-therapist interaction on the basis of patient activation. The clinical utility of the CICS is then assessed with a series of following studies: The convergent validity is assessed with therapeutic alliance (using the observer-rated working alliance inventory: Horvath & Greenberg, 1989) and therapist competence (using the CTS-R; Blackburn et al., 2001). The predictive validity of the CICS for a range of health outcomes is then assessed in a larger sample. Finally the sensitivity of the CICS to session-by-session changes is evaluated. These assessments are taken from CBT sessions with patients who are high users of either primary care or urgent care and experience comorbid physical and mental health problems.

3.2 Aims

The overall aim of this thesis is to evaluate the CICS as a possible enhancement to therapist feedback methods by clarifying what types of therapist-patient interaction are most important to different categories of clinical outcome. It also aims to broaden the application of patient activation, by evaluating whether the concept remains a reliable predictor of important physical and mental health outcomes when applied to interaction-level, observer rated data from CBT sessions. This aims to be a stern test by evaluating the predictive power of the CICS amongst patients that healthcare

services have struggled to consistently help: those with multimorbid physical and mental health problems who use services at a high rate. Furthermore, the strict focus on very early interactions makes for a stringent assessment of clinical utility.

The CICS itself aims to address some key issues with other measures with similar aims:

2. It overcomes several issues with self-report measures. There are problems with the reliability of self-reports and positive skews on associations with outcome, due to common methods variance.
3. This validation will establish IRR turn-by-turn rather than on total scores or behaviour counts. This would give assurance that agreement on specific types of patient-therapist interactions has been reached, so more reliable inferences can be drawn. No other measures reviewed have achieved this form of reliability.
4. This validation will place important clinical outcomes at the centre of the evaluation of the CICS. This means evaluating outcome association with measures that are rarely used in process-outcome research in psychological therapies (i.e. measures of quality of life, general health and physical symptoms). A focus on outcomes also means evaluating the relationship in a variety of ways so that more robust conclusions can be drawn. This includes extreme-case comparison; correlational methods; use of sophisticated statistical methods to account for covariance of outcomes over time within and between patients and therapists, and use of time series outcome data to establish temporal precedence in case-comparisons.

5. Contrary to most previous studies, the validation of the CICS will address the interaction-styles and types most important to outcome for patients experiencing both physical and mental health comorbidities. Furthermore, clinical outcomes that are important to this population will be assessed. This is a very influential patient population for healthcare services as a whole and an under-researched group in psychological therapy process-outcome research.
6. The CICS aims to account for the interplay between patient presentations and therapist responsiveness in a single assessment, whereas most studies of this type have used individual evaluations of each. This occurs despite agreement that an understanding of the interface between patient and therapist interaction-styles and types is important for predicting outcome. The CICS aims to address this issue through combined assessment of both therapist and patient factors from in-session interactions.

These features are addressed in the CICS to support development of a more clinically valuable tool in the identification of important process markers from in-session interactions and effective development of therapist responses that promote outcome improvement.

3.3 Objectives

The thesis objectives can be broken down into those that apply to the *development* of the CICS and those that apply to evaluating the *clinical utility* of the CICS.

3.3.1 Objectives and hypotheses: The CICS development

Given the paucity of research focused on psychological therapy interactions amongst high service utilisers with multimorbidity, an inductive thematic analysis is used to establish interaction themes arising in CBT sessions for the population. In a secondary, deductive thematic analysis these themes are sub-categorised into those hypothesised to improve outcome and those that hinder improvement on the basis of observable patient activation level. This differentiation is translated into a scoring system, with higher scores where interactions were observed to show greater levels of patient activation. The main hypothesis for this developmental stage was that those with higher patient activation observed in session interactions would have better outcomes at six month follow-up than those with lower scores. Secondary hypotheses posited that CICS scores rated in the above described way would be associated with standardised, observer ratings of therapeutic alliance and therapist competence.

3.3.2 Objectives and hypotheses: Clinical utility of the CICS

The CICS was then applied to a larger, similar but independent sample of patients, again receiving CBT. The key objective for this study was to establish the clinical utility of the CICS from the very start of therapy. That is, could the CICS predict clinical outcomes up to one year later from initial therapy sessions. This would be an important, demanding and stringent test. But if achieved, the CICS could be a helpful guide for the prevention of therapeutic dropout, given that most dropout occurs very early in therapy. Furthermore, most therapeutic improvement occurs in the early sessions, so clarifying helpful and hindering processes at this stage for an important patient

group would be informative. Secondary to this overall objective, the evaluation aimed to clarify any interaction themes that consistently predicted outcomes, which could be instructive to therapists. Lastly, this evaluation aimed to clarify what outcomes (if any) this type of assessment might predict. Hypotheses tested in this study were:

1. That interaction-styles and types identified by the CICS would predict a range of outcomes up to six months later.
2. That interactions in the first and second session would still be important predictors of clinical outcomes up to one year later, as rated by the CICS.

There were no a priori hypotheses set about which outcomes would be predicted or which interactions would be more predictive of outcome than others, give that this was a developmental, exploratory set of studies.

As a detailed development and test of the findings from the above study, session-by-session outcome ratings were used to assess whether the CICS predicted proximal outcomes (session-by-session changes in outcome using ROM) as well as distal outcomes (outcomes assessed several months post-therapy). The primary hypothesis in this study was that increases in patient activation would precede improvements in session-by-session outcomes. To inform the value of this assessment, the relationship between session-by-session outcomes and a range of longer-term outcomes was then evaluated. Therefore, the credibility of any value the CICS showed in predicting proximal outcomes would also be linked to distal outcomes.

At each stage of the thesis a summary of each section is given. At the end of the thesis the results, implications and future directions are discussed (Chapter 9).

4. DEVELOPMENT OF THE CICS: THEMATIC ANALYSIS

4.1 Introduction

Patient activation is defined as the degree to which a person feels confident and able to be actively involved in managing their own health (Hibbard et al., 2005). The concept has received little attention within psychological therapy research, despite widespread use in behavioural medicine (NHS England, 2018). There are currently no measures of patient activation at the level of psychological therapy interaction and no studies have explored its ability to predict outcome in psychological therapy. Qualitative analyses have often been used in the initial development of psychometric measures; indeed qualitative methods drove the development of the PAM used to assess patient activation (Hibbard et al., 2004).

This chapter presents the qualitative analyses used to develop the Consultation Interaction Coding Scheme (CICS), an instrument designed to rate individual psychological therapy interactions based on patient activation. The primary aim was to initially develop a detailed understanding of what is talked about in CBT sessions for high service utilisers with multimorbidity. It is important to note that motivational processes and particularly patient activation were not chosen as guiding theoretical frameworks prior to engagement with the data, but were selected after initial descriptive analyses. Qualitative analysis of psychological therapy interactions has been lacking in general, but particularly amongst patients with multimorbidity where detailed research has often failed to explore mental health treatments in detail (Naylor et al., 2016; Naylor et al., 2012). The results of this analysis are then used to conduct a theory-informed approach to sub-categorising interaction themes into those

likely to be helpful and those likely to be hindering in relation to outcome, on the basis of patient activation. The next chapter describes the mixed methods used to integrate results from this study into quantitative assessments of reliability and validity.

4.1.1 Rationale for a qualitative approach

"We should adapt our methods, so far as we can, to the object [of study] and not define the object in terms of our faulty methods" (Allport, 1963, cited in Smith, 1996).

Allport's statement reflects the philosophical criticisms levelled at the quantitative approach over many years. Critique from feminism, post-structuralism, Marxism and psychoanalysis suggest that the, at times, "naïve positivist" assumptions and reductionist methods may be unsound (summarised in Table 4.1; Parker, 2005). For example, taking patient self-reports at face value can underplay the context in which reports are made and the influences at work in what is said and done. Furthermore, the way quantitative research variables are operationalised can narrow the potential for new understandings (e.g. the reduction of attitudes to a five point scale; Coolican, 2019). This can mean that important detail is lost in quantitative analysis and results, which might enhance understanding, particularly in new or emerging fields.

Table 4.1. Summary of perspectives contributing to the qualitative research approach (adapted from Parker, 2005)

Theoretical Perspective	Ideology	Relevance to CICS development
Feminism	3. Traditional psychology is dominated by masculine characteristics, leading to illusionary claims of precision and measurement in psychological research. 4. Feminine characteristics, including acknowledgement of feelings and experiences are ignored despite their importance.	Initial aims were to explore experience rather than measure any component of psychological therapy experience.
Foucauldian Post-Modernism	1. The researcher's vested interests influence how knowledge is sought and the knowledge produced. 2. The role of the researcher and their methods are as important as the research results.	Iterative researcher reflexivity was central to the quality enhancement methods used. These aimed to best account for views and biases brought by researchers to data analysis.
Psychoanalysis	1. Psychology maintains rigidity and structure in its research due to anxiety about the true unordered nature of psychological experience.	Flexible application of qualitative methods was used to best address the research question, rather than rigidly applying formulaic methods.
Marxism	2. Reality is socially constructed between people, particularly through language. 3. Participants of psychological study cannot be separated from the contexts in which they are found.	The interpersonal, interactional elements of psychological therapy conversations were acknowledged in the analytic methods used.

When considered with the aims of this study: gaining a detailed understanding of patient-therapist interactions in CBT, within an under-researched population; there is a risk that an exclusively quantitative approach may miss important elements of the phenomena under study. Alternatively, qualitative approaches highlight the importance of evaluating the underlying assumptions in psychological research. For example, there may be personal and professional reasons that influence a researcher's approach and the questions they ask, which in turn affect the way a study is conducted. Furthermore, quantitative methods can reduce the room for new discoveries,

partly due to necessary and rigorous assessment of specificity (Nowell, Norris, White, & Moules, 2017). Qualitative approaches, on the other hand, contribute to the *generation* of hypotheses rather than the *testing* of hypotheses (Jones, 1995). Overall, qualitative methods, particularly those of thematic analysis, give opportunity for discovery of new phenomena and allows the flexibility to integrate analysis with relevant theory, which is appropriate for areas where much remains unknown (V. Braun, Clarke, Hayfield, & Terry, 2012). However, qualitative research is not a single approach and represents a number of philosophical perspectives; Reicher (2000) summarised these perspectives into two groups:

1. *Discursive approaches*, which question the stability and reality of experience interpretations, assuming that they are constructed as part of a social function.
2. *Experiential approaches*, which focus on human experiences within their context alongside consideration of reflexivity (taking a self-critical account of the research process and the contribution made by the researchers' own interpretations).

Both approaches produce research which attempts to take account of the complexity of human studies, but in different ways. This study takes an experiential approach, because it is primarily focused on the human experience of psychological therapy and less emphasis is placed on the external, contextual factors which contribute to the construction of psychological therapy narratives.

Patton (1980) argued that qualitative research should not aim to nullify the usefulness of quantitative research, but highlight the potential for other,

more applicable methods to some research questions. This underlines the need for pragmatism in research methodology, not letting dominant traditions come before appropriate research design. It was from this perspective, of finding a methodology which fitted the research question that qualitative methods were chosen and the specific method of thematic analysis selected.

4.1.2 Qualitative methods in psychological therapy

The use of qualitative methods to inform theory and practice of psychological therapy has grown over the past 15 years and made important contributions to the field (Levitt, Pomerville, & Surace, 2016). However, most studies have focused on the use of interviews *about* psychological therapy, rather than using the interactive content of therapy sessions in itself (Georgaca, 2001; Rennie, 1992; Timulak, 2009). The almost exclusive use of self-reported semi-structured interviews has the potential to be limited by the same criticisms as self-reported process-outcome research. Similar to quantitatively orientated self-report methods, interviewees may be affected by demand characteristics and social desirability biases (Diefenbach, 2009). The direct use of psychological therapy interactions has been used much less frequently, despite much value and emphasis being placed on this type of data in recent years (e.g. Aviram et al., 2016). The use of qualitative methods for instrument development can be valuable because it leads to measures which have been:

- Open to novel findings that may not have been identified in previous research
- Based on and emergent from rich, original, qualitative psychological therapy interaction data specific to the patient group in question.

- Developed using repeated reflection, checking and consensus evaluations for quality enhancement.
- *Informed* by relevant theory, but not *driven* by theory. This leaves room for greater interpretative orientation towards outcome that may not be fully explained or currently understood in specific psychological therapy theories.

Overall, the balance of current evidence suggests that, although qualitative research methods are growing in use within psychological therapy research and making a unique contribution, there are patient groups and qualitative data sources which have not been sufficiently explored. Psychological therapy for high users of healthcare with multimorbidity would be one of these patient groups and direct therapy interactions an under-used data source.

4.1.3 Psychological therapy and high service use

A small group of patients are consistently in the top decile of frequent GP consulters over at least a three year period (Smits, Brouwer, van Weert, Schene, & ter Riet, 2009). Such patients typically have a median of three long-term conditions, two psychiatric conditions and severe health anxiety, often fuelling medical reassurance(Patel et al., 2015; Salkovskis et al., 2003) . Patients within this group rarely seek mental health care and can have negative experiences of health services (Gask, Dowrick, Salmon, Peters, & Morriss, 2011; P. Tyrer, Eilenberg, Fink, Hedman, & Tyrer, 2016). Addressing mental health needs of patients with this kind of multimorbidity can slow disease progression and improve outcomes, whilst ignoring mental health worsens the situation (De Ridder, Geenen, Kuijer, & van Middendorp, 2008; Duarte et al., 2015). Providing appropriate mental healthcare and

consequently making even small reductions in care use, would provide an improvement in overall health and significant gains in primary care capacity.

Psychological therapies such as mindfulness-based cognitive therapy and individual CBT have shown short-term clinical effectiveness but not cost-effectiveness for subgroups of high utilising patients in primary care (Allen, Woolfolk, Escobar, Gara, & Hamer, 2006; van Ravesteijn et al., 2013). This may be because the majority of GP frequent attenders do not consult frequently beyond 12 months (Smits et al., 2009). This means that most frequent attenders return to a normal level of attendance after one year, typically due to a temporary escalation in consultation rates after health problems. However, the small proportion that continue to consult at a high level have much more complex, multimorbid, physical and mental health needs. Only one study has explored the use of psychological therapy to improve health and reduce service use amongst long-term frequent attenders in primary care (Malins et al., 2016; Morriss et al., 2012). In a feasibility case series, individual CBT was offered to long-term high users of primary care at five GP practices, purposively sampled for maximum variation. Results showed that health-related quality of life improved and primary care service use halved over 12-month follow-up. This suggests that targeted psychological intervention holds promise as a means of improving health and creating service capacity amongst high service utilisers with multimorbidity. However, little is known about the specific, interaction-level mechanisms of change or why change was limited in other outcomes.

In summary, moving straight to quantitative methods would be premature due to the novelty of the type of assessment to be developed. In

addition, the experience under study is not currently well understood. Previous research is likely to be relevant to this topic, but adopting an initially inductive qualitative approach makes room for new, unexpected, confirmatory or contradictory findings. This qualitative study aims to use recordings of the CBT sessions from Malins et al.'s (2016) project to delineate the themes of interaction within CBT for long-term frequent attenders in primary care with multimorbidity.

4.2 Method

4.2.1 Thematic Analysis

Thematic analysis was used as the analytic method for this study, because it was the most suitable to address the research question (V. Braun & Clarke, 2006). The flexibility of the approach combined with an adaptable structure made thematic analysis the most appropriate for this study. Furthermore, the method was consistent with the epistemological position of the lead researcher. It has been argued that there are key strands of thematic analysis in most qualitative methods, but drawing them together means that a coherent, trustworthy, yet adaptable set of methods can be rigorously applied to relevant data (V. Braun & Clarke, 2006). In particular, methods within thematic analysis allowed for movement from inductive analytic approaches to deductive, theory informed approaches in a way that most other methods would not.

This approach would be possible within interpretative phenomenological analysis (IPA), which is another well-used qualitative method focused on human experience (Smith, Flowers, & Larkin, 2009). However, IPA has idiographic foundations, which could limit the generalisability of any emerging

findings (Brocki & Wearden, 2006). Therefore, if strictly adhering to IPA's theoretical foundations, IPA results could not be applied more generally as was the hope from this study. Furthermore, IPA's interpretative focus could over-emphasise the researcher's perspective which might not fit with a study aiming to develop a more generalisable assessment tool. McAdams (1993) has also highlighted that the IPA methodology is primarily interview focused and orientated towards the individual's experience. This would not be the best fit for a study aiming to elaborate observations of the interactions between a patient and a therapist.

Thematic analysis was chosen over grounded theory, despite similarities in approach and potential for compatible philosophical positions, because development of new theory was not the primary purpose of this study. Furthermore, iteration between new concepts and existing theory, such as patient activation, would not fit with archetypal ground theory studies wherein existing theory is brought in towards the end of analytic processes (Charmaz, 2014). Lastly, a discursive approach (e.g. Discourse Analysis) would have linked with the aim of this study in endeavouring to "treat people as if they were human beings" (Harré, 1972, p.84). However, the focus on relational discourse – prioritising the contextual influences on narratives between participant and researcher - could detract from understanding the content of patient-therapist's interactions. Moreover, the multiple truths associated with qualitative methods based in a social constructionist epistemology is not consistent with the current approach which aims to explore a presumed underlying truth (Parker, 2005); a penetrative analytic account of this type

needs to privilege a particular reading over another, but this can undermine the discourse analyst's philosophical position.

Notwithstanding these differences in methodology, Reicher (2000) warned against "methodolatry" (p.1) in qualitative research: rigidly adhering to or proscribing one set of methods over others with insufficient consideration of the impact on research quality. This emphasises the need to "respect the reflexive nature of the human subject and organize a methodological stance to reflect that respect" (Blumer, 1969, p.60, cited in Reicher, 2000, p.1). With this in mind, thematic analysis appeared to be the best fit between appropriate, feasible research methods and the subject of study. However, the methods are used as a guide to serve the overarching need to focus upon and adapt to the research aims, data and participants (Parker, 2005).

4.2.2 Epistemological position

A critical realist epistemological position was derived from the research aims. This position acknowledges the significance of beliefs and experiences contributed by researchers and participants, whilst holding that there is a "reality" mediated by these factors (Fairclough, Jessop, & Sayer, 2002). According to this perspective research can lead to an understanding of participants' perspectives, whilst acknowledging that any understanding is influenced by context and other interpersonal and intrapersonal factors. The lack of clarity about the degree to which elements of "reality" can be known could make it difficult to assess the usefulness of results. Nonetheless, critical realism provides a useful grounding point for a thematic analysis aiming to make best use of researcher interpretations whilst incorporating quality assurance methods to account for researcher biases. Alternative

epistemologies were a poorer fit given the research aims to derive some approximation of a “truth”.

4.2.3 Participants

Thirty-two patients were offered CBT in a feasibility and acceptability case series of CBT for long-term frequent consulters in primary care (Malins et al., 2016). Eleven of the 32 participants gave written, informed consent to sessions being recorded and session data being anonymously reported (see Appendix C for participant consent form). All had primary care consultation rates above the 90th centile for the previous two years and experienced multimorbidity, with an average of five chronic health conditions (Table 4.2). As participants presented with a range of difficulties, treatment was adapted to presenting problems, with CBT for health anxiety as the default starting point (Salkovskis et al., 2003). Participants received an average of 13 session, ranging from 4 to 40. For each participant, a true random number generator (https://www.random.org) was used to select one session recording from the earlier half of their therapy (mean session 3), and one from the latter half (mean session 9). Recordings were transcribed verbatim, with identifying information disguised, in total giving a dataset of 22 session transcripts. Researchers who collected outcome data from participants operated independently from those developing the CICS. Researchers developing the CICS were blind to participant outcomes.

Table 4.2. Participant primary care consultation rates and health problems.

ID	Face-to-face consultations in the previous 2 years[a]	Total sessions	Total current morbidities[b]	Morbidity details
Participant 1	46	5	4	dyspepsia, IBS, atrial fibrillation, peripheral vascular disease
Participant 2	27	15	5	depression, GAD, IBS, psoriasis, eating disorder
Participant 3	39	40	5	chronic pain, asthma, CHD, GAD, IBS
Participant 4	28	22	3	depression, GAD, IBS
Participant 5	33	18	3	chronic pain, tendinitis, GAD
Participant 6	35	9	5	hypertension, myofascial pain syndrome, thyroid problems, IBS, glaucoma
Participant 7	28	6	3	hypertension, chronic pain, GAD
Participant 8	26	4	6	hypertension, arthritis, asthma, diabetes, CKD, IBD
Participant 9	60	4	6	hypertension, arthritis, diabetes, IBS, prostrate disorder, IBD
Participant 10	35	5	5	hypertension, asthma, diabetes, arthritis, COPD,
Participant 11	29	11	8	hypertension, depression, diabetes, GAD, IBS, stroke, CKD, blindness
MEAN (SD)	**35 (10)**	**13 (11)**	**5 (2)**	

Note. [a]The average UK primary care patient had 5.5 consultations per year at the time of data collection (Hippisley-Cox & Vinogradova, 2009).
[b]Morbidities calculated through examination of patient medical records. Abbreviations: CHD, coronary heart disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary diseases; GAD, Generalised Anxiety Disorder; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome.
[c]Follow-up sessions facilitating onward referral were not included in session selection. Sessions were selected from the active treatment period.

4.2.4 Therapists

Two psychological therapists provided CBT to participating patients. One was a female CBT therapist with Masters-level training and five years' qualified experience. The other was a male clinical psychologist with two years' doctoral-level qualified experience.

4.2.5 Data storage and management

The data used in this chapter and all following chapters were collected, processed and stored in line with the University of Nottingham's Research Code of Conduct. In compliance with the General Clinical Practice guidelines (Grimes et al., 2005), the Chief Investigator (Professor Richard Morriss) maintained responsibility for records and documents regarding the conduct of the two overall projects from which this thesis emerged. Data will be retained for at least 7 years or for longer if required. If the responsible investigator is no longer able to maintain the study records, a second person will be nominated to take over this responsibility. The study documents held by the Chief Investigator shall be finally archived at secure archive facilities at the University of Nottingham. This archive shall include all study databases and associated meta-data encryption codes.

All identifiable patient data was removed from research data at recruitment and replaced with a unique identifier. Personal data was stored separately from research data and an encrypted identification key was kept and only accessible to the Programme Manager (Dr Catherine Kaylor-Hughes). Data from both source projects were monitored through confirmation of informed consent; source data verification; data storage and data transfer procedures; local quality control checks and procedures, back-up and disaster

recovery of any local databases and validation of data manipulation. These were carried out by study researchers and the Programme. The lead researcher (Miss Shireen Patel) checked the data quality assurance procedures at each site annually. Study data and evidence of monitoring and systems audits were made available for inspection by the approving Research Ethics Committee, the relevant Trust research and development department and a representative of the Sponsor organisation (the University of Nottingham) as required.

Ethical approval was granted by Nottingham 1 National Research Ethics Service Committee (11/EM/0392).

4.2.6 Procedure

Given the staged nature of the procedure and the numerous measures taken to ensure transparency of approach, elements of the results are present at each stage to illustrate how the analysis progressed. The overall results are then presented in the results section.

A number of quality assurance measures were used at each stage of the analysis to facilitate a trustworthy and credible analysis (Elliott, Fischer, & Rennie, 1999; Nowell et al., 2017). The stages of analysis and associated quality assurance methods at each stage are described in Table 4.3. Reflective journals were kept by thematic analysts to help them consider how existing knowledge and beliefs might influence their reading of transcript data. For example, both thematic analysts were trained psychological therapists so when attempting to identify interaction themes they noted the urge to label interactions using psychotherapeutic terminology. In particular, they initially described “problem analysis” interactions as “formulation” interactions.

Reflecting on this type of influence (using their journals at regular meetings) helped them to consciously set aside biases and, as best they could, engage with the data more openly. As suggested in guidance for trustworthy thematic analysis, diagrams of emerging theme structures were developed at each meeting and reflective notes were made on each case discussed (see Appendix D for examples of emergent theme diagrams and Appendix E for examples of reflective notations followed by an example of overall reflections from each coder).

Multiple coding and researcher triangulation added quality assurance to the themes developed, as similar forms of major themes were present in both coders' analyses and differences were resolved through critical discussion. Coders maintained an auditable process to confirm that emerging themes could be tracked back to the original data. This ensured that all analysis remained grounded in the source data. External credibility and trustworthiness was added through regular consultation with two PhD supervisors who were not involved in the analysis, but commented on theme development and transparency. This was extended in the stages described in the next chapter, where external researchers new to the whole project were included and their comments contributed to theme adaptations. Developing a report of the analysis helped achieve consensus amongst researchers, clarify the audit trail and make final refinements to themes in line with inconsistencies identified.

Table 4.3 Phases of Thematic Analysis (V. Braun & Clarke, 2006) and quality assurance methods used at each stage (adapted from Elliott et al., 1999; Nowell et al., 2017)

Phase	Description of the Process	Quality Assurance Methods To Establish Trustworthiness
1. Familiarising yourself with your data:	Reading and rereading the data, noting down initial ideas.	*Prolonged engagement with data* - ensuring sufficient time is allowed for detailed knowledge of the data. *Reflexive journaling* - Document thoughts about potential codes and associated theories and reflections.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.	*Researcher triangulation* - Different researchers working independently on different parts of the data set and then triangulating their analyses through discussion and review of the data. The convergent areas of two viewpoints are more likely to have a meaningful relevance to those outside the study. *Reflexive journaling* - Maintaining a journal of reflections from the analytic process helps to account for the researcher's impact on the study and results.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.	*Providing an auditable process* - ensuring that the identification and development of all themes and their place in theme structures can be traced back to source data from which it is derived in a coherent, auditable and accountable fashion. *Documentation of all researcher meetings.*
4. Reviewing themes:	Checking in the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic “map” of the analysis.	*Researcher triangulation.* *Diagramming to make sense of connections between themes.* *Keeping detailed notes about theme development and hierarchies.* *Peer debriefing* - Themes vetted by external researchers (PhD supervisors).
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells; generating clear definitions and names for each theme.	*Researcher triangulation.* *Peer debriefing.*
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.	*Providing an auditable process.* *Researcher triangulation.* *Peer debriefing.* *Report on reasons for theoretical and methodological choices.*

Note: Italics denote quality assurance methods.

A two-stage thematic analysis was conducted using an adapted version of V. Braun and Clarke's (2006) method. In the first stage of analysis, each of two thematic analysts took half of the sample (five participants and 11 transcripts each with one participant shared) and completed a descriptive, inductive thematic analysis. This open approach aimed to identify "what is talked about in this session and how?" Having each looked at different parts of the dataset, similarities and differences in emerging themes were used to develop a theme template. Nine themes were developed by movement from initial descriptive notations, through 49 initial theme categorisations, until transcript contents could be described in a smaller set of 20 cluster themes and nine overall superordinate themes (Table 4.4 shows an example audit trail for the structuring and task focused theme; Table 4.5 shows a summary of all themes developed from initial themes through cluster themes to superordinate themes). A joint thematic model of interaction themes was refined through discussion and review between the two thematic analysts. Thematic analysis continued until a set of superordinate themes were established with each represented in at least half of the transcripts analysed, with coherence amongst themes.

Table 4.4. Example audit trail for the Structuring and Task Focus theme development.

Transcript Text	**Initial Notations Summary**	**Initial Theme**	**Cluster Theme**	**Superordinate Theme**
T: OK. Well let's go back to the pain then. So how has the pain been over the last 2 weeks? P: Hasn't been too bad (Participant 11)	moving discussion on to important issues	Directing/Guiding	Structuring	Structuring and Task Focus
T: So back to your diverticulitis, you said you were diagnosed about a year or so ago did you say? P: About two years ago. (Participant 10)	Directing conversation, avoiding deviations			
T: Okay. Okay. I tell you what we'll do. We will do this bit and then I will give you this bit to take home with you…. P: Okay. (Participant 3)	Being explicit about movement within the session, consent from participant	Explicit task shift	Task Focusing	
T: I guess if we go back to those top ones just kind of pin them down a little bit more. [ ] P: I think we're at the beginning stages kind of [ ] accepting them when they're happening. (Participant 2)	Subtle refocus on therapeutic task, positive response from participant	Implicit task shift		

Note. Abbreviations: P, Patient; T, Therapist.

Table 4.5. Summary of the overall theme template

Initial Theme	Cluster Theme	Superordinate Theme
Expectation management Agreeing goals Refining goals Committing to plans	Action planning	Action Planning and Idea generation
Exploring possible plans considering choices and ideas Addressing barriers	Idea generation	
Self-supporting statements Self-critical statements	Client evaluations of self	Evaluations of Self or Therapy
Encouragements and affirmations	Therapist evaluations of client	
Valuing previous interactions Therapy meta-comments Uncertainty about therapy	Evaluations of therapy	
Explaining/providing rationale Correcting Normalising Advising Persuading Reassurance	Information-based interventions	Information Discussion
Requesting/seeking information Checking out understanding	Information exploration and discussion	
Benefits/improvements from intervention Noticing difference	Noticing positive change	Noticing Change or Otherwise
Deterioration/Losses No change in spite of efforts Feeling stuck	Noticing negative/no change	
Identifying problem causes Analysing problems Clarifying the problem	Problem analysis	Problem Analysis and Understanding
Linking problems and solutions	Linking problems and solutions	
Exploring alternative perspectives New understanding	New understanding of problem	
Problem description Problem assessment	Problem description without analysis	Problem and Context Description
General enquiries Problem context description General talk	Description of context around problems	
Directing/guiding	Structuring	Structuring and Task Focus
Explicit task shift Implicit task shift	Task focusing	
Mutual agreement Humour Joint idea building Rapport	Collaboration	Collaborative Flow
Flow Engagement	Flow	
Session transitions Integration of content	Integrated session movements	Integration of Themes
Content coverage	Coverage of themes	

In a second stage, theory-informed, outcome-orientated, deductive analyses of established themes was carried out. Whilst blind to outcomes, the two thematic analysts reanalysed and grouped data from each super-ordinate theme into three types of theme presentation:

1. Those deemed likely to contribute to therapeutic improvement
2. Those deemed likely to contribute to non-improvement or deterioration
3. Those where no observable distinction could be made

Patient activation theory was used to decide which types of interaction were likely to lead to therapeutic improvement or otherwise. In particular, patient activation theory suggests that better outcomes will ensue when someone is more engaged and confident in their ability to manage their health (Hibbard & Gilburt, 2014). Therefore, where there were observable, verbal indications of engagement with, confidence in and/or facilitation of health improvement, these were categorised positively:

> **Therapist:** Is it [ ], like, this is it, you know, this is just life, it's always like this?
>
> **Patient:** No, I ought to make more of an effort to do my exercises and look after my back. I have been looking in my pain management book recently and thinking, yes I must do that...
>
> [Coded: Action planning idea generation +1; Participant 5]

Conversely, where there were observable, verbal indications of reluctance to engage and/or low confidence about health management, these were categorised negatively:

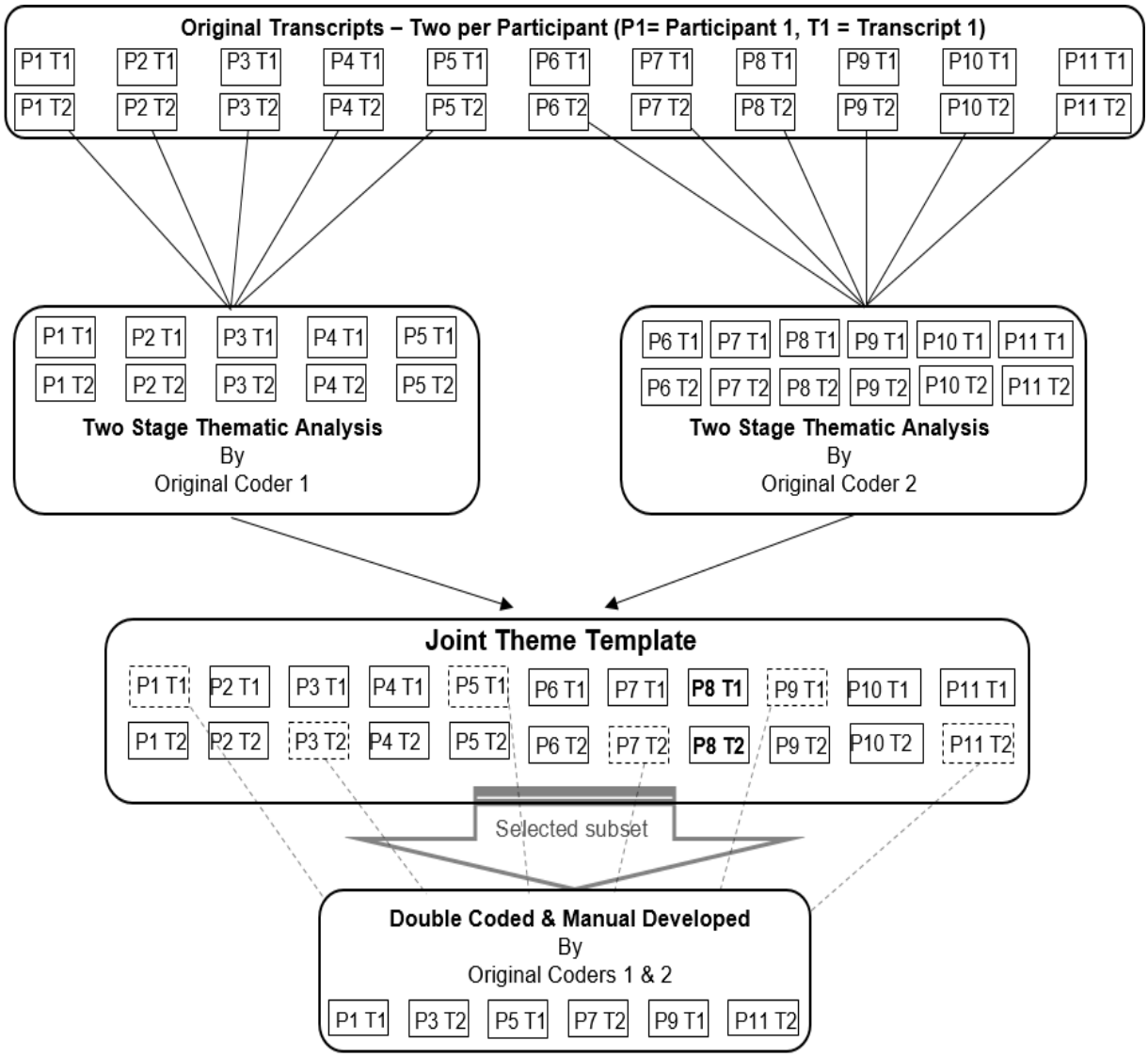
> **Therapist:** I think maybe the starting point might be to get a better understanding of what's going on for [your daughter]. Start off with what she thinks and work from there [ ].

Patient: I've tried all that as well [ ] I've tried so many approaches and there's just nothing seems to shift her.

[Coded: Action planning idea generation -1; Participant 4]

The results of this analysis were used to develop a coding manual, so that the methods used could be replicated by new and independent coders. A subset of transcripts were then double-coded as an initial assessment of IRR (Figure 4.1 depicts the coding process).

Figure 4.1. The coding process.



4.3 Results

4.3.1 Thematic analysis

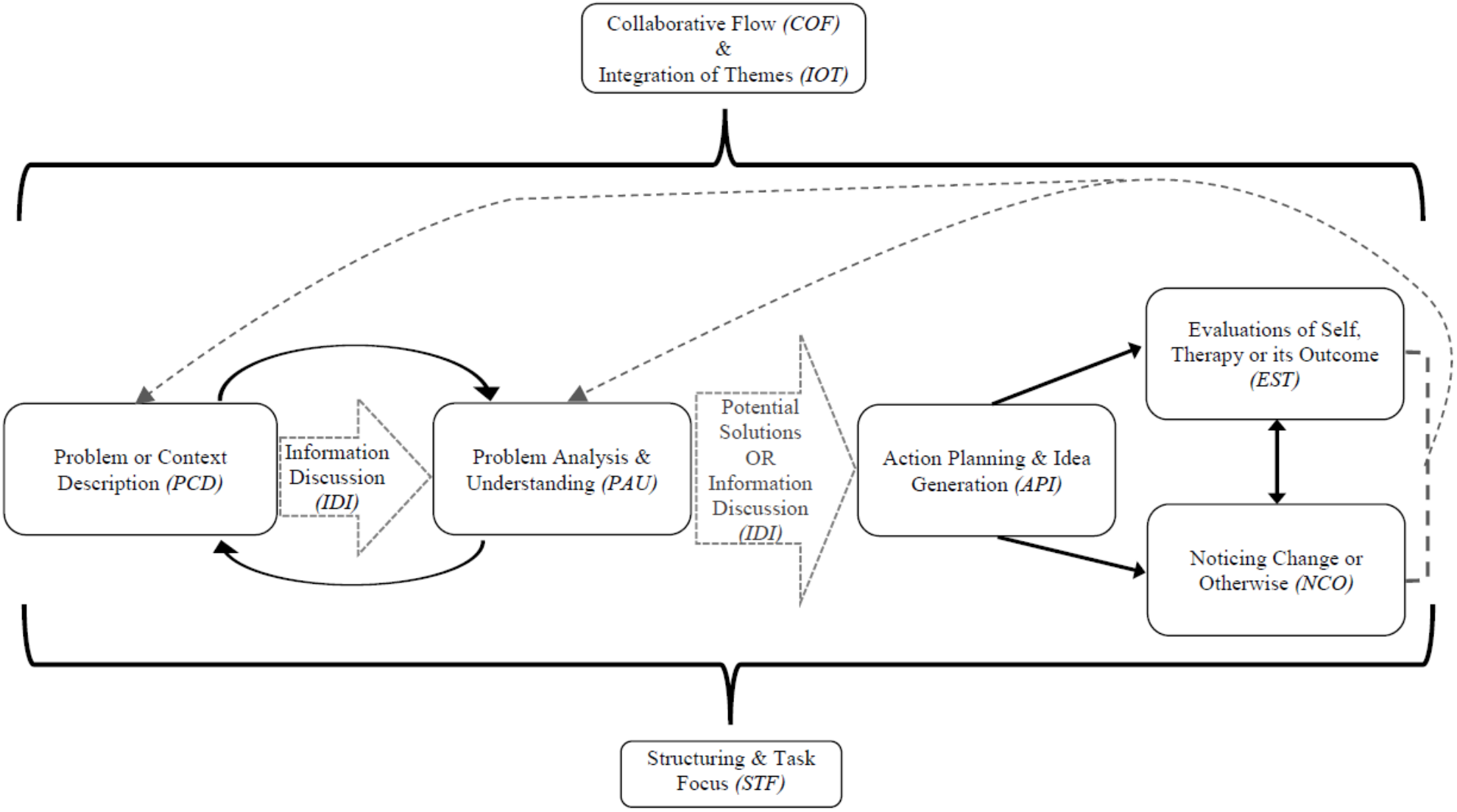
From the initial inductive thematic analysis, a set of super-ordinate interaction themes emerged, describing topics discussed within and across sessions. These themes were organised into an interactive process model, described below, and depicted in Figure 4.2.

In terms of interaction contents, the therapy conversation process was seen to typically begin with descriptions of problems or their context (*Problem or context description: PCD),* in which therapist and client delineated the nature of a difficulty, the context in which it occurred, and/or its impact. If session discussions moved on from *PCD*, an analysis of problems often ensued *(Problem analysis and understanding: PAU)*, sometimes mediated by relevant *Information discussion (IDI)*. Identification of potential solutions or related information could lead the conversation into action plans for change or ideas about what could be done to improve the situation *(Action planning and idea generation: API).* Feedback from action plans and ideas was sometimes given, typically after they had been tested out in the session or between sessions. Feedback included *Evaluations of the self or therapy (EST)* alongside reports of change or the absence of change from therapeutic interventions *(Noticing change or otherwise: NCO).* Feedback from action plans also informed the next cycle of the conversation process – for example the changes noticed could lead to a conversation about a new understanding reached *(PAU)* and consequent action plans *(API).* Conversely, if no change was observed, continuous description of related problems could ensue *(PCD)* or an analysis of the failure might take place *(PAU).* The degree of joint

structuring, focusing and task choice in these movements between themes was identified as an anchoring theme present at specific interactions across the session *(Structuring and task focus: STF).*

In global terms, the degree to which the therapy conversation integrated and moved naturally between multiple themes seemed important to the overall flow of the session *(Integration of themes: IOT).* The mutuality of task and idea development between therapist and client also appeared to have an important impact on how themes were expressed *(Collaborative flow: COF).*

Figure 4.2. Process model from thematic analysis.



4.3.2 Moving from thematic analysis to interaction coding instrument

Thematic analysis resulted in a set of super-ordinate themes with subthemes categorised into those contributing to: (1) improvement, (2) non-improvement, or (3) unable to discern. Therapist-patient interactions were categorised in this way based on the level of patient activation observed. Coding methods were aggregated into a manual named the Consultation Interactions Coding Scheme *(CICS; since the completion of the study described, the CICS has been made publically available online: Malins et al., 2018; The CICS manual is also available in Appendix F)*. The CICS provides clear descriptors for rating patient activation within each interaction – rather than global session scores alone.

4.3.3 Coding Scores

Each super-ordinate theme was sub-categorised in terms of hypothesised contribution to outcome. That is, lower scores were hypothesised to link with poorer outcome and vice versa. Ratings were assigned to one of five ordered levels, ranging in valence and magnitude: from -2 for interactions deemed strongly detrimental to outcome, to +2 indicating interactions deemed strongly beneficial to outcome with a midpoint of 0 denoting neutral contribution to outcome. Level descriptors were characterised in a similar way across themes, broadly mapping on to the levels of patient activation identified in the PAM. Level descriptors for the CICS and related PAM level descriptors are summarised below in Table 4.6. The specific level descriptors for the problem analysis and understanding theme are presented in Table 4.7 with exemplar quotes. A full set of exemplar quotes for all themes is presented in Table 4.8.

Table 4.6. CICS coding scores and equivalent, mapped patient activation levels (adapted from Deeny, Thorlby & Steventon, 2018)

CICS level	CICS level descriptor	Mapped PAM level	PAM level descriptor and percentage of patients at each level[a]
+2	Hypothesised to be interactions most predictive of positive outcome. Key attributes are a high level of patient activation and focus; an interaction usually led by the patient. This would include patient initiated therapeutic activity not cued or primed by the previous therapist interaction.	Level 4	*Level descriptor: "I'm my own advocate".* Patients have adopted behaviours and practices to manage their condition, such as good medication adherence, care planning or self-monitoring. Such individuals may be accessing support peers and their community. Patients may still need extra support, for example to recover from a relapse or a hospital procedure (13% of respondents).
+1	Hypothesised to be a positive interaction, with regard to outcome, but with a lower level of patient activation and leadership. Typically, this would be a therapeutically active interaction, led or guided by the therapist, which the patient endorses and/or develops with their contributions.	Level 3	*Level descriptor: "I'm part of my healthcare team."* Patients who appear to be taking action, for example setting goals for their health (such as adhering to a medically-advised diet) or creating a care plan with health care providers, but may still lack the confidence and skill to maintain these (46% of respondents).
0/Neutral	These are interactions where few or no observable positive or negative interactive factors are apparent and is deemed to be neutral – neither beneficial nor detrimental to outcome. The same is coded if the theme is absent. This includes interactions where therapists make suggestions or comments with little or no sense of how the patient receives these.	N/A	
-1	Hypothesised to be therapeutically unhelpful interactions in a minor way. This includes interactions which show the start of unaddressed disagreements or reluctance to engage with therapeutic activities. Low levels of patient activation and involvement are observed.	Level 2	*Level descriptor: "I could be doing more"* Patients who may manage some low level aspects of their health (for example, take their medication) but struggle in other aspects of their care, such as engaging with care-planning. (19% of respondents).
-2	Hypothesised to be interactions that would be contradictory to most therapeutic guidance. This would include argumentative or obstructive interactions where the patient and potentially the therapist appear disengaged, unfocused, and/or oppositional to therapeutic activity.	Level 1	*Level descriptor: "My clinician is in charge of my health."* Patients tend to feel overwhelmed by managing their own health, and may not feel able to take an active role in their own care. They may not understand what they can do to manage their condition better, and may not take their medication, attend preventative appointments, or see the link between healthy behaviours and good management of their condition. (22% of respondents)

Note: [a]Data taken from a UK sample of 9,348 primary care patients (Deeny et al., 2018) CICS: Consultation interaction coding scheme; PAM: Patient activation measure

Table 4.7. Problem analysis and understanding coding level descriptors and exemplar quotes.

Level	Level Descriptor[a]	Exemplar Quote
+2	*Patient led* articulation of a new understanding with links between components of problems, where the solutions may be clearer (e.g. I've realised; I now see). There is some evidence that therapeutic concepts are understood. Endorsed or encouraged by therapist with development.	**P**: I think as well because I think I am thinking differently, I'm not feeling as much pain. **T**: Really, right. **P**: I think I would have brought more on–because I get myself all tensed up and everything. [Participant 7]
+1	Therapist presents the above or jointly with patient with some clarity as to what the solutions may be and *it is developed, summarised or corrected* by the patient OR patient-led new understanding with *limited clarity e*ndorsed or encouraged by therapist.	**T**: The fear isn't realised when you don't do [the compensatory strategy] and you feel more confident and [ ] satisfied with yourself. **P**: I think I've always tried very, very hard to [ ] somehow change people's perception of me. Whereas I think I've come to realise that maybe the people that I was aspiring for to like me and to have those friendships with aren't the people that [I] probably naturally should be running after. [Participant 2]
0	Therapist or patient presents a problem analysis of some apparent therapeutic value and this is acknowledged by the other with *no development*	**T**: Where you've stayed with it long enough, that the anxiety starts to drop. That's, that's how you know. **P**: Okay, okay. [Participant 3]
-1	Therapist presents a problem analysis and the patient displays *ambivalence in their response* without explicit correction. OR the patient presents a problem analysis with the potential to be counterproductive.	**T**: I: So with those few factors in mind, your son staying at home during this week [ ], you're having your breakfast later and a different breakfast during this week and earlier [ ] and then I guess us meeting and talking about pain and testing it out. Are there any of those things…what of any of those things might have…? **P**: And I think also I mentioned to you last week my family have wanted me to have a wheelchair. [Participant 6]
-2	Therapist presents a formulation and this is refuted but no alternative is given by the client and no agreement is reached OR the client presents an analysis that is anti-therapeutic and this is not opened for correction by the therapist.	**T**: While you have been drinking more though, you've had one night of full sleep. So what does that say? **P**: Yeah, mm, you're trying to say your system works and it doesn't. [Participant 9]

Note. [a]Italics add emphasis to the key component of the level descriptor; Abbreviations: P: Patient; T: Therapist.

Table 4.8: Exemplar quotes with corresponding CICS themes and level coded.

Theme	Level	Exemplar Quote
Action planning and idea generation **(API)**	+2	**T**: is it [ ], like, this is it, you know, this is just life, it's always like this? **P**: No, I ought to make more of an effort to do my exercises and look after my back. I have been looking in my pain management book recently and thinking, yes I must do that... [Participant 5]
	+1	**T**: Yeah, what we did. So kind of like, we did about 12 minutes, so maybe a little bit more than what we just did, if you can do that each day. Like a gradual building process. **P**: If I get up 10 minutes early and do it in the morning, it could set me up for the day really couldn't it? [Participant 4]
	0	**T**: There's 2 key bits to it. One about, kind of, today and tomorrow, planning ahead what's going to make life easier, what's going to be kind of pacing it for the weekend. **P**: OK, right. [Participant 11]
	-1	**T**: So is this all about keeping mobile? **P**: I can't walk far so I'll have to have a taxi. Can't do it. [Participant 8]
	-2	**T**: Do you ever use kind of warmth or anything? What did the pain clinic say? **P**: This is another thing. I just cannot get warm. I sent for (Dr Smith), it was last year and I says my feet and my legs are frozen [Participant 8]
Evaluations of self or therapy **(EST)**	+2	**P**: My initial feeling was I'm pleased with myself for being honest with myself [ ] you know feeling, this is ironic! Confident enough to be honest with them about how I'm feeling. **T**: Why is that ironic? [Participant 2]
	+1	**T**: Mmm, how [has] it been particularly off the back of last time? **P**: I mean, this has been very helpful, very useful, it's been good. [Participant 5]
	0	**T**: you have been brave [ ] those are really big steps and really valuable steps [ ] and that is what you're telling yourself every time you [ ] let water go on your face a little bit more [ ]. Did you get round to kind of doing [a diary]? **P**: Well, I didn't... [Participant 7]
	-1	**P**: Back to square one aren't we. **T**: Is that what it feels like to you? [Participant 6]
	-2	**P**: I just don't think I've done anything right in life, I really don't. I don't see my point being here at all. And I keep going to see [my GP] and I feel like a hypochondriac. [ ] I just feel like nobody needs me. **T**: Yeah. [Participant 4]

Information Discussion **(IDI)**	+2	**P**: It would be really helpful if we could talk about how panic attacks work like you mentioned last time, so I know what I'm dealing with. **T**: Yeah, great idea. Shall we look at a recent example? *[Hypothesised example]*
	+1	**T**: What you think about, how you feel physically and emotionally [ ] and what you do are all linked [ ] so for example if you kept on making time for yourself [ ] over time you will feel less guilty, [ ] you will have less of those thoughts of feeling selfish [ ]. **P**: I need to practise that then, that's the first step then… [Participant 3]
	0	**T**: In a kind of very basic formulation that we do in cognitive behavioural therapy, so it can be thoughts about the situation, OK, physiology is about body sensations, how you feel, behaviour basically is what you do, OK, and emotion is how you feel. [ ] Is that OK? **P**: Uh-huh. [Participant 1]
	-1	**T**: You can also get headaches with dehydration. So it might be possible that you know with your - **P**: That would be a miracle then. [Participant 9]
	-2	**T**: But it is still good to drink good amounts, 'cause it's good for your kidneys, it's good for your brain. **P**: I haven't got a brain so I. [Laughter] My eyes are back to normal. I see them again, soon. We've got it all sorted out now. You know that funny letter I wrote... [Participant 9]
Noticing change or otherwise (**NCO)**	+2	**P**: I have noticed this week actually that when I've been ill and been coughing a lot and not been able to breathe I got very anxious. And I thought well there's no need to do this, you know it's not life threatening, it's nothing really bad. And I sort of talk myself down again. **T**: Good, good. More self-aware about it. [Participant 10]
	+1	**T**: Yeah, and you are saying like there is a drop in the pain actually after [the graded pain killer medication reduction]? **P**: Yes. As I say it did improve. [Participant 6]
	0	**T**: Yeah, so [your sense of achievement] was feeling different then anyway. **P**: Yeah. [Participant 11]
	-1	**T**: Well, aside from that how have things been? **P**: Oh, not too good I don't think in the fact that I have been in a lot of pain, I mean the worst it's ever been [Participant 7]
	-2	**P**: So I keep pushing myself and pushing myself and pushing myself but I'm getting now that it's getting too much to push myself. **T**: Ahh. What is it about that then? [Participant 8]

Problem analysis and understanding **(PAU)**	+2	**P**: I think as well because I think I am thinking differently, I'm not feeling as much pain. **T**: Really, right. **P**: I think I would have brought more on – because I get myself all tensed up and everything. [Participant 7]
	+1	**T**: Evaluating how worthwhile that the compensatory strategy is [ ] the fear isn't realised when you don't do it and you feel more confident and [ ] satisfied with yourself. **P**: I think I've always tried very, very hard to [ ] somehow change people's perception of me. Whereas I think I've come to realise that maybe the people that I was aspiring for to like me and to have those friendships with aren't the people that are probably naturally should be running after. [Participant 2]
	0	**T**: Where you've stayed with it long enough, that the anxiety starts to drop. That's, that's how you know. **P**: Okay, okay. [Participant 3]
	-1	**T:** I: So with those few factors in mind, your son staying at home during this week [ ], you're having your breakfast later and a different breakfast during this week and earlier [ ] and then I guess us meeting and talking about pain and testing it out. Are there any of those things…what of any of those things might have…? **P:** And I think also I mentioned to you last week my family have wanted me to have a wheelchair. [Participant 6]
	-2	**T**: While you have been drinking more though, you've had one night of full sleep. So what does that say? **P**: Yeah, mm, you're trying to say your system works and it doesn't. [Participant 9]
Structuring and task focus **(STF)**	+2	**P**: I also want to look at the goal setting thing as well, because I think that might be useful for me. **T**: OK. [Participant 5]
	+1	**T**: I was thinking it [ ] might be useful to have some kind of goal, general goals about (P: Yeah, definitely) what we're aiming for and that would be ok [ ]. **P**: Ok. [ ] **T**: We talked a bit about mindfulness [to] do that as well [ ] about half and half in terms [ ] of time. Is there anything [ ] other than that, that would you like to talk about? **P**: No that sounds pretty much probably what I was kind of expecting so that's cool. [Participant 2]
	0	**T**: OK, well what we were going to do today if that's OK as part of our session is kind of walk through the process that you have for dealing with your thrush, what you do to try to make it better and then when do you decide to go to the GP? So is that OK? **P**: Uh-huh. [Participant 1]
	-1	**T**: What did you go and see? **P**: Oh, the Woman in Black. **T**: Oh, that's quite scary. **P**: Really scary, yes, but I do like films that keep me on my toes. [Participant 3]
	-2	**P**: I give the neighbours a smoke signal [ ] I've got an incinerator – I give them a bit of smoke. I wait till they get their washing in – yeah. **T**: Oh well that's nice at least you will wait for the washing. **P**: Well yeah, you know the legal on that: there is no set time when you can have a bonfire as long as it isn't causing any danger. [Participant 9]

The problem and context description (PCD) theme was the exception to the above-defined scoring system. Interactions where problems were described without reflections, analysis or evaluations (which would be coded problem analysis or evaluations of self or therapy) were deemed to neither predict positive nor negative outcome. It was inferred that, as a neutral theme, *PCD* could become problematic if larger proportions of the session were dedicated to problem description, because it would give less opportunity and space for interactions where patient activation could be more readily expressed and observed. Therefore, *PCD* was hypothesised to be a negative predictor of outcome and deemed to be either present or absent rather than rated on a scale:

> **Therapist**: OK. And how did that day go in terms of like pain?
> **Patient**: My lower back was hurting because I was sitting up to do this searching for the papers.
>
> [Participant 11, *PCD* present]

> **Therapist**: Mmm, how [has] it been particularly off the back of last time?
> **Patient**: I mean, this has been very helpful, very useful, it's been good.
>
> [Participant 5, *PCD* absent; *EST* +1 present]

Global themes Collaborative Flow (COF) and Integration of Themes (IOT) were overall, single ratings for each session. This was because they were seen to illustrate important themes that characterised interaction qualities across several/all interactions.

4.4 Discussion

This chapter has described the initial development of the CICS using a two stage thematic analysis. A hierarchical theme template was established

through inductive thematic analysis of 22 CBT sessions with 11 long-term high users of healthcare with multimorbidity. A set of six themes were developed and sub-categorised turn-by-turn using criteria based on patient activation level (Action planning and idea generation; Evaluation of self or therapy; Information discussion; Noticing change or otherwise; Problem analysis and understanding, and structuring and task focus). Two additional themes were rated on the same criteria, but across the whole session (Integration of themes and Collaborative flow). One other theme was developed and categorised on the basis of “presence” or “absence” (problem and context description). It is notable that agreement could not be reached across the research team on a way to code the problem and context description theme using patient activation levels. It was agreed that in interactions where low levels of patient activation were observable during patient descriptions of their problems it was more appropriately coded as a low level of the structuring and task focus theme. Alternatively, descriptions of problems and contexts where there were observable signs of high patient activation were best coded as either high levels of problem analysis and understanding or high levels of action planning. Therefore, problem descriptions were deemed neutral events in terms of patient activation, but large amounts of problem description would be problematic through a knock on effect to other themes.

It is also noteworthy that two global, cross-cutting themes still emerged (Collaborative Flow and Integration of Themes), despite the explicit focus on interaction-level ratings. Thematic analysts agreed that these were important superordinate themes, but could not be captured adequately by ratings of individual speech turns. This experience perhaps helps to explain why most

process-outcome research has focused on global scores of whole sessions (Crits-Christoph, Connolly Gibbons, & Mukherjee, 2013). The relative value of these themes to outcome will be initially evaluated in the following chapter.

The next chapter describes the initial testing of the coding instrument developed in this chapter and uses the same dataset. This will entail introducing and training novice coders. Then IRR for new coders will be assessed. This will be followed by evaluation of convergent validity with two established instruments, broadly assessing interaction qualities in psychological therapy: the observer-rated working alliance inventory (Horvath & Greenberg, 1989) and the CTS-R (Blackburn et al., 2001). Then associations between 6-month follow-up outcome scores and coding scores will be assessed. Lastly, an extreme case comparison is presented to illustrate the similarities and differences between the highest and lowest scorers on the CICS.

5. DEVELOPMENT OF THE CICS: RELIABILITY AND VALIDITY

5.1 Introduction

This chapter describes the assessment of reliability, convergent validity and clinical utility for the CICS. Having developed the CICS using varied qualitative methods, this chapter describes the quantification of the CICS' reliability and validity. Established research methods are used alongside standard benchmarks to describe the degree of agreement achieved in each assessment (Cicchetti, 1994; Landis & Koch, 1977; Pituch & Stevens, 2015). Conclusions are then drawn on the relative strengths and weaknesses of the CICS and ultimately, whether further investigation is warranted.

There are several well-used methods to assess the reliability and validity of new psychometric instruments. In terms of reliability for observational instruments, IRR is a central assessment: the degree to which different, independent raters using the same instrument would rate the same observational data in the same, similar and/or consistent ways (Hallgren, 2012). This measure helps to establish the dependability of an assessment. Greater correlations amongst raters indicate that there is a higher level of agreement on what is being measured and a lower level of measurement error or noise (Novick, 1966). However, IRR can say little about *what* is measured, a measure's validity – the degree to which an assessment measures what it claims is the construct under study (Pituch & Stevens, 2015). There are a number of tests of validity. Assessment of convergent validity is a common and important means of establishing the validity of a new instrument, but it cannot establish validity alone (Messick, 1995). Convergent validity was sought in this study, because convergence with related measures is an important first step in

evaluating the relevance of a new instrument. Predictive validity was also sought, because the CICS aimed to have clinical utility in aiding therapist responsiveness, by identifying interactions that best predict outcome. Convergent validity compares the relationship between ratings on a new instrument with those achieved using an established instrument which assesses a similar construct. Close relationships between scores on the new instrument and those of the established instrument suggest that there is convergence in the constructs assessed. Inter-rater reliability and convergent validity are assessed for the CICS in this chapter. Initial assessment of predictive validity is also conducted, using correlations between CICS scores and residualised change in clinical outcome scores.

In recent years greater emphasis has been place on outcomes in psychological therapy (Wampold & Imel, 2015). In particular, this refers to psychological therapy practices and process measures that are similarly orientated towards improvement in outcomes. Therefore, high value is placed on the prediction of outcomes from processes occurring within psychological therapy (Grant et al., 2014; Lambert, 2017; Lorenzo-Luaces, German, & DeRubeis, 2015). A measure with strong predictive validity is deemed to have clinical utility – meaning that it has a clinical use for adapting psychological therapy in line with processes identified. This kind of predictive validity can be assessed in a number of ways. Historically, the most common method has been to correlate scores on the process measure with changes in validated, important outcome measures relevant to both the therapy and the process (e.g. Paivio & Bahr, 1998). Although multiple regression and hierarchical linear modelling methods have somewhat replaced simple outcome correlations

recently, associations with outcome remain an important initial step in evidencing clinical utility (Luborsky, Crits-Christoph, Mintz, & Auerbach, 1988).

The potential value of the CICS and the unique contribution it could make as an assessment of individual psychological therapy interactions is illustrated when comparing the CICS with an established psychological therapy process measure. The observer-rated working alliance inventory (Horvath & Greenberg, 1989) is the most commonly used observational assessment of psychological therapy interactions in research. Scores are a robust predictor of clinical outcomes, though the effect can be smaller for earlier sessions than later sessions and there is significant variability between study results (Flückiger et al., 2018). As with most process measures, the working alliance inventory provides an overview score of the session as a whole. However, within-session changes can significantly affect outcome and might not be picked up by global-level assessment (Safran et al., 2011). Furthermore, summarising many exchanges with a single score does not allow for the recent finding that the same type of interaction may be deemed helpful or unhelpful at different points in a session (Swift, Tompkins, & Parkin, 2017).

Taken together, this means that summary scores alone - like those gained from the working alliance inventory and most other psychological therapy process measures - may not be sensitive enough to differentiate specific interaction-types that require corrective attention for therapists to improve their practice. Furthermore, the reliability of the working alliance inventory has only been assessed for overall scores of either whole sessions or session segments. So little dependable comment could be made about smaller, specific groups of in-session interactions. Therefore, assessments

based on single ratings for whole sessions are unlikely to provide the level of detail required to fully contribute to the improvement of therapist interaction-level skills. This strengthens the case for reliable and valid interaction-level assessment of psychological therapy sessions with the CICS as the current candidate.

By contrast to most psychological therapy process measures, the CICS aims to reliably evaluate each interaction in therapy, with categories and scores allocated for each turn of speech. If found to hold clinical utility, this would mean that the CICS could offer therapists with sufficiently detailed feedback to inform them of particular interaction-types that require corrective practice.

Motivational Interviewing is a therapeutic modality that has specifically addressed the issue of detailed interaction-level feedback to enhance clinical skill (W. R. Miller & Rollnick, 2012). Motivational Interviewing skills assessments address the need for detailed assessment of in-session interactions through ratings of each therapist and patient utterance on the basis of CT/CCT (for patients) and specific MI skills such as complex reflections (for therapists). However, these assessments still rely on total behaviour counts over a session and cannot differentiate whether specific interactions enhanced motivation or not. For example, a complex reflection is a therapist response which is consistent with the MI approach and so would be rated positively, but is unlikely to always be the most effective intervention – much relies on the response it receives. The most recent version of the MI treatment integrity scale (Moyers et al., 2016) attempts to address this issue by adding ratings of therapist responsiveness to patient language, but these are

only rated on global scales not at the more sensitive interaction-level. The CICS is consistent with this approach and has the potential to enhance it, through attention to the dyadic nature of patient activation. This approach suggests that the interplay between therapist and patient speech is important in determining the clinical effectiveness of the conversation. The concept is illustrated in several study findings which indicate that there can be as much or more outcome variability *within* therapists' caseloads as *between* therapist effects (e.g. Imel, Baer, Martino, Ball, & Carroll, 2011), which suggests an interaction between therapist and patient effects. Given the importance placed on in-session speech for outcome (see Chapter 2), it is possible that in-session speech may be an important part of the interface between therapist and patient effects.

Overall, the CICS aims to assess whether the combined effect of therapist interventions and patient responses leads to increased patient activation, at specific in-session interactions as well as total scores. This approach accounts for the same type of interaction being viewed as helpful or unhelpful at different points in a session (Swift et al., 2017), which is not addressed by total behaviour counts or global scores. In this way, the CICS allows therapists to gain feedback on their own interaction patterns and responses received, which are helpful or hindering in given situations with specific patients. Crucially, problematic characteristics in this type of over-learned, habitual behaviour are likely to go unnoticed without feedback (Ericsson, 2009; Kahneman, 2011; Kahneman & Klein, 2009).

The type of mixed methods applied in the development of the CICS (informed by the qualitative results of the previous chapter) can use the

strengths of both qualitative and quantitative methods to answer important health research questions (Shorten & Smith, 2017). Furthermore, mixed methods can illuminate findings that neither qualitative nor quantitative research could reveal alone (McKim, 2017). However, mixed methods require close integration of different research modalities in order to achieve the potential benefits (Johnson, Onwuegbuzie, & Turner, 2007). As such, this chapter aims to assess whether the themes emerging from qualitative analyses can be reliably rated by novices (that is CBT therapists who had no prior knowledge of the CICS or related concepts), are associated with established assessments of psychological therapy process and have a quantifiable value in relation to outcome.

This study aimed to establish the IRR and convergent validity of the CICS among a sample of high health service utilisers with multimorbidity. Patient activation is likely to have a clinical impact on this population (Greene & Hibbard, 2012). An exploratory analysis of the CICS clinical utility was also conducted, with the aim of assessing predictive validity by comparison to established predictors of outcome. Comparable outcome associations to those achieved with established measures would need to be achieved for the CICS to hold sufficient promise as a new measure. The overall purpose of the study was to contribute to the development of more detailed therapist feedback processes, potentially using the CICS as a supporting tool. This method could enhance deliberate practice techniques and improve psychological therapists' effectiveness.

5.2 Method

5.2.1 Participants

The same 11 participants from Malins et al.'s (2016) Cognitive Behavioural Therapy (CBT) case-series used in the previous chapter were included in the current analysis. The same 22 transcripts were also used in this study. They were selected in the same way: one of a session recording was randomly selected from the earlier half of therapy and one was selected from the latter half of therapy for each participant.

5.2.2 Outcome measures

A range of clinical outcomes were selected to assess common problems experienced by high service users and those with multimorbidity. All outcomes were collected in person by researchers who operated independent of psychological therapy provision and CICS development. Self-reported outcome measures were collected at baseline and six-month follow-up:

1. The Acceptance and Action Questionnaire II (AAQ II) is a 7-item measure of psychological inflexibility or experiential avoidance. It demonstrates satisfactory test-retest reliability (*r* = .79) and concurrent validity with a range of established mental health assessments (*r*s = .59 to .71)(Bond et al., 2011).
2. The seven-item Generalised Anxiety Disorder scale (GAD-7) is a measure of generalised anxiety symptoms based on diagnostic criteria. It has excellent test-retest reliability (ICC = .83) and good convergent validity with other validated anxiety measures (*r*s = .72 to .74) (Spitzer, Kroenke, & Williams, 2006).

3. The 18-item, short Health Anxiety Inventory (SHAI) is a measure of health anxiety assessing worries about health and bodily changes. The SHAI shows excellent test-retest reliability (r = .90) and convergent validity with assessment of hypochondriasis (r = .85) (Salkovskis, Rimes, Warwick, & Clark, 2002).
4. The Euroqol - 5 dimensions - 3 levels (EQ5D-3L) is a measure of quality of life (EuroQol Group, 1990). It assesses functional, physical and mental health components of life quality. It has shown acceptable test-retest reliability (ICC = .52) and good convergent validity with established assessments of wellbeing (r_s = .77) (Janssen, Birnie, Haagsma, & Bonsel, 2008; Janssen et al., 2013).
5. The 15-item Patient Health Questionnaire (PHQ-15) is a measure of somatic symptoms, wherein psychological distress is manifest in physical symptoms. A satisfactory level of test-retest reliability has been determined (α = .65) and reasonable convergent validity with measures of general health (r = .44) and depression (r = .56) (Kroenke, Spitzer, & Williams, 2002).
6. The nine-item Patient Health Questionnaire (PHQ-9) measures depression symptoms based on diagnostic criteria for major depression. Excellent test-retest reliability is confirmed (α = .84) and good convergent validity with general mental health assessment (r = .73) (Kroenke, Spitzer, & Williams, 2001).
7. The Ruminative Responses Scale (RRS) is a 22-item measure assessing the degree to which rumination is a typical response when feeling sad, down or depressed. Acceptable test-retest reliability is

established (.60) alongside replicated evidence that increased rumination is associated with greater depression chronicity and severity (Treynor, Gonzalez, & Nolen-Hoeksema, 2003).

8. The Short-Form 36 items (SF-36) is a measure summarising a series of validated instruments, which are combined into assessments of general mental health (mental component scale: MCS) and physical health (physical component scale: PCS) The specificity of each scale has been effectively validated through association with common physical and mental health problems (Ware, Kosinski, & Gandek, 2005).

5.2.3 Observer-rated instruments

5.2.3.1 The CICS

The CICS is an observer-rated assessment of individual psychological therapy interactions on the basis of patient activation (Malins et al., 2018). The CICS categorises and rates patient activation for each in-session turn of speech. First a topic is assigned for the turn of speech from the CICS content themes:

- Action Planning and Idea generation (API) – Discussion of specific plans or potential plans for activities outside the session.
- Evaluations of Self or Therapy (EST) – Offering a personal assessment of therapy or one of the parties in therapy.
- Information Discussion (IDI) – Giving, receiving or requesting specific information.

- Noticing Change or Otherwise (NCO) – Where changes are reported, or a lack of change is described, despite efforts to bring them about.
- Problem Analysis and Understanding (PAU) – An analysis or understanding of a problem is given or received.
- Problem or Context Description (PCD) – A default theme for descriptions of problems and their contexts, which do not deviate into other themes.
- Structuring and Task Focus (STF) – Where verbal efforts to structure, plan, or progress the session are offered or sought. Conversely, where sessions deviate from any relevant topic without intervention from either party.

Once an interaction topic is allocated, the level of patient activation present in that interaction is rated. Scores range from +2 for interactions showing observable, high levels of patient activation and engagement to -2 for interactions showing observable indications of low patient activation and disengagement. The rating and score methods are described in detail in Chapter 4 and in Appendix F.

5.2.3.2 Working alliance inventory

The observer-rated working alliance inventory measures the therapeutic bond between therapist and patient; agreement on tasks of therapy, and agreement on goals of therapy (Horvath & Greenberg, 1989). The working alliance inventory was selected to assess convergent validity with the CICS because it is the most commonly used assessment of interaction qualities in psychological therapy (Horvath et al., 2011). Excellent inter-rater reliability and

internal consistency (α = .81 to .87, with similar ICCs for IRR) has repeatedly been achieved and more studies have shown a robust and replicable relationship between the working alliance inventory and clinical outcome than any other measure in psychological therapy (Flückiger et al., 2018). The original 36 item scale has been adapted to 12 and 6 item measures, achieving very similar results for internal consistency and excellent convergent validity with the original scale (*r* = .91 to .95) (Falkenström, Hatcher, Skjulsvik, Larsson, & Holmqvist, 2015; Hatcher & Gillaspy, 2006; Tichenor & Hill, 1989; Tracey & Kokotovic, 1989). This study used a six-item observer-adapted working alliance inventory (WAI-O; Falkenström et al., 2015; Horvath & Greenberg, 1989). This included six items - all from the original observer-rated working alliance inventory - each rated on a 7-point scale (1 = never to 7 = always). The WAI-O was used as an established and reliable predictor of outcome from many previous studies: $r_{adjusted}$ = .24 across 150 studies, indicating a moderate effect size (Flückiger et al., 2018). In particular, one might expect that increased patient activation could co-occur with improved therapeutic alliance. Therefore, concurrent assessment of the WAI-O and CICS was used to assess convergent validity. So the WAI-O was used to assess every session that was rated with the CICS.

5.2.3.3 Cognitive therapy rating scale – revised

The CTS-R (Blackburn et al., 2001) is the most widely used assessment of competence for CBT therapists in the UK. It is routinely used in training of CBT therapists in the national IAPT programme (e.g. University of Exeter, 2019). It is therefore likely to be the most widely used assessment of psychological therapy interactions in routine practice. The scale rates different

areas of therapeutic interaction (e.g. agenda setting) on a 0 to 6 scale (0 = therapy component absent or very poorly conducted by the therapist; 6 = therapy component is performed with a very high level of skill and adherence, even in the face of patient difficulties). Adequate IRR (ICC = .63) and excellent internal consistency (αs = .92 to .97) has been reported and trainee CBT therapists show significant improvement in scores as training progresses (Blackburn et al., 2001; Milne et al., 1999). In this study the CTS-R was used as a secondary measure of convergent validity. Although the CTS-R has had variable success in predicting clinical outcomes (e.g. Branson et al., 2015), it is plausible that increased therapist competence would relate to increased patient activation.

5.2.4 Coders

The two original thematic analysts who developed the initial CICS theme template acted as coders in the current study. However, to fairly assess both the IRR and the generalisability of the CICS, two independent therapists were trained to use the coding manual. Neither of the naïve coders had been involved in any aspect of the coding manual development or any of the previous related work and were blind to outcome. Both new coders were trained CBT therapists with experience of working with adults with a range of common mental health difficulties. Training involved three weekly training sessions that were two hours in length, with each session requiring two further hours of coding to be completed between sessions. The first session included an introduction and explanation of the CICS, alongside its theoretical and empirical underpinnings. This was followed by an explanation of the coding system and brief demonstration of how coding would be completed. The new

coders were trained using six of the transcripts that had been rated by both original coders, because adequate agreement had already been established. Therefore, existing scores were used as the standard against which the new coders' ratings were assessed. Between each training session a new transcript was coded independently. At the following training sessions difficulties with coding were addressed and corrective feedback was given on their performance. A transcript was then worked through together in the session with the trainer (one of the original coders) and a new transcript was given at the end to work on independently.

After training, feedback was sought from the new coders, including reflections on ambiguity in coding definitions and overlap between themes. This helped establish an independent view of the CICS' usability. The CICS was amended, accounting for feedback. All four coders were then given a final version of the CICS and the 10 transcripts not subject to previous coding to assess IRR. These coders also rated each session with the WAI-O using the standard published guidance.

An experienced CBT therapist and trainer used the CTS-R to rate recordings of each included session. The CTS-R rater was also blind to outcome and was not involved in the development of the CICS. The CTS-R rater had more than 10 years' experience of using the CTS-R to assess CBT therapist competence and worked separately from the coders. Nonetheless, rating reliability is likely to be reduced by single-coding.

5.2.5 Procedure

Initially, IRR was established between four CICS coders. The four coders used the CICS to independently rate 10 CBT sessions from five of the

11 participants. Individual scores for each turn of speech were used to assess IRR at a turn-by-turn level as well as using an overall score for the session.

All session transcripts for all patient-participants (one from earlier in therapy and one later) were rated by at least two of the four coders using the CICS and the WAI-O. Each session was also assessed by a different independent rater using the CTS-R. These assessments were used to evaluate convergent validity with the CICS.

The CICS, WAI-O and CTS-R scores for earlier and later sessions were separately correlated with the residualised change in clinical outcome scores for each measure from baseline to six-month follow-up. That is using the change in score residualised with the baseline score in a linear regression, giving a change score which accounts for baseline severity. This gave initial comparative correlations with clinical outcomes for the CICS, WAI-O and CTS-R.

To illustrate the CICS discriminative validity, an extreme case comparison was undertaken. The two participants with the highest and lowest scores on the CICS were respectively compared on clinical outcomes, WAI-O and CTS-R scores.

5.2.6 Statistical Analysis

For each CICS theme, IRR assessment was based on four independent ratings of each theme at each turn-of-speech. This meant that data units were defined as transcribed turns-of-speech and content units were CICS themes rated by patient activation level (-2 to +2). If neutral assessments of patient activation were made or where themes were rated as absent, these were scored as 0, to lie between positively- and negatively-rated patient activation.

Transformed scores were part of the focus in the IRR analysis. Transformed scores were calculated by multiplying the percentage of transcribed words coded at each level of each theme by the score for the given level. For example, if 10% of a transcript was coded as Action Planning and Idea generation (*API*) at the +2 level, this would contribute a score of 20 to the total for the *API* theme (+2 x 10). Conversely, if 10% of the same transcript was scored as *API* -1, this would contribute a score of -10 to the total for the *API* theme (-1 x 10). If these were the only levels of *API* coded for the transcript a total score for *API* on that transcript would be 10 (20-10). This calculation was carried out for all themes on all transcripts, and the total theme scores were then compared across coders.

As described in the previous chapter, the exception to this rule was the problem and context description theme (*PCD*), which was instead converted to a binary rating: present or absent at each turn of speech. The *PCD* score was the percentage of the transcript coded present.

Assessment of IRR was also conducted at the level of individual speech turns. Therefore, additional ICCs were calculated for each theme on an interaction turn-by-turn basis.

Two-way, random effects, consistency, average-measures ICCs were used to assess IRR (Shrout & Fleiss, 1979) using SPSS 24. Binary turn-by-turn ratings of the *PCD* theme were assessed using Siegel and Castellan's (1988) kappa, averaged across each coding pair using Light's (1971) method. This meant that absolute agreement on presence or absence of PCD for each turn of speech was assessed with a kappa statistic and the scores were averaged across all possible pairs of coders. Each session was represented

by one summary score for each CICS theme by averaging scores between coders. Correlations among process and outcome scores were calculated from scores for all sessions using two-tailed Spearman's rho correlation coefficients (i.e. one set of scores for sessions earlier in therapy and one set for later sessions). In acknowledgement of the limitations associated with outcome correlations as a means of assessing outcome prediction, the component of this chapter reporting predictive validity is deemed initial evidence. More rigorous methods of outcome prediction are assessed in the next chapter with a larger patient sample.

To elaborate the clinical significance of extreme case comparisons, reliable changes that cross published clinical cut-off boundaries and/or meet published criteria for minimal clinically important change are of particular interest. For example, scores above 9 on the PHQ-9 would fall within a defined clinical range and a reduction of six or more points is recognised as reliable improvement (Clark & Oates, 2014). Where published cut-offs were not available, Jacobson and Truax's (1991) criteria were used to compute clinical cut-offs at two standard deviations below the mean, based on published data from the clinical reference sample. Published information on the internal consistency and standard deviation of such measures was used to derive the minimal change-score required to meet the Reliable Change Index. The Reliable Change Index demonstrates (with 95% confidence) that change could not be accounted for by measurement imprecision alone. It is used here as an indication of minimal clinically important change.

5.3 Results

5.3.1 Inter-rater reliability

Analysis included ratings of 3,134 turns of speech from 10 CBT sessions with five patients and two therapists, rated by four coders working independently (results are summarised in Table 5.1). The data used had been previously used for thematic analysis, but had not been coded with the CICS. All transformed theme scores achieved ICCs within the “excellent” range of inter-rater agreement (ICCs = .84 to .97), except IDI (ICC = .60) which fell within the “good” range (Cicchetti, 1994). The wide 95% confidence interval on IDI (-.05 to .89) suggests that more coding data may be required for a reliable ICC. Excellent range ICCs were also achieved for both the combined average of global themes (collaborative flow and integration of themes) and the average of content themes (all remaining CICS themes apart from problem and context description) (both ICCs = .97). All turn-by-turn theme ratings achieved ICCs within the good to excellent range, (ICCs = .69 to .80; Cicchetti, 1994). Turn-by-turn, the problem and context description theme (PCD) showed moderate to substantial agreement, (κ = .54 to .61; Landis & Koch, 1977).

Table 5.1. Intra-class coefficients for transformed scores and turn-by-turn ratings.

Coded Theme	ICC	*95% CI*		[a]Turn-by-turn coding ICC	*95% CI*	
		Lower	Upper		Lower	Upper
Global Ratings						
Collaborative Flow (COF)	0.95	0.88	0.99	-	-	-
Integration of Themes (IOT)	0.95	0.86	0.99	-	-	-
Totals						
Combined content ratings	0.96	0.88	0.95	-	-	-
Combined global ratings	0.97	0.92	0.99	-	-	-
Content Ratings						
Structuring and Task Focus (*STF*)	0.97	0.91	0.99	0.67	0.65	0.70
Action Planning and Idea Generation (*API*)	0.89	0.72	0.97	0.82	0.81	0.83
Noticing Change or Otherwise (*NCO*)	0.94	0.84	0.98	0.77	0.75	0.78
Problem and Context Description (*PCD)*	0.87	0.65	0.96	[b]0.57	[b]0.54	[b]0.61
Problem Analysis and Understanding (*PAU*)	0.84	0.59	0.96	0.69	0.67	0.71
Evaluations of Self or Therapy (*EST*)	0.85	0.60	0.96	0.72	0.70	0.74
Information Discussion (*IDI*)	0.68	0.13	0.91	0.79	0.77	0.80

Note: Abbreviations: CI, Confidence Interval; ICC, intra-class correlation coefficient

[a]876 turns of interaction were excluded from turn-by-turn analysis as they were part of a pilot method where turns were coded in agreed groups

[b]Siegel and Castellan's (1988) Kappa was calculated for agreement on PCD and 95% CI is estimated on the basis of 1,000 bootstrapped samples

5.3.2 Convergent validity

Correlations across all 22 rated sessions showed significant, strong to very strong Spearman's rho correlations between all CICS themes and the WAI-O (r_ss = .72 to .91; p < .001), with the exceptions of structuring and task focus (STF; r_s = .36; p = .102) and information discussion (IDI; r_s = .08; p = .719). Significant, moderate to strong correlations were found between five of the nine CICS themes and CTS-R scores (Table 5.2). The WAI-O and CTS-R had significant negative correlations with the problem and context description theme (WAI-O: r_s = -.71, p <.001; CTS-R: r_s = -.47, p = .029).

Table 5.2. Spearman's rho correlation between therapeutic alliance, therapist competence and CICS themes.

Consultation Interaction Coding Scheme theme	**WAI-O** Correlation *p value*	**CTS-R** Correlation *p value*
Action Planning and Idea Generation (*API*)	.72** *<.001*	.39 *.071*
Evaluations of Self or Therapy (*EST*)	.83** *<.001*	.73** *<.001*
Information Discussion (*IDI*)	.08 *.719*	-.03 *.883*
Noticing Change or Otherwise (*NCO*)	.83** *<.001*	.40 *.066*
Problem Analysis and Understanding (*PAU*)	.80** *<.001*	.55** *.008*
Problem and Context Description (*PCD)*	-.71** *<.001*	-.47* *.029*
Structuring and Task Focus (*STF*)	.36 *.102*	.41 *.058*
Collaborative Flow (*COF*)	.89** *<.001*	.57** *.006*
Integration of Themes (*IOT*)	.84** *<.001*	.68** *<.001*
Global average	.90** *<.001*	.63** *.002*
Content average	.90** *<.001*	.62** *.002*
WAI-O	- -	.67** *.001*

Note. *p <.05 (2-tailed) **p <.01 (2-tailed). Abbreviations: WAI-O: Working Alliance Inventory – Observer-rated; CTS-R: Cognitive Therapy rating Scale – Revised.

5.3.3 Predictive validity

The valence of all correlations reported below has been adjusted so that positive correlations consistently indicate improvement in clinical outcome and higher CICS scores. Significant, strong Spearman's rho correlations were found between structuring and task-focus (*STF*) interactions in the earlier half of therapy and residualised improvements in: overall mental health (Table 5.3; SF-36-MCS: .76; $p = .006$); psychological flexibility (AAQ: $r_s = .71$; $p = .022$); health anxiety (SHAI: $r_s = .68$; $p = .042$) and rumination (RRS: $r_s = .64$; $p = .048$). Non-significant, moderate correlations were observed with improvements in depressive symptoms (PHQ-9: $r_s = .61$; $p = .060$) and generalised anxiety (GAD-7: $r_s = .52$; $p = .102$). No significant outcome correlations were observed with the WAI-O, the CTS-R or other CICS themes for earlier or later sessions.

Table 5.3. Spearman's rho correlations between early session structuring score and change in outcome score at six month follow-up.

Residualised outcome change (baseline to 6 months)	**Early session Structuring and task-focus** Correlation *p value*
Acceptance and Action Questionnaire	.709* *.022*
Euroqol – 5 Dimensions – 3 levels	-.345 *.298*
Generalised Anxiety Disorder – 7 items	.518 *.102*
Patient Health Questionnaire – 9 items	.612 *.060*
Patient Health Questionnaire – 15 items	-.064 *.852*
Ruminative Responses Scale	.636* *.048*
Short Form – 36 items Mental Component Scale	.764** *.006*
Short Form – 36 items Physical Component Scale	-.227 *.502*
Short Health Anxiety Inventory	.683* *.042*

Note. * p< 0.05 (2-tailed); ** p< 0.01 (2-tailed). The valence of correlations has been adjusted so that positive correlations relate to improvement in outcome and higher CICS scores.

5.3.4 Outcomes from extreme case examples

Brief descriptions of the participants with the highest and the lowest CICS scores are given followed by a summary of the clinical outcomes reported at baseline and six-month follow-up[4]. These were compared to CICS, WAI-O and CTS-R scores.

5.3.4.1 Highest CICS scorer

Participant 2 was a 35-year-old mother of two who had periodically suffered with anxiety, depression and bulimia for several years. This became more severe after the birth of her second child six months prior to the start of psychological therapy. Stress-related skin psoriasis also emerged at a similar time. She became very concerned about potential health problems and struggled to find a sense of purpose. Therapy sessions focused on addressing rumination and included assertiveness training, as rumination often followed passive responses to relationship difficulties. Outcomes indicated reliable improvements in generalised anxiety (GAD-7, 20 to 10), health anxiety (SHAI, 23 to 17), depressive symptoms (PHQ-9, 16 to 10), psychological flexibility (AAQ-II, 44 to 33), rumination (RRS, 71 to 55) and the physical and mental component scales of the SF-36 (MCS, 19.8 to 26.9; PCS, 47.2 to 58.4). Contrary to the trend of improvement, a minor deterioration in health-related quality of life (EQ5D, .85 to .81) was also reported, though scores began and remained well above the clinical cut-off. No reliable change was shown in somatic symptoms (PHQ-15, 11 to 11).

[4]Demographic details of both cases are similar but disguised to protect anonymity.

5.3.4.2 Lowest CICS scorer

Participant 9 was a 78-year-old retired business man who had developed multiple health problems over several years. These included hypertension, arthritis and diabetes. Therapy sessions focused on psychoeducation about the impact of anxiety, coupled with strategies to be more purposeful in consultations with the patient's GP. Outcomes indicated reliable improvement in health-related quality of life (EQ-5D, .16 to .62), but reliable deterioration in psychological flexibility (AAQ, 7 to 31), depression (PHQ-9, 3 to 10), generalised anxiety (GAD-7, 0 to 6) and health anxiety (SHAI, 29 to 44). No clinically important change was shown in rumination (RRS, 36 to 27), somatic symptoms (PHQ-15, 18 to 15) or the physical or mental component of the SF-36 (MCS, 50.4 to 50.8; PCS, 26.1 to 21.8).

Comparison of CICS scores show that Participant 2 had strongly positive scores (CICS content average: 13.33; global average: 1.31) whilst Participant 9's average global and content scores were negative (CICS content average: -3.61; global average: -1.06). Participant 9 also had a larger proportion of the session coded *problem and context description* (84.45 vs. 47.16), as was hypothesised to be the case for interactions with poorer outcomes. The biggest contrast in CICS scores was on the structuring and task-focus theme (*STF*, 8.55 vs. -28.7), with verbal signs that Participant 2 was engaged in structuring sessions, choosing tasks, and correcting deviations:

> **Therapist**: Is there anything that from… other than [goals], that what would you like to talk about?
>
> **P2**: No that sounds pretty much probably what I was kind of expecting so that's cool.
>
> **Therapist**: Ok cool. So shall we start with the kind of goals?
>
> **P2**: Yeah.

[P2 and therapist start to describe a separate issue]

P2: Sorry are we digressing! [laughs]

Therapist: No, no.-

P2: Supposed to be talking about goals here! [laughs]

Conversely, Participant 9's sessions were characterised by deviations from agreed tasks to unrelated topics with neither party correcting course for several turns:

[A discussion about burning hedge clippings]

Therapist: Oh well that's nice at least you will wait for the washing.

P9: Well yeah, you know the legal on that – there is no set time when you can have a bonfire as long as it isn't causing any danger. I just have one of those dustbins with holes in it – incinerator thing – I just stick the lid on it and it's got a funnel.

Therapist: But I suppose its politeness though isn't it to wait for people to bring their washing in?

P9: Well, I wouldn't want my washing smoking.

Therapist: No, not at… it's just consideration.

P9: Not when I've put [fabric conditioner] in it.

Participant 2's averaged WAI-O score (5.8) was nearly double that of Participant 9 (3.1), indicating a contrast in therapeutic alliance. Diverging scores on the CTS-R were also received. Participant 2's sessions receiving an average rating of 45 which fell within the "proficient" range, whereas Participant 9's average rating of 29 fell below the "competent" threshold of 36 (Blackburn et al., 2001). The two participants also had the highest and lowest consultation rates in the sample (P9 = 60 consultations; P2 = 27 consultations).

5.4 Discussion

The current study suggests that individual psychological therapy interactions can be reliably rated based on patient activation and scores closely related to those from well-used assessments of interaction qualities. There may be an important relationship between turn-by-turn interactions and clinical outcomes, particularly patient activation in structuring and task focus interactions. The study was carried out amongst long-term high users of healthcare with multimorbidity, for whom patient activation plays an important role in health improvement (Greene & Hibbard, 2012; Greene et al., 2015).

Novice coders were trained to use the CICS with good-to-excellent IRR. Composite scores for the CICS and most individual themes showed strong to very strong correlation with therapeutic alliance and therapist competence. This indicates good convergent validity. Importantly, therapeutic alliance and therapist competence were both negatively correlated with general problem description interactions. This indicates that more time spent on problem description (in contrast with more active interaction themes such as problem analysis) may be associated with reduced therapeutic alliance and therapist competence in CBT for people with multimorbidity.

Perhaps the most significant finding is that structuring and task choice interactions in early sessions significantly correlated with multiple health outcomes, whilst no other interaction themes or process measures did. This suggests that actively engaging patients in the process of structuring sessions and choice of therapeutic tasks early in therapy may be important to overall outcome. These findings were supported by an extreme case illustration where the biggest contrast in scores were found in structuring interactions.

Importantly, the structuring and task choice theme outperformed the WAI-O and CTS-R in terms of outcome prediction. This gives initial indication that some components of the CICS may have value compared to established instruments in outcome prediction. The contrasting consultation rates may indicate that CICS ratings of patient activation are associated with service use in a similar way to PAM ratings.

5.4.1 Strengths

Typically, interaction rating schemes lead to total behavioural counts or global scores on identified dimensions. Turn-by-turn analysis establishes the specificity of coders' judgements for each unit of measurement. This provides a stronger sense of how well each coder is agreeing on a specified data point. Otherwise, spurious chance agreement on overall scores could not be ruled out. This approach also reduces the potential for bias in outcome prediction, because patient comments made during rated sessions about clinically relevant improvement/deterioration would only affect ratings of one interaction not session-level scores; whereas global scores may be adjusted on the basis of a single comment that draws raters' attention.

Achieving reliable turn-by-turn analysis also allows easier identification of clinically important changes that occur within the session. Such changes can affect outcome and may not be picked up by total score analyses (Safran et al., 2011; Swift et al., 2017). The scores derived from ratings of each individual interaction can, if required, be used to look at detailed analyses of particular parts of a session (which may be the focus of a deliberate practice review) without having to re-rate or rescore. In addition, reliability has been achieved amongst a sample from a heterogeneous population with multimorbid

difficulties and without a specific primary diagnosis. This is a helpful indication for the potential application of the CICS amongst patients with complex difficulties. Overall, it suggests that a coding manual and training method has been developed with sufficient clarity, specificity, and reliability to provide a foundation upon which further study of interaction-level outcome prediction can be built.

5.4.2 Limitations

Adequate IRR was not achieved for interactions where information was offered or received, which may relate to the comparatively small number of turns to which it was applied. Alternatively, it could relate to confusion between information discussion and problem analysis interactions highlighted by novice coders in their training feedback. Coders identified that information discussion could be framed in a personalised manner, which was at times interpreted as a type of problem analysis. Attempts were made to redefine both themes for clearer delineation, but this may not have been adequately achieved. These problems may have contributed to poor convergent validity between information sharing interactions, therapeutic alliance and therapist competence. However, the strong convergent validity maintained for problem-analysis interactions suggests that coding of these was less affected by the confusion.

It is unclear why structuring and task focused interactions also did not achieve convergent validity with therapeutic alliance or therapist competence. It may indicate that structuring interactions are a considerably different construct to the other types of interaction, given the differential in predictive validity. Alternatively, it may be an artefact of multiple comparisons made with

several CICS themes. It is also unclear why some outcomes correlated with structuring interactions but others (e.g. quality of life) did not. Such information would advance understanding of which interaction types are most important and for which particular clinical problems. Common method variance is a possible competing argument for the high rates of convergent validity: As the same people rated patient activation and therapeutic alliance, correlations may be inflated. However, convergent validity was similar for therapist competence, which was rated independently.

Given that all interaction themes were rated similarly, it is notable that no other CICS theme correlated with any outcome. As with convergent validity, this may highlight structuring interactions as a particularly important and different type of interaction. However, it may also suggest that some themes are redundant for outcome prediction. Chance correlations between structuring interaction scores and outcomes are also a possibility given that multiple correlations were conducted. Larger-scale assessment using more sophisticated methods of outcome prediction would clarify whether low power or chance agreement affected the current results.

Overall, it is important to temper any confidence in the conclusions drawn given that this was, in quantitative terms, a pilot sample for analysis. For example, included participants received a wide range of CBT sessions (4 to 40), which could limit the generalisability of findings. Furthermore, convergent validity with therapeutic competence used only used a single coder so rating reliability could be questioned. The evidence from extreme cases must also be held lightly, as they might overplay the degree of difference that would be observed on average.

5.4.3 Conclusion

The high level of consistency in scoring suggests the CICS will be sufficiently dependable for future outcome-prediction studies. The high level of convergent validity and potential for clinical utility indicates that the CICS warrants further study. Therefore, a larger-scale assessment of the CICS' predictive validity within a similar population is presented in the following chapter. This will clarify whether the initial evidence of predictive validity presented above remains when tested with a more robust design. In terms of predictive validity, simple outcome correlations as used in this chapter have been superseded by more refined means of outcome prediction. Multilevel Modelling is a sophisticated statistical method that allows outcome prediction to control for other important variables, whilst accounting for the hierarchical structure of data (Hox, Moerbeek, & Van de Schoot, 2017). These methods will be applied in a more rigorous test of the CICS in the next chapter.

6. THE PREDICTIVE VALIDITY OF THE CICS

6.1 Introduction

This chapter expands the initial evidence of predictive validity shown in the previous chapter with a more extensive investigation of the CICS clinical utility. The chapter introduces a new, larger but similar sample of high service utilisers who received CBT for health anxiety remotely – via video conferencing or the telephone – as part of an RCT. Using a secondary analysis of this trial data a more rigorous methodology is used to evaluate the CICS' predictive validity. Furthermore, this study predicts outcome using CICS ratings from only the first and second sessions of CBT, given the potential value of earlier sessions to outcome identified in the previous chapter. In sum, this means that the current study is a much more rigorous test of the CICS. But it also means that more informed comment can be made about the type of interactions that occur early in psychological therapy and how they affect eventual outcomes.

6.1.1 Early interactions in psychological therapy

The way interactions proceed between CBT therapists and anxious patients early in treatment has potential to be an important factor determining clinical outcomes (Aviram et al., 2016). However, it is unclear to what degree motivational factors overlap with therapeutic alliance. Early alliance change has also been shown to predict subsequent symptom change and this could be a similar mechanism to motivation (Falkenström, Ekeblad, & Holmqvist, 2016). For example, low ratings of agreement on tasks could well correlate with greater patient resistant or CCT.

Early interactions are likely to be important in CBT for health anxiety, because multimorbidity is common amongst health anxious patients and dropout is often higher than for other conditions (Greeven et al., 2007). Furthermore, poor engagement of patients with health anxiety can have extensive and costly consequences for the whole healthcare system (Bermingham, Cohen, Hague, & Parsonage, 2010).

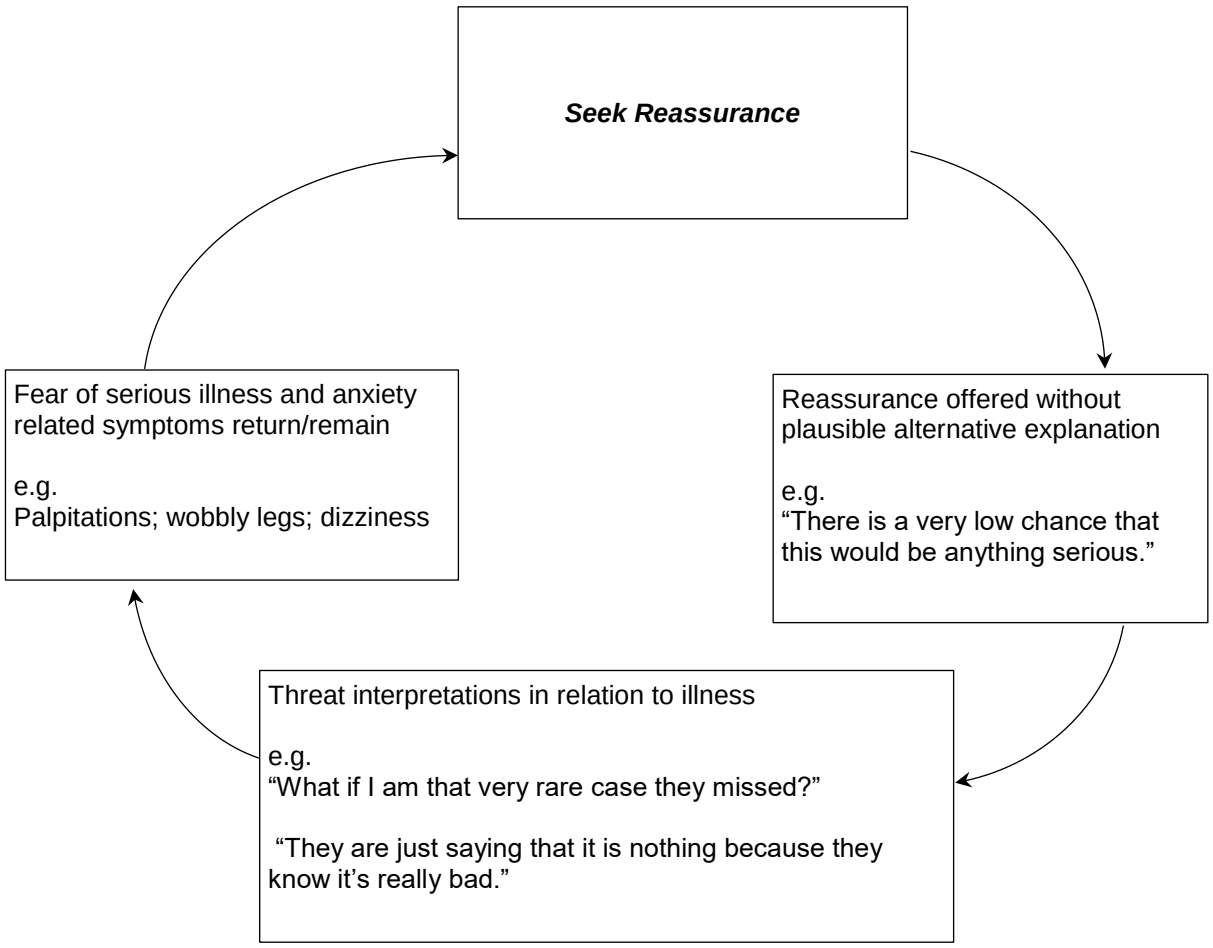
6.1.2 Health anxiety and high service use

Health anxiety is defined as chronic preoccupation with health worries or a belief that one has a serious illness and the problem is not resolved by medical reassurance (Fink, Ørnbøl, & Christensen, 2010; Salkovskis & Warwick, 1986). In the most recent version of the diagnostic and statistical manual of mental disorders (DSM-5) health anxiety is defined as somatic symptom disorder if anxiety occurs with somatic symptoms and illness anxiety disorder if occurring without somatic symptoms (DSM-5; American Psychiatric Association, 2013). Approximately 5.7% of the population are affected by health anxiety across their lifetime (Sunderland, Newby, & Andrews, 2013). However, rates of health anxiety increase dramatically in medical clinics, where prevalence can be up to 24% (P. Tyrer, Cooper, Crawford, et al., 2011). There are severe and disabling consequences to health anxiety, including increased risk of many chronic physical and mental health conditions and a sometimes dramatic increase in healthcare use (Berge et al., 2016; Fink et al., 2010). The high healthcare use that can occur with health anxiety combined with the increased risk of health anxiety amongst those seeking medical help, has led to its identification as a possible cause of high service utilisation.

Mechanisms have been identified which explain the relationship between high health anxiety and high service use: Health anxiety can be fuelled by a vicious cycle of repeated reassurance seeking (see Figure 6.1; Salkovskis, 1989; Salkovskis & Warwick, 1986). In this cycle the fear of a serious health problem leads health anxious patients to seek healthcare consultations. If no serious illness is identified and the patient is reassured verbally by the healthcare professional it provides temporary relief from anxiety. However, anxiety often quickly returns when a plausible alternative explanation is not given for anxiety symptoms, which fuels further medical reassurance. For example, Lucock, Morley, White, and Peake (1997) found that health anxiety returned within 24 hours of reassurance being given amongst patients with pre-existing high health anxiety. Over time, reassurance can become like a drug where more reassurance is required more quickly, because the relief achieved becomes much shorter-lived.

This issue is also complicated by health-related information on the internet, which can contribute to high healthcare use amongst health anxious patients (Eastin & Guinsler, 2006). Further difficulties occur when healthcare professionals send health anxious patients for (sometimes costly) medical investigations in order to reassure the patient that nothing serious is wrong. This strategy can easily backfire as health anxious patients are more likely to hold threat-based beliefs about their health. As a result, medical investigations can heighten health anxiety rather than reassure, as it can give the message that there is sufficient risk of a serious problem to warrant the test in the first place (Salkovskis et al., 2003).

Figure 6.1. The vicious cycle of reassurance seeking.



In sum, there is much theoretical and experimental evidence to suggest that health anxiety could be a significant contributor to high service use. The issue is exacerbated, because health anxiety sufferers often have comorbid, chronic physical health conditions (e.g. Rode, Salkovskis, Dowd, & Hanna, 2006), which can lead to high service use in itself (Deeny et al., 2018). Patient activation can play a key role in health improvement and reduction of service use amongst patients with chronic health conditions, through improved self-management of health (Barker et al., 2018). However, health anxiety is rarely recognised by patients or healthcare professionals, with most sufferers unaware that they have the problem (P. Tyrer et al., 2016). Despite effective treatment being available it is seldom offered or taken up (Kirmayer & Looper,

2006; Sunderland et al., 2013; P. Tyrer et al., 2014). One of the best-evidenced treatments is CBT (P. Tyrer et al., 2017).

6.1.3 Cognitive Behavioural Therapy for health anxiety

Cognitive Behavioural Therapy for health anxiety focuses on reducing the cognitive and behavioural habits that inadvertently maintain anxiety (Salkovskis et al., 2003). This is achieved through identification of anxiety-maintaining beliefs about health, illness and associated issues then using a range of strategies to reality-test these beliefs. For example, behavioural experiments are used to explore the effects of specific behaviours on anxiety. In the case of reassurance-seeking a patient may be asked to seek much more reassurance for a limited period of time to see what effect this has on anxiety. Often this will lead to an increase in anxiety, due to the cycle described previously. The patient may then test out seeking little or no reassurance for a period of time and evaluating the effect this has on anxiety – usually resulting in reduced anxiety. In this way the patient and therapist collaborate to help the patient reach their own conclusions about what would be most helpful for them, rather than simply doing what the therapist tells them (J. S. Beck, 2011). Although there is a systematic process to CBT for health anxiety, procedures are selected and carried out in an order and at a pace that is responsive to patient presentations (see Table 6.1 for a summary of procedures; P. Tyrer, Cooper, Tyrer, et al., 2011).

Table 6.1: Essentials elements of cognitive behaviour therapy for health anxiety (adapted from P. Tyrer, Cooper, Tyrer, et al., 2011)

- Listing of all symptoms and how the patient interprets them
- Formulation of worst fears
- Introducing possibility of alternative explanations (e.g. pie charts and pyramids)
- Evaluation of risk of disease
- Considering the price paid for health anxiety
- Exploring the hypothesis of fear of disease being more likely than actual disease
- Building evidence by identifying symptom patterns that are related to those of anxiety (i.e. diary keeping)
- Reducing behaviour that maintains health anxiety such as excessive bodily checking and reassurance seeking

One of the mechanisms through which CBT for health anxiety can affect physical health problems is the way in which it seeks to reduce habits that make physical symptoms worse. For example, when highly anxious about health a patient may repeatedly check, prod and manipulate parts of their body. Concurrently, an attentional bias is likely to occur in which patients pay more attention to body parts which have become the focus of concern. Checking of the body means that any pain, rashes or unusual symptoms are exacerbated through manipulation. Attentional biases mean that the patient is much more likely to notice bodily changes or symptoms that would otherwise have occurred outside of awareness. Combined with a health-related threat focus, this leads to misinterpretation of normal bodily changes as threatening symptoms of serious illness. The situation is made worse if the patient has a pre-existing chronic health condition, as they are more likely to experience

problematic symptoms and are at greater risk of anxieties about other health problems (Rode et al., 2006). Therefore, CBT for health anxiety can both reduce anxiety and improve symptom management, particularly amongst sufferers with other comorbidities.

The effectiveness of CBT for health anxiety is well-evidenced through RCTs and meta-analyses. It has performed as well or, most frequently, better than other bona fide treatments (e.g. Clark et al., 1998). The most recent meta-analyses indicated that CBT for health anxiety is highly effective across healthcare settings and patient populations (Cooper, Gregory, Walker, Lambe, & Salkovskis, 2017; Olatunji et al., 2014). Since these meta-analyses, the largest trial of CBT for health anxiety to date has demonstrated an intervention effect at five year follow-up following 6 to 10 sessions of CBT for health anxiety offered in medical clinics (P. Tyrer et al., 2017).

At times confusion has obscured the focus of trials and meta-analyses over the difference between health anxiety, somatisation, medically unexplained symptoms and other similar conditions. This has meant that in some studies different conditions are combined and in others treated separately (Henningsen, Zimmermann, & Sattel, 2003; Olatunji et al., 2014; olde Hartman et al., 2009). Regardless, of the overlapping and idiosyncratic characteristics of these problems, CBT remains an important and effective treatment for health anxiety specifically.

6.1.4 Accessibility of CBT

Although the effectiveness of CBT for health anxiety is well demonstrated, perhaps the biggest challenges are supporting appropriate patients to engage and access treatment. For example, in the initial role out of

IAPT services only 0.1-0.4% of patients seen by IAPT in England were treated for health anxiety, indicating low uptake by patients and referrers (Clark et al., 2009; Glover, Webb, & Evison, 2010). One possible means of improving engagement is through attention to early session interactions, as will be explored with the CICS in this study. This is because some patients do not fully benefit from CBT for health anxiety. Another means of improving accessibility, engagement and reach of CBT for health anxiety is through remote delivery of treatment. Remotely delivered CBT for health anxiety can improve access to psychological therapy and overcome the anxiety and stigma experienced by some patients that prevents them accessing mental health services (Hollis et al., 2015; Kirmayer & Looper, 2006).

Remote, internet-delivered CBT for health anxiety has been successfully and effectively delivered (Hedman, Axelsson, et al., 2014). Furthermore, remotely delivered CBT can achieve outcomes similar to face-to-face therapy (Andrews et al., 2018; Carlbring, Andersson, Cuijpers, Riper, & Hedman-Lagerlöf, 2018). However, unguided or self-guided psychological therapy is generally less effective than when it is delivered by/with a therapist (Baumeister, Reichler, Munzinger, & Lin, 2014; M. G. Newman, Szkodny, Llera, & Przeworski, 2011). This suggests that therapist supported, remotely delivered CBT for health anxiety is a suitable means of increasing engagement and access in therapy for a hard-to-reach patient group, whilst maintaining clinical effectiveness. One recent study investigated using principles of face-to-face CBT for health anxiety whilst delivering treatment remotely, over video conferencing or the telephone (Morriss et al., 2019; Patel et al., 2016). This study aimed to maintain clinical effectiveness through therapist contact in a

manner similar to face-to-face therapy, whilst improving accessibility, maintaining patient confidentiality and overcoming stigma via video conferencing or the telephone.

6.1.5 Video conferencing and telephone delivered CBT for health anxiety

Morriss et al.'s (2019) trial indicated that CBT for health anxiety was effective in reducing health anxiety, generalised anxiety, depression and general health at 12-month follow-up compared to usual care. It also demonstrated cost-effectiveness, with an estimated net monetary benefit of £3,164 per patient at a willingness to pay threshold of £30,000. This included a "hard" healthcare cost saving of £1,064 per patient, which solely included the reduction in healthcare use compared to usual care and excluded the monetary benefits associated with improved health and quality of life. However, there was great variability in patient response to treatment.

Although 62% of the patients who received CBT showed recovery from health anxiety at 12 month follow-up, this meant more than a third did not recover. Furthermore, the range of health anxiety score changes at the end of follow-up (on the SHAI) spanned from a six point deterioration to a 24 point improvement. Although this may be partly explained by differences in baseline severity, it still indicates great variability in responsiveness. Moreover, nearly a third (29%) of those offered CBT in the trial did not attend sufficient sessions to be deemed an adequate treatment does (≥5 sessions). Taken together this suggests that CBT for health anxiety is clinically effective when delivered over video conferencing or the telephone, but there is great unexplained variability in engagement and effectiveness. This chapter uses the above described trial to investigate the role that interaction-level patient activation may play in

explaining the variability in outcomes. The study particularly focuses on initial interactions in the earliest sessions, as interactions at this point can have long-standing effects on outcome (Poulin et al., 2018).

6.1.6 Summary

Health anxiety can be seen as a condition which is emblematic of health problems that coincide with multimorbidity and high service use, but are amenable to psychological therapy. Health anxiety can be effectively treated with CBT and similar effectiveness can be maintained when delivered via video conferencing or the telephone. However, problems with engaging patients and optimising clinical effectiveness remain in remotely delivered CBT. A clearer understanding of early interactions that predict clinical outcomes could guide CBT therapists in maximising treatment effectiveness. The CICS might aid understanding of early interactions in CBT for health anxiety. The previous chapter indicated an association between more active patient involvement in structuring and task-choice interactions during early CBT sessions and greater improvements in overall mental health, rumination, psychological flexibility and health anxiety - the primary outcome in the RCT described. The CICS may help identify important interaction-styles and types early in remotely delivered CBT for health anxiety.

6.1.7 Aims and hypotheses

This study investigates whether patient activation in early session interactions of remotely delivered CBT for health anxiety predicts clinical outcomes; and, if so, which interaction types hold consistent predictive validity across outcomes. The CICS will be used to assess this relationship. Early session therapeutic alliance may overlap with interaction-level patient

activation, but the independent predictive value of each has not been studied. Therefore, all CICS-outcome predictions will control for therapeutic alliance, to evaluate whether the CICS has unique predictive value. The study was predominantly exploratory in nature, as there is insufficient evidence to make confirmatory hypotheses about specific CICS themes or specific outcome relationships. However, two hypotheses were formed from evidence gathered in the CICS development:

- That the CICS would perform similarly to therapeutic alliance in predictive validity, as was the case in the development project
- That structuring and task-choice interactions would be of particular importance to outcome given previous findings.

6.2 Method

6.2.1 Design

An exploratory, observational design was applied to participant data from an RCT of remotely delivered CBT versus usual care for health anxious patients who were using urgent care at a high rate (Morriss et al., 2019). The study objective was to use an observational assessment of patient activation (the CICS) to rate therapist-patient interactions in the first and second CBT session from session recordings and transcripts. Treatment outcome data were then used to assess whether CICS ratings predicted clinical outcomes over time, controlling for baseline outcome score. Observational assessment of the therapeutic alliance was conducted for the same sessions using the WAI-O (Horvath & Greenberg, 1989), so analysis could control for the overlapping predictive value of the WAI-O and CICS.

6.2.2 Source RCT

Patient data came from a single-blind, patient-level, parallel group, multicentre RCT. The RCT recruited patients from UK primary and secondary care settings. This included GP practices, emergency departments, walk-in centres and hospital outpatient clinics including neurology, endocrinology and cardiology. Participants were adults (≥18 years old) who had received two or more unscheduled/urgent consultations with any healthcare provider in the previous 12 months where no underlying pathology had been found. Participants were included if they scored ≥18 on the SHAI when screened by researchers, which is the clinical cut-off for severe health anxiety in UK IAPT services (Clark & Oates, 2014). Participants were excluded if they were experiencing an acute medical condition requiring further ongoing assessment. Those with moderate to severe intellectual disabilities or serious mental illness, including substance use disorders, were excluded if their difficulties meant that engagement in the intervention would not be possible. Participants were also excluded if they had received specialist mental health care within the previous six months. Those with comorbid common mental health problems or chronic, physical conditions were intentionally included.

Participants were typically approached by their treating clinician at the recruiting site. A non-stigmatising description of the intervention was developed in collaboration with patient advisors and delivered by referring clinicians. Most participants were recruited opportunistically during clinical consultations, but some were contacted by GP practices following a records search identifying high users of urgent care. If participants consented to be contacted about the study a researcher telephoned potential participants, gave

them further information about the study and screened them with the SHAI. It was explained that participants had a 1 in 2 chance of receiving CBT for health anxiety and that the study might provide an *additional* source of support rather than replacing their current healthcare provision. Eligible participants were then assessed at a face-to-face baseline interview, during which written informed consent was sought, including consent to record CBT sessions should they receive them and anonymously report extracts (see Appendix G for the baseline assessment pack; Appendix H for the study's participant information sheet, Appendix I for the CBT information sheet and Appendix J for the patient-participant consent form). Following baseline assessment, confirmation of eligibility and consent to take part in the study, participants were randomly allocated to remotely delivered CBT plus usual treatment or Treatment As Usual (TAU) alone. Researchers also asked participants whether they wished to receive CBT via video conferencing or over the telephone, should they be allocated to the CBT arm of the trial.

6.2.3 Remote CBT

Participants allocated to receive remote CBT were sent an information sheet about the intervention (Appendix I); a patient guide to CBT for health anxiety (http://www.clahrc-em.nihr.ac.uk/clahrcs-store/cognitive-behavioural-therapy-health-anxiety); a manual to support the setup and use of video conferencing software if that was their preferred delivery modality and a contingency management plan for when internet connection was poor or not possible. If participants requested session to be provided over the telephone, relevant CBT worksheets were sent out to participants, either as a whole pack, or more commonly on an individual basis. In this way participants and

therapists could work through appropriate therapeutic activities together even if over the telephone.

Participants were contacted by a study CBT therapist within 10 days of randomisation. Patient preference for a male or female therapist and preferred timings of sessions were considered when allocating patients to therapists. If no preference was expressed, patients were allocated on the basis of therapist availability to enable the quickest initiation of therapy. Six to 12 weekly sessions of CBT for health anxiety were offered, with up to three booster sessions if required. This included a “setup” session during which the methods for adapting CBT to remote delivery were discussed and any patient concerns were addressed. Session-by-session progress monitoring was conducted using the Partners for Change Outcome Measurement System (PCOMS; S. D. Miller, Duncan, Sorrell, & Brown, 2005). This method was used to assess treatment response and to support the tailoring of treatment methods. The number of sessions offered was based on the level and rate of treatment response: More sessions were offered where treatment response was slower. An established face-to-face CBT for health anxiety treatment protocol was adapted for remote use in the trial (P. Tyrer et al., 2014). The number of sessions offered aimed to replicate the P. Tyrer et el. (2014) trial (5 to 10 sessions, plus booster sessions – two additional sessions were added to accommodate patient adjustment to the remote treatment modality).

Standard CBT for health anxiety procedures were used in remote CBT (see Table 6.1); where necessary these were adapted to remote delivery. For example, a template Excel spreadsheet was developed in order to conduct the “cause pie chart” collaboratively over video conferencing software (see

Appendix K for an example). The cause pie chart was used to highlight plausible explanations for bodily symptoms such as headaches, which patients feared were a symptom of serious illness. The patient would generate the most common causes of headache and guess at the percentage of headaches caused in this way (e.g. 20% of headaches caused by dehydration). Adapting the intervention for remote use entailed sharing the template within the video conferencing software module so that both patient and therapist could jointly control the mouse and keyboard. Then both parties could contribute fully to the exercise.

Recordings were made available to patients to consolidate learning (Shepherd, Salkovskis, & Morris, 2009). Recordings were also used to improve the effectiveness of clinical supervision. Specifically, the most recent treatment sessions for patients with poor treatment response were reviewed in group supervision and appropriate adaptations to treatment approach were discussed. Every six months, two randomly selected recordings for each therapist were used to evaluate, maintain and improve therapist competence and treatment fidelity. This process was conducted using a version of the CTS-R adapted to CBT for health anxiety (Blackburn et al., 2001; P. Tyrer et al., 2014). Recordings were also used for the secondary analysis described below, in line with patient consent.

6.2.4 Therapists and supervision

Four CBT therapists delivered remote CBT to study participants. Two were male and two were female, all were aged between 30 and 50. Two had received doctoral-level training as clinical psychologists and two had Masters level training. One therapist had less than 5 years qualified experience and

one had more than 10 years qualified experience at the start of study recruitment. One of the clinical psychologists was identified as the lead therapist who designed, developed and adapted many of the treatment and supervision methods used. The lead therapist also allocated patients to therapists on referral.

Clinical oversight and supervision of CBT therapists was conducted by the lead therapist from Tyrer et al.'s (2014) large-scale trial (H. Tyrer). The supervisor was medically trained as a GP and then trained as a CBT therapist 7 years prior to the start of recruitment. The supervisor facilitated monthly group supervision with all therapists using the video conferencing technology applied to remote CBT. The supervisor also led evaluations of CBT therapists' competence using the CTS-R every six months. The supervisor used their medical training to advise therapists on whether symptoms were likely to be caused by an underlying medical pathology where confusions were apparent.

6.2.5 Participants

The original trial recruited 156 participants with 79[5] allocated to the remote CBT arm (see Figure 6.2 for the trial's CONSORT diagram). Most opted to have CBT sessions delivered via video conferencing (n = 53; 67%) with 14 receiving CBT over the telephone (18%). Eleven participants did not attend any sessions (14%). Most participants were not seeking psychological treatment when recruited (69 of 79; 87%) and most reported multimorbid difficulties (75 of 79; 95%).

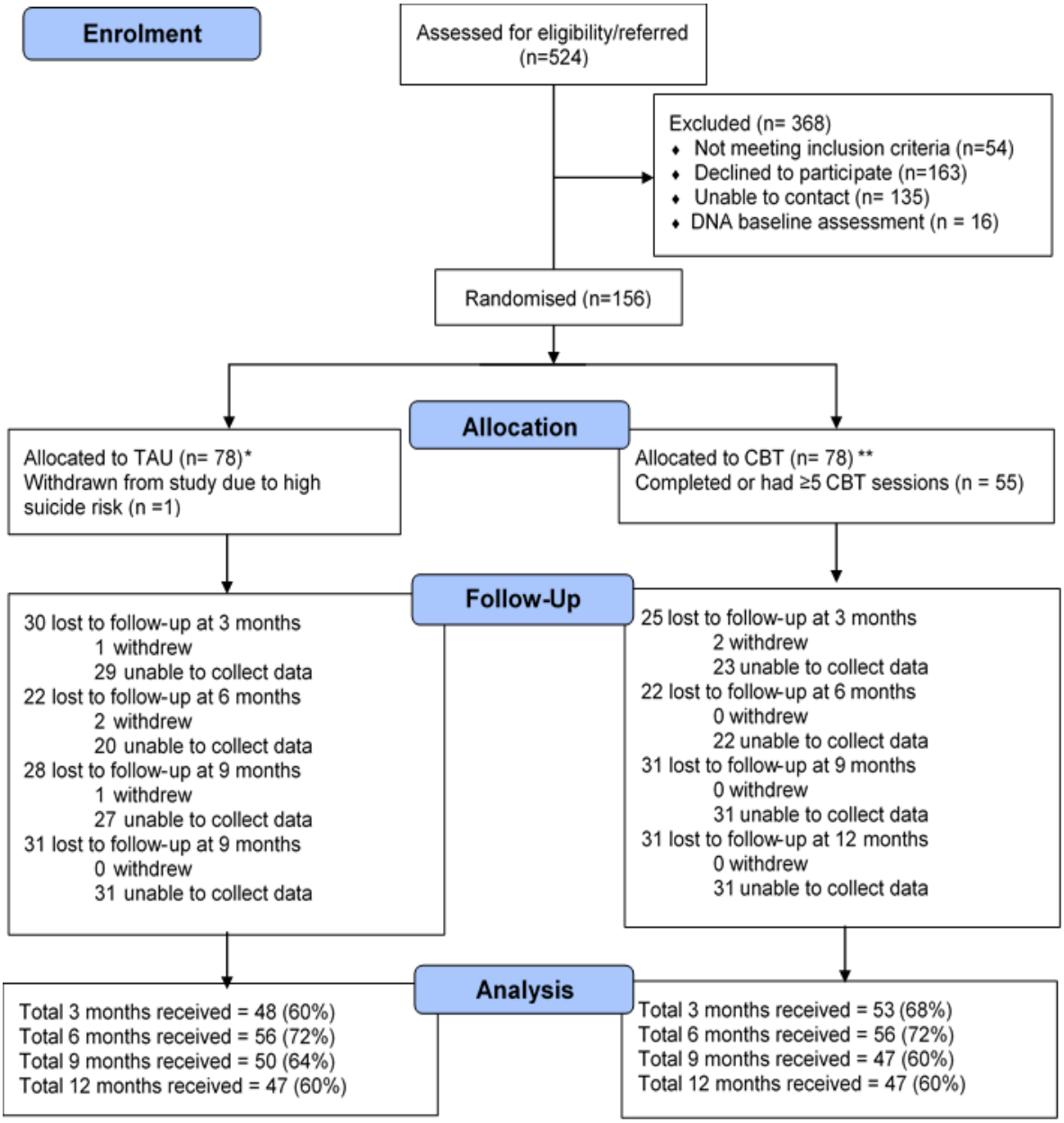
[5]There were 78 participants randomised to CBT and one participant was randomised to usual care, but offered CBT in error. Their data is included in the analysis. Therefore the total sample is 79.

6.2.5.1 Inclusion and exclusion criteria

Fifty of the 79 participants offered CBT (63%) were included, having:

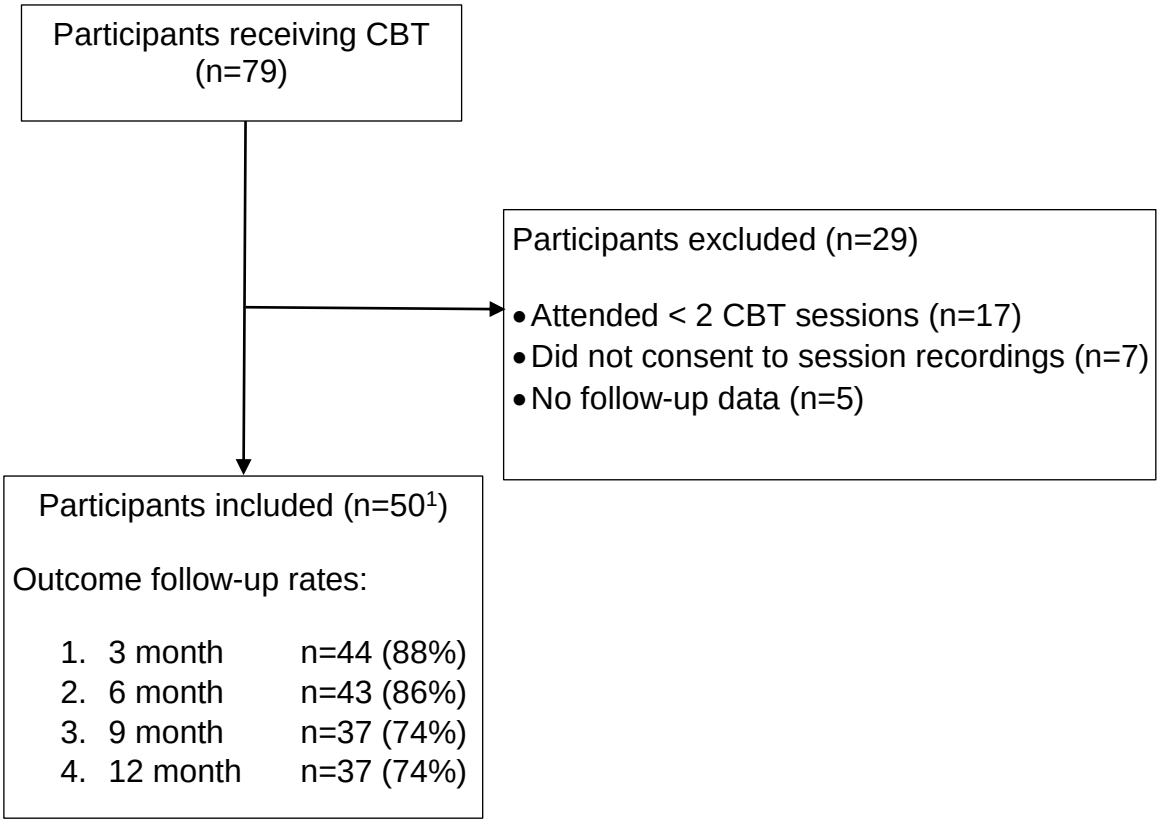
1. Received CBT as part of the source RCT;
2. Attended two or more CBT sessions (providing sufficient data for the analysis);
3. Consented to treatment sessions being recorded and extracts anonymously reported;
4. Completed at least one follow-up assessment on clinical outcomes. All those meeting the criteria were included (Figure 6.3).

Figure 6.2: Morriss et al.'s (2019) CONSORT diagram: Participant flow into the RCT



Note. *There was one randomization protocol violation. One participant who was allocated into TAU by the randomisation system was accidentally sent the incorrect treatment allocation letter resulting in them receiving the remote CBT therapy. This error was identified following the completion of treatment. The participant completed outcome data only at 3 months.
** There was an enrolment protocol violation. Two participants in the CBT group did not meet the criteria of ≥18 on the SHAI. This error was not identified until final analysis and as such both participants were included in the analysis.

Figure 6.3. Participant Flowchart.



Note: [1]One of the participants who did not meet the criteria of ≥18 on the SHAI was included, as the analysis controlled for baseline severity.

6.2.6 Outcome measures

Outcomes were collected by independent researchers blind to treatment group allocation at 3-, 6-, 9- and 12-month follow-up time points**.** Follow-up assessments were carried out by telephone, email, video conferencing, post or face-to-face dependent on participant preference. Given the high rates of multimorbidity amongst health anxiety sufferers, a range of physical and mental health outcomes were used to assess change:

- Short form Health Anxiety Inventory (SHAI; Salkovskis et al., 2002), assessing health anxiety. Change in health anxiety from baseline to six month follow-up was the original trial's primary outcome.

- Generalised Anxiety Disorder 7 items (GAD-7; Spitzer et al., 2006), assessing generalised anxiety.
- Patient Health Questionnaire 15 items (PHQ-15; Kroenke et al., 2002), assessing somatic symptoms.
- Patient Health Questionnaire 9 items (PHQ-9; Kroenke et al., 2001), assessing depressive symptoms.
- Work and Social Adjustment Scale (WSAS; Mundt, Marks, Shear, & Greist, 2002) assessing functional impairment.
- Euroqol – 5 levels (EQ5D-5L, Herdman et al., 2011) assessing quality of life.
- Visual analogue scale (VAS, Herdman et al., 2011) assessing general health.

The structured clinical interview for the Diagnostic and Statistical Manual of Mental Disorders IV (SCID; First, Spitzer, Gibbon, & Williams, 1998) was completed at baseline. Long-term physical health conditions were also recorded from baseline patient interviews.

The outcome rating scale (ORS; S. D. Miller, Duncan, Brown, Sparks, & Claud, 2003) was also used at each session of remote CBT, where participants consented to its use. However, outcome prediction was not conducted with the ORS in the current study as proximal outcomes are more thoroughly investigated in Chapters 7 and 8. This study focuses on prediction of more distal, long-term outcomes.

6.2.7 Procedure

Therapy sessions were routinely recorded if participants consented. Audio or video recordings of the first and second CBT sessions were

transcribed verbatim by an approved third party transcription service (see Appendix L for a copy of the signed transcriber confidentiality agreement). Session recordings and transcripts were then allocated to two trained raters[6] and assessed using the CICS and the WAI-O. For three participants where session two recordings were not available, recordings of the third session were used instead. Once all data was collected, MLM analysis was used to separately assess whether the WAI-O and/or CICS scores from the first or second CBT session predicted follow-up outcomes over time.

6.2.8 Process measures

6.2.8.1 The CICS

The CICS categorises each in-session turn of speech and then rates the level of patient activation. First a topic is assigned for the turn of speech from the CICS content themes:

Action Planning and Idea generation (API); Evaluations of Self or Therapy (EST); Information Discussion (IDI); Noticing Change or Otherwise (NCO); Problem Analysis and Understanding (PAU); Problem or Context Description (PCD), or Structuring and Task Focus (STF). Once an interaction topic is allocated, the level of patient activation present in that interaction is rated. Scores range from +2 to -2 (Table 6.2 gives examples from the current data set of level descriptors from noticing change interactions). The CICS coders were trained using established methods (Malins et al., 2018).

[6]Double coding was completed for 94% of participants rated (47/50).

Table 6.2. Patient activation level descriptors and exemplars for the Noticing Change or Otherwise (NCO) CICS theme.

Level	Level Descriptor[a]	Exemplar Quote
+2	Patient notices a positive change, *articulates it without prompting*. Likely to include: its impact on their life; some sense of personal ownership, or a sense of control over the change. Endorsed or encouraged by therapist with development.	P: So what I've been doing is I've set my clock for 7.15am and I get up at 7.30am, and I've been doing that every day. T: Oh really? P: Yes, so I'm really chuffed with that one, because that's a big change. [P 03014]
+1	Therapist presents the above or jointly with patient and *it is strongly endorsed, developed, summarized or corrected* by the patient.	T: What is your experience of kind of being around people more this week? [ ] Is there anything that you kind of take home from that, kind of take home messages for you? P: [ ] I think I feel like more connected and less cut off and it feels a bit better. [P 1034]
0	Therapists initiated acknowledgement of change, patient acknowledges with *no development* OR very minor acknowledgment of change by the patient not personally "owned", perhaps relating the change to sources outside of their control.	T: Okay. So that sounds-, that looks like things are slightly better than they were when we last met. P: Yes. [P 1023]
-1	Therapists initiated acknowledgement of change, *patient displays ambivalence* towards suggested change in response OR Efforts to make a change are reported but none is seen.	T: Do you notice any difference from having done that from having stopped checking your body, reduced how much you check the internet? P: I only really check my body [ ] if I get some sort of new pain. [ ] I don't check my body that often. T: OK, yeah, but that wasn't a major change to not be checking your body? P: Yeah. [P 1017]
-2	Therapists initiated acknowledgement of change patient *refutes positive change* OR Efforts to make a change are reported but deterioration results.	T: [Things are] slightly worse, like things have not been as good as when we first met. It would be great to hear a bit about what's gone on this week P: I'm finding this more difficult than I thought I would [ ] so I feel I've had a bit of a dip. [P 3002]

Note. [a]Italics add emphasis to the key component of the level descriptor; Abbreviations: P: Patient; T: Therapist.

Following the development studies described in previous chapters, a refinement process was conducted before the start of the current study. All coders met to discuss areas of difficulty with coding from the previous study. The key changes emerging from these discussions were as follows:

Firstly, the trial data had much more session content where patients were led through structured CBT activities than the development dataset. It was agreed that if the therapist was leading or guiding the patient through a set of instructions (e.g. “Now I want you to imagine that you are back at your house.”) the ensuing interaction would be coded as a structuring and task-focused interaction. The level of patient activation would be determined by the response received, even if after a very long turn of therapist guidance. If there were clear observable verbal signs of patients actively “going with” the activity this was rated positively (e.g. “Yes I can picture that now.” Rated STF +1). If the level of engagement was less clear (e.g. “Ok.”). It was rated at an STF 0 level and verbal signs of ambivalence about the activity were negatively rated (e.g. “Where are we going with this?” Rated STF -1).

Secondly, the rating of “Other” interactions had been rarely used in development work, but in the remote CBT study there were quite a lot of interactions about the use of the technology. It was agreed that these interactions would be rated as Other (OTH). This meant that the bulk of “Other” ratings were actually discussions in which the use of remote technology was being facilitated:

> **Therapist:** So if I make you presenter [sharing screen] can we both look at it?
>
> **Patient:** Yes, if I can find it.
>
> **Therapist:** Basically now you go to ‘share’, the one next to ‘connected to audio’, to be able to share your screen or share your application sharer sort of thing. [Pause 16 secs].

Patient: Okay.

Therapist: There we go.

Patient: Excellent.

[P 01023, OTH]

Lastly, the difference between information discussions (IDI) and problem analysis (PAU) interactions was further delineated, using observable language use markers: Problem analysis was defined by language markers indicating that bespoke information was being delivered, specific to the receiver. In particular, this would be analytic information that would not make sense if spoken to a room of people:

Therapist: I remember we talked about that feeling of like, I need to be in control here [ ] I was wondering if underneath that there is kind of like almost like a rule like if I'm not in control something is more likely to go wrong and-. I didn't know, well what do you think about that? Is that kind of how, the way that you view it?

Patient: [coughs] Oh yes, that's completely right.

[P 01007, PAU +1]

Whereas, information discussion interactions involved more general information – perhaps using more general pronouns (e.g. "we", "humans" - as below). This would more likely be information which would feel more relevant if spoken to a room of people:

Therapist: You're exactly right that when we, as human beings are anxious we cannot think straight. A part of the reason for that is what's called the fight or flight response, is that something you've heard? What do you already know about the fight or flight response?

Patient: Is it like when adrenalin is released?

[P 01078, IDI +1]

Notes from the coders' discussion and agreement on CICS amendments can be found in Appendix M.

6.2.8.2 The WAI-O

Given the non-significant outcome correlations with therapeutic alliance observed in the development study, the full length WAI-O (Horvath & Greenberg, 1989) was used in this larger study to maximise validity and reliability. The larger scale also offered slightly more reliable sub-scale analysis. This included 36 items equally weighted on the assessment of therapeutic bond (e.g. “There is a good understanding between the client and therapist”) agreement on tasks (e.g. “There is agreement about the steps taken to help improve the client's situation”) and agreement on goals (e.g. “There is a shared perception of the client's goals in therapy”). In this study the WAI-O is again used as an established benchmark for reliability and predictive validity*.* It also controls for overlap between therapeutic alliance and CICS-rated patient activation in predictive modelling. Raters were trained using published guidance and assessed each session on 36 items about therapeutic alliance, each rated on a 7-point scale (1 = never to 7 = always).

6.2.9 Coders

The same coders were used in this study as in the development project reported in chapters 4 and 5. All were trained in CBT and other modalities. Three of the coders were therapists in the current trial, so sessions to be rated were allocated so that therapists did not rate their own sessions.

6.2.10 Method of analysis

6.2.10.1 Sample characteristics and IRR

Included and excluded participants were compared to clarify any significant differences caused by the sampling methods. Mann-Whitney U tests were used to compare continuous variables (e.g. age). Chi-squared tests were

used to compare characteristics for categorical variables (e.g. gender) between included and excluded participants. This study provided the opportunity to assess the CICS IRR for two (as opposed to four) coders. The IRR could also be compared to an established instrument in the WAI-O. Inter-rater reliability was assessed for the CICS and WAI-O using two-way, random effects, consistency ICCs. Spearman's rho correlations were used to demonstrate initial correlations between CICS themes and residualised 12-month follow-up outcome scores. That is, outcome scores adjusted for baseline severity.

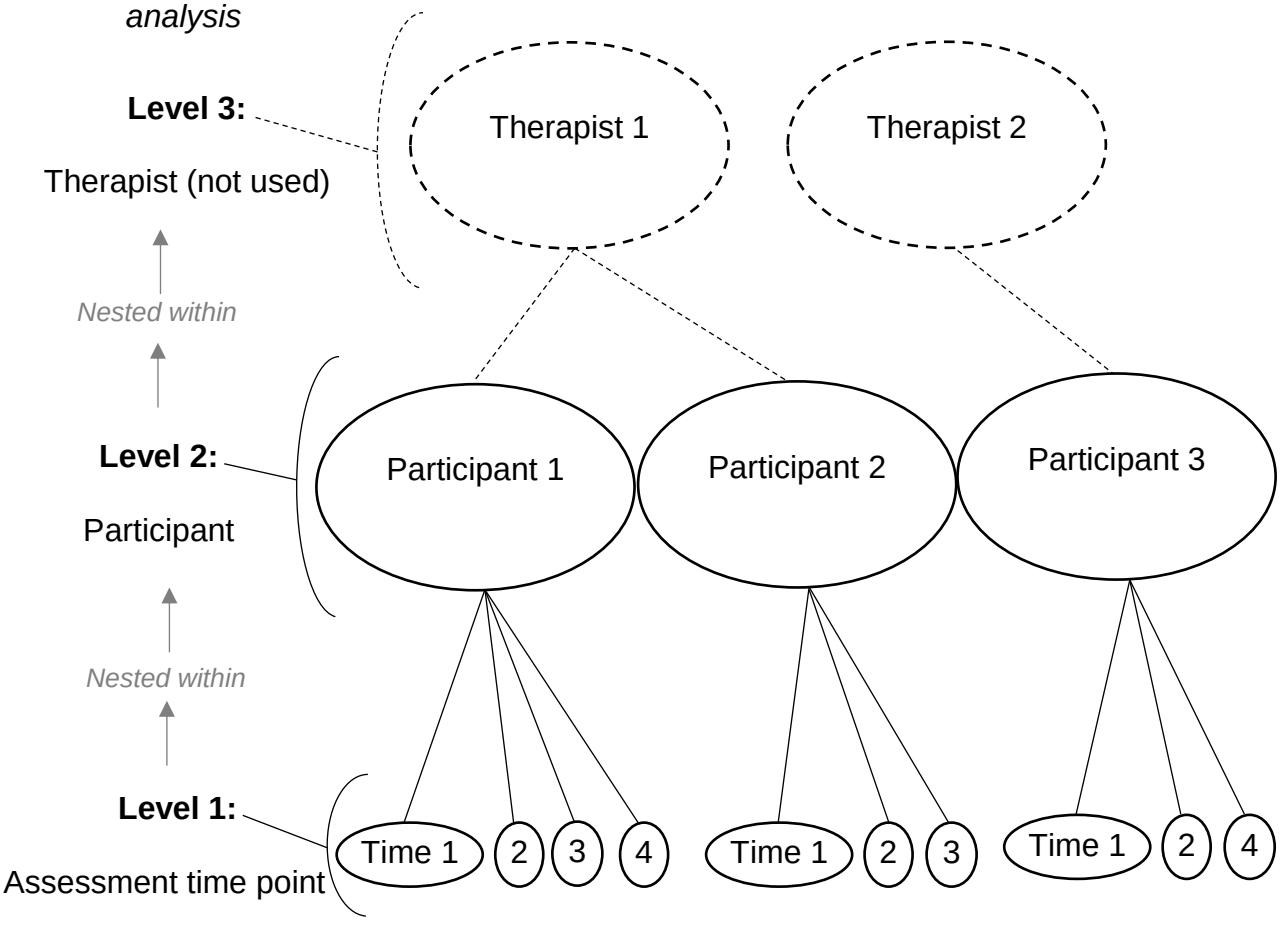
6.2.10.2 Multilevel modelling

Multilevel modelling was used to assess the predictive validity of each CICS theme and the WAI-O with follow-up scores for each outcome separately. This method was used due to the repeated measurement design. Multilevel modelling is a quantitative method of analysis that allows hierarchical nesting of data to better account for variance within the sample. Within the current study this means that change can be measured from the four repeated measurement time points *within* participants (i.e. covariance of outcome due to individual level changes in score across time) and *between* participants (i.e. changes that are explained by changes across different participants) (Tasca & Gallop, 2009). An MLM approach also has the advantage over traditional repeated measures analyses (e.g. multiple linear regression) of including cases where one or more data points are missing. Using multiple linear regression, models would be needed for each time point and would exclude cases with missing data. Furthermore, more accurate estimates can be made of change over time rather than at a single time point.

The method is also sometimes used to nest patients within the respective treating therapists (e.g. Okiishi et al., 2006). However, the current dataset did not meet existing standards for assessing therapist effects in this way (Schiefele et al., 2017). There were too few level 3 variables (only four therapists) and some had too few level 2 variables (e.g. one therapist only saw four participants).

As a three-level model was not appropriate, a two-level MLM structure was used as there was significant clustering of covariance within participants across follow-up time on all outcomes (follow-up time point [“time” hereafter] nested within patients; Figure 6.4). Covariance structures, linearity of trends and normality of data were assessed. Goodness of fit was compared using -2 log likelihood. All final models included random intercepts and fixed slopes. Parameter estimation was made using maximum likelihood.

Figure 6.4. Pictorial illustration of the hierarchical structure of data in the MLM analysis



Established MLM development methods were used for repeated measurement designs (Hox et al., 2017; Tasca & Gallop, 2009). However, as MLM is a relatively new and complex method, confusions can emerge about the steps taken in model development (Field, 2018). Therefore, each step of model development is illustrated with one model on the primary outcome, as all models were developed similarly (Table 6.3). Hierarchical structures of the data were examined in unconditional models. The significance of random slopes and random intercepts were examined in each model. Level 2 covariates were added one at a time and change in -2 log likelihood explored. However, as each variable was added for theoretical purposes, all components were kept in final models even if there was not a significant change in -2 log likelihood. For example, assessment of therapeutic alliance remained in the final model even if it was not a significant predictor of outcome. This was because assessing the predictive value of both the CICS and the WAI-O in the same model was central to addressing the study aims.

Fully adjusted models included the following level 2 predictors: *CICS theme (first session); WAI-O (first session); baseline outcome score (i.e. controlling for baseline severity); (all fixed effects)* and *time (fixed and random effects).* These models were repeated with second-session CICS themes and WAI-O scores. Centring was not used in the analysis, as CICS scores (the primary predictors of interest) do not yet have a meaningful independent value. Therefore, mean CICS scores do not have a pre-specified meaning, but they all crossed zero. So intercepts in MLM models would have a relevant value. However, time was re-scored from time 1-4 (3 month through 12 month follow-up) to 0-3. This made the intercept more meaningful, as the starting follow-up

value. The model developed in Table 6.3 is used to illustrate the equations used across outcomes and CICS themes:

$$y_{ij} = \beta_{0i} + \beta_1 \text{time}_{ij} + \beta_2 \text{Baseline SHAI}_{ij} + \beta_3 \text{Session 1 WAI-O} + \beta_4 \text{Session 1 CICS} + e_{ij}$$

Where y_{ij} is the SHAI score for the $_j$th participant at the ith follow-up time point. As such j is participants (level 2 unit) and i is follow-up time order (level 1 within level 2 unit). β_{0i} indicates that estimates of the intercept for both the overall model and variability of level 1 (i) intercepts were calculated. β_1 through β_4 were fixed level 2 effects, adjusting for differences between participants. e_{ij} is the residual or unexplained variance in the model.

Table 6.3. Steps in the development of the MLM model assessing first session structuring interactions as a predictor of health anxiety severity.

Step	Model	Deviance (-2LL)	Change in Deviance, new dfs	Parameter statistically significant?[a]	Residual variance	Intercept variance	Slope variance
1	Unconditional	1245	-	-	16.19	35.27	-
2	Fixed effects and random intercept for time	1215	30, 1	yes	12.98	35.53	-
3	Random slope for time (parameter removed)	1214	1, 1	No	12.01	34.83	.67[b]
4	Control variable 1: Fixed effect of baseline SHAI	1208	8, 1	yes	13.00	31.79	-
5	Control variable 2: Fixed effect of session 1 WAI-O	841	367, 1	yes	12.56	24.82	-
6	Fixed effect of session 1 structuring and task focus	837	25, 1	yes	12.54	22.47	-

Note. Abbreviations: -2LL: -2 log likelihood; dfs: degrees of freedom; SHAI: short health anxiety inventory; WAI-O: working alliance inventory – observer rated.
[a]Statistical significance was assessed using chi-squared with the change in -2LL and the change in dfs. [b]Variance in slope was not statistically significant (p = .312).

6.2.10.3 Sample size

Following the American psychological association's guidance for journal article reporting standards for quantitive research in psychology, a comment on sample size for MLM is included (Appelbaum et al., 2018). There is no simple, overall agreement on required sample size in MLM analyses, partly because much depends on the purpose and focus of the analysis (Kreft & De Leeuw, 1998). However, change in level 2 variables (participants) was the central focus of this analysis, so the sample size at level 2 is deemed most important (McNeish & Stapleton, 2016). Sample size recommendations at this level, for longitudinal analysis has varied from as low as 20 (Snijders & Bosker, 2012) to 50 (Maas & Hox, 2005). Given these recommendations, it was deemed that a suitable sample size was achieved for robust models. However, even with a sample of 50 models can still be overloaded with excessive predictors. Therefore, only essential variables for the research questions of interest were added as level 2 covariates. For example, age and gender might be important predictors, but given the sample size and the exploratory nature of this study they were not included as covariates.

Estimates of required sample size can be obtained retrospectively. However, post-hoc sample size estimation is strongly discouraged, particularly in circumstances where the sample size cannot be changed (Hoenig & Heisey, 2001). As this study is a secondary analysis of data where the sample size was powered for an RCT and further data collection was not possible, post-hoc sample size was not estimated. As recommended, it is acknowledged that the sample was not powered for the current analysis and therefore results should be treated with suitable caution (Hoenig & Heisey, 2001). Nonetheless, a

sample of 50 is at the top end of minimum sample sizes for MLM analysis, so the analysis can still be conducted with some confidence.

6.2.10.4 Missing data

Missing data were imputed in the original trial, but these data were imputed on the basis of the overall sample, including participants who were not offered CBT. Therefore, missing data was not imputed and only observed values were used in the analysis reported.

Analysis was carried out with SPSS 24. Ethical approval was obtained from the National Research Ethics Service, London-Riverside Committee (reference 14/LO/1102).

6.3 Results

6.3.1 Sample characteristics

The 50 participants included in the study did not differ significantly from excluded participants on any demographic or clinical characteristic (Table 6.4). The included sample had a larger proportion of female participants and a lower mean age than has been typical in previous trials of CBT for health anxiety (e.g. Tyrer et al., 2014: mean age 50; 52% female). However, representation of minority ethnic groups was similar. When applying the clinical thresholds for the clinical outcome measures, baseline clinical characteristics indicated that mean scores for all disorder-specific assessments fell within the respective measures' clinical ranges. On average, participants were experiencing, severe health anxiety, moderate severity clinical depression, generalised anxiety and somatic symptoms, and clinically significant function impairment. No clinical cut-offs are identified for the quality of life (EQ5D) and general health (VAS) measures, but mean general health scores fell below the 25th centile of

normative data for even those aged over 70 (van Reenen & Janssen, 2015). This indicates very poor self-reported general health at baseline.

On average participants met the criteria for seven mental disorders on the SCID with generalised anxiety disorder and somatoform disorders both affecting more than a third of the sample alone. The sample also reported a mean of one chronic physical health condition, with chronic pain affecting more than a quarter of patients (26%). Overall, this suggests that the sample predominantly experienced multiple chronic physical and mental health problems. This type of multimorbidity where patient activation is identified as an important prognostic factor (Blakemore et al., 2016).

6.3.2 Inter-rater reliability

The CICS achieved ICCs between .71 and .97 across interaction themes, indicating good-to-excellent IRR overall for two coders (Table 6.5; Cicchetti, 1994). In particular, good IRR was achieved in the information discussion (IDI) theme, with a confidence interval in the good to excellent range. This suggests that the issues identified in the development work were overcome and sufficient data were coded as IDI to warrant outcome prediction. Despite ICCs being retained within the good-to-excellent range for two coders, the global scales had notably lower agreement, with the weakest agreement for all themes being integration of themes (IOT).

The WAI-O demonstrated generally lower IRR than the CICS. The subscales of the WAI-O achieved ICCs of .78 (goal), .70 (task) and .46 (bond) with an ICC of .69 for the overall score (Table 6.6). This indicated excellent, good and fair IRR respectively with good IRR overall (Cicchetti, 1994).

Table 6.4. Demographics and clinical characteristics of included and excluded participants.

Demographics	*Included* (n = 50)	*Excluded* (n = 29)
Females	39 (78%)	18 (62%)
Mean age (SD)	36 (15)	42 (20)
Ethnicity		
White British	38 (76%)	21 (72%)
Other	12 (24%)	8 (28%)
Unemployed	6 (12%)	6 (21%)
Clinical Characteristics		
Baseline SHAI (SD)	27.0 (5.2)	28.0 (5.7)
Baseline PHQ-9 (SD)	13.2 (6.3)	13.5 (6.6)
Baseline GAD-7 (SD)	12.9 (5.7)	13.2 (5.3)
Baseline PHQ-15 (SD)	14.5 (5.3)	13.0 (5.8)
Baseline WSAS (SD)	17.3 (11.4)	17.3 (11.1)
Baseline EQ5D-5L utility Index (SD)	0.614 (.261)	0.602 (.312)
Baseline VAS (SD)	54.0 (21.7)	54.9 (22.6)
Mean number of SCID Diagnoses (range)	7 (0-16)	5 (0-11)
Generalised anxiety disorder	34 (68%)	19 (66%)
Somatoform disorders	33 (66%)	12(41%)
Hypochondriasis	32 (64%)	15 (52%)
Current depressive episode	31 (62%)	15 (52%)
Panic disorder	31 (62%)	12 (41%)
Other disorders	48 (96%)	26 (90%)
Mean chronic physical health problems (range)	1 (0-6)	1 (0-4)
Chronic pain	13 (26%)	6 (21%)
Functional neurological disorders	5 (10%)	0
Chronic fatigue	5 (10%)	1 (3%)
Irritable bowel syndrome	4 (8%)	2 (7%)
Diabetes	4 (8%)	0

Note. Abbreviations: SD, standard deviation; SHAI, short form health anxiety inventory; PHQ-9, patient health questionnaire 9-items; GAD-7, generalised anxiety disorder 7-items; PHQ-15, patient health questionnaire 15-items; WSAS, work and social adjustment scale; EQ5D, Euroqol 5-dimensions 5-levels; VAS, visual analogue scale; SCID, structured clinical interview for the diagnostic and Statistical Manual of Mental Disorders IV.

Table 6.5. Consultation Interactions Coding Scheme IRR.

CICS Theme	ICC	95% CI		Agreement descriptor[a]
		Lower	Upper	
Global Ratings				
Collaborative Flow (COF)	0.73	0.59	0.82	Good
Integration of Themes (IOT)	0.66	0.49	0.78	Good
Totals				
Combined content ratings	0.76	0.63	0.84	Excellent
Combined global ratings	0.72	0.58	0.82	Good
Content Ratings				
Other (typically facilitation of remote therapy)	0.97	0.95	0.98	Excellent
Noticing Change or Otherwise (*NCO*)	0.9	0.85	0.93	Excellent
Evaluations of Self or Therapy (*EST*)	0.85	0.78	0.9	Excellent
Problem and Context Description (*PCD)*	0.79	0.69	0.86	Excellent
Action Planning and Idea Generation (*API*)	0.75	0.61	0.83	Excellent
Information Discussion (*IDI*)	0.74	0.61	0.83	Good
Structuring and Task Focus (*STF*)	0.71	0.55	0.81	Good
Problem Analysis and Understanding (*PAU*)	0.71	0.57	0.81	Good

Note. Abbreviations: CICS: Consultation interactions coding scheme; CI: Confidence Interval; ICC: intra-class correlation coefficient.

[a]Established using Cicchetti's (1994) qualitative descriptors of ICCs.

Table 6.6. Observer-rated working alliance inventory IRR

WAI-O Scale	ICC	95% CI		Agreement descriptor[a]
		Lower	Upper	
Goals	0.78	0.66	0.85	Excellent
Task	0.70	0.57	0.81	Good
Total scale	0.69	0.53	0.8	Good
Bond	0.46	0.2	0.65	Fair

Note. Abbreviations: WAI-O: Working alliance inventory – observer rated; CI: Confidence Interval; ICC: intra-class correlation coefficient.

[a]Established using Cicchetti's (1994) qualitative descriptors of ICCs.

6.3.3 Outcome correlations

Correlations between session 1 CICS scores and residualised 12-month follow-up outcomes indicated a consistent association between greater patient activation in interactions where changes were noticed (NCO) and improvement in five of the seven outcomes (Table 6.7). No other CICS themes showed any similarly consistent correlations. Session 2 outcome correlations indicated significant moderate associations between noticing change interactions and all outcomes (Table 6.8). Multiple 12-month outcomes were correlated with problem description (PCD), problem analysis (PAU) and general/facilitative interactions (OTH). The global theme integration of themes (IOT) was also significantly correlated with multiple outcomes. Action planning interactions, collaborative flow and therapeutic alliance did not correlate with any outcomes from either first or second session ratings.

Table 6.7. Session 1 CICS and WAI-O correlations with 12-month follow-up outcomes

	SHAI	EQ5D	GAD-7	WSAS	VAS	PHQ-9	PHQ-15
Action Planning and Idea Generation (*API*)	.02	-.148	.263	-.046	-.246	.101	.094
Evaluations of Self or Therapy (EST)	-.337	.300	-.395*	-.125	.359*	-.239	-.182
Information Discussion (IDI)	-.109	.265	-.333	-.122	.574**	-.165	-.181
Noticing Change or Otherwise (NCO)	-.206	.359*	-.472**	-.510**	.542**	-.385*	-.271
Problem Analysis and Understanding (PAU)	-.172	.266	-.144	-.341	.166	-.299	-.316
Structuring and Task-Focus (STF)	-.163	-.121	.051	-.035	.019	-.13	-.052
Other (general/facilitative interactions)	-.196	-.082	-.103	-.135	-.012	-.126	-.124
Problem and Context Description (PCD)	.272	.007	.044	.027	-.113	.185	.113
Collaborative Flow (COF)	-.087	.302	-.103	-.296	.158	-.100	-.285
Integration of Themes (IOT)	-.149	.081	-.064	-.192	.044	-.127	-.115
Working Alliance Inventory – Observer rated (WAI-O)	-.127	.266	-.208	-.287	.177	-.247	-.332

Note. *P<.05 **p<.01. Abbreviations. SHAI: Short health anxiety inventory; EQ5D: Euroqol – 5 dimensions, 5 – levels; GAD-7: Generalised anxiety disorder – 7 items; WSAS: Work and social adjustment scale; VAS: visual analogue scale; PHQ-9: Patient health questionnaire – 9 items; PHQ-15: Patient health questionnaire – 15 items.

Table 6.8. Session 2 CICS and WAI-O correlations with 12-month follow-up outcomes

	SHAI	EQ5D	GAD-7	WSAS	VAS	PHQ-9	PHQ-15
Action Planning and Idea Generation (*API*)	-.019	.212	-.152	-.187	.189	.040	-.049
Evaluations of Self or Therapy (EST)	-.260	.033	-.079	-.189	.079	-.085	-.242
Information Discussion (IDI)	-.260	.238	-.119	.118	.276	-.065	-.350*
Noticing Change or Otherwise (NCO)	-.369*	.339*	-.403*	-.374*	.402*	-.407*	-.470**
Problem Analysis and Understanding (PAU)	-.431**	.379*	-.269	-.363*	.419**	-.232	-.359*
Structuring and Task-Focus (STF)	-.171	-.005	-.050	-.353*	-.084	-.225	-.092
Other (typically facilitation of remote therapy)	.007	.249	-.382*	.179	.269	-.441**	-.302
Problem and Context Description (PCD)	.340*	-.359*	.308	.181	-.421**	.183	.066
Collaborative Flow (COF)	.003	.185	-.061	-.141	.133	-.115	-.267
Integration of Themes (IOT)	-.130	.377*	-.165	-.229	.331*	-.225	-.338*
Working Alliance Inventory – Observer rated (WAI-O)	-.196	.026	-.191	-.229	.112	-.049	-.195

Note. *P<.05 **p<.01. Abbreviations. SHAI: Short health anxiety inventory; EQ5D: Euroqol – 5 dimensions, 5 – levels; GAD-7: Generalised anxiety disorder – 7 items; WSAS: Work and social adjustment scale; VAS: visual analogue scale; PHQ-9: Patient health questionnaire – 9 items; PHQ-15: Patient health questionnaire – 15 items.

6.3.4 Predictive MLM

6.3.4.1 Normality assumptions

Similar assumptions of normality are taken in MLM as multiple linear regressions: normal distribution of data (particularly normal distribution of the model's unstandardized residuals); linearity of relationship (that the relationship between predictor and outcome is linear); homoscedasticity (that the same error variance occurs around the outcome at each predictor value), and independence of variance (that error variances are independent of one another). From the normality assessment of the models reported, the assumption of normality was violated as residuals were slightly skewed. Ghasemi and Zahediasl (2012) have argued that the normality assumption can be dismissed if hundreds of data points are included, except in extreme cases (more than 200 data points were included in the models analysed). Furthermore, examination of histograms and Q-Q plots are required alongside statistical tests of normality, given the tests' propensity to be over-sensitive in larger samples (Field, 2018; Raudenbush & Bryk, 2002). Histograms indicated approximate normality, Q-Q plots indicated homoscedasticity (Q-Q plots show the observed variance against the expected variance at each data point, if the distance between expected and observed is relatively uniform across data points homoscedasticity has been achieved). Nevertheless, a bootstrap of 1,000 samples was used to more accurately estimate 95% confidence intervals on all models, as is recommended under conditions where normality is violated (Davison & Hinkley, 1997). The bootstrapped results confirmed all results from the observed data and are presented in Appendix N.

6.3.4.2 Modelling results

In unconditional models for all outcomes 21% to 41% of the covariance was explained by clustering within patients, which was significant for all models. This indicated that a two-level hierarchical structure was appropriate to account for this effect. Overall, 42 participants' first sessions and 49 participants' second sessions were rated and included in predictive models. As several models were run, a strategic approach to analysis was taken, which sought to best achieve the study's aims. Results were focused on CICS themes where the theme coefficients in final models showed at least a trend in the predicted direction across all seven outcomes. That is, CICS themes which were negatively correlated with all assessments of disorder severity (disorder symptoms reduced as patient activation increased) and positively correlated with the EQ5D and VAS (quality of life and general health increased as patient activation increases). Furthermore, only CICS themes that predicted either the primary outcome or multiple secondary outcomes were considered meaningful results. This approach aimed be sufficiently open to novel findings for an exploratory study, but at the same time protect against chance significant results. Three CICS themes met the criteria described and will each be addressed in turn:

Positive interactions about changes that patients noticed at the second session of CBT significantly predicted every clinical outcome measured over time (Table 6.9).

Therefore, when more of the second session was spent with patients actively identifying changes that had occurred and discussing them, greater improvements were observed across every clinical outcome:

> **Therapist:** Can you tell me a bit about that, what's different?
>
> **Patient:** I just feel more relaxed in myself and, like, some things are just clicking into place, like I've got loads of things that I've planned that I'm going to do, like, I have loads of things to look forward to. I think I'm also thinking about the New Year and to try and, like, change the way I think about things.
>
> [P 01035, NCO +1]

Conversely, negative interactions about change at the second session were predictive of worse outcome, long-term.

> **Patient:** I did the thing where you said not to go to the health centre, like, for small things and things like that, but I feel like that's backfired a bit! [ ]
>
> **Therapist:** Right. So it might not have helped you [ ] what is it that makes it feel like it's backfired?
>
> **Patient:** It probably wouldn't have helped me but I just feel like I did something different and now I'm ill, so I feel like that it was a bad thing to do.
>
> [P 01002, NCO -1]

This was the case despite positively scored interactions about change only taking up an average of 3.0% of spoken words at session two (range: 0-21.6%; SD: 4.2%). Interactions about negative changes noticed (or the absence of change) took up only 0.2% of spoken words (range: 0-2.9%; SD, 0.5).

Table 6.9. Multilevel model of noticing change interactions at session two predicting all health outcomes across time.

Health anxiety (SHAI)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	11.48	8.47	.182	-5.55	28.51
Assessment time point	-1.56	0.27	<.001	-2.10	-1.02
Baseline severity (SHAI)	0.56	0.15	.001	0.25	0.86
Therapeutic alliance at session 2 (WAI-O)	-0.87	1.14	.449	-3.16	1.42
Interactions about changes noticed at session 2 (CICS)	*-0.40*	*0.20*	*.048*	*-0.80*	*-0.004*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	13.61	1.84	7.40	<.001
Intercept variance	23.77	5.77	4.12	<.001

Note. Abbreviations: SHAI: short health anxiety inventory; B: unstandardized β coefficient; SE: standard error; p: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme; Z: Walt's z test for significance of variance.

Quality of life (EQ5D-5L)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	0.31	0.22	.170	-0.14	0.76
Assessment time point	0.02	0.01	.017	0.00	0.04
Baseline QoL (EQ5D)	0.64	0.11	<.001	0.42	0.85
Therapeutic alliance at session 2 (WAI-O)	-0.02	0.04	.553	-0.10	0.06
Interactions about changes noticed at session 2 (CICS)	*0.02*	*0.01*	*.020*	*0.00*	*0.03*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	0.02	0.00	6.94	<.001
Intercept variance	0.03	0.01	3.91	<.001

Note. Abbreviations as above plus: QoL: quality of life; EQ5D: Euroqol – 5 dimensions – 5 levels

Generalised anxiety (GAD-7)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.75	5.49	.300	-5.27	16.76
Assessment time point	-1.02	0.24	<.001	-1.49	-0.55
Baseline severity (GAD-7)	0.41	0.11	<.001	0.20	0.62
Therapeutic alliance at session 2 (WAI-O)	-0.18	0.85	.835	-1.87	1.52
Interactions about changes noticed at session 2 (CICS)	*-0.50*	*0.15*	*.002*	*-0.80*	*-0.20*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	9.73	1.34	7.28	<.001
Intercept variance	10.46	2.77	3.77	<.001

Note. Abbreviations as above plus: GAD-7: generalised anxiety disorder – 7 items

Depression (PHQ-9)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.68	6.09	.356	-6.56	17.92
Assessment time point	-0.56	0.23	.018	-1.02	-0.10
Baseline severity (PHQ-9)	0.49	0.11	<.001	0.28	0.70
Therapeutic alliance at session 2 (WAI-O)	-0.19	0.95	.846	-2.11	1.73
Interactions about changes noticed at session 2 (CICS)	*-0.51*	*0.17*	*.003*	*-0.84*	*-0.18*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	9.36	1.31	7.17	<.001
Intercept variance	14.97	3.77	3.97	<.001

Note. Abbreviations as above plus: PHQ-9: patient health questionnaire – 9 items.

Somatic symptoms (PHQ-15)

Fixed effects					
				95% CI	
Parameter	B	SE	*p*	Lower Bound	Upper Bound
Intercept	8.13	4.58	.083	-1.12	17.37
Assessment time point	-0.91	0.21	<.001	-1.33	-0.49
Baseline severity (PHQ-15)	0.61	0.09	<.001	0.42	0.79
Therapeutic alliance at session 2 (WAI-O)	-0.52	0.71	.467	-1.96	0.92
Interactions about changes noticed at session 2 (CICS)	*-0.35*	*0.13*	*.008*	*-0.60*	*-0.09*

Random effects				
Parameter	Variance	SE	Z	*p*
Residual variance	8.22	1.16	7.10	<.001
Intercept variance	8.43	2.51	3.36	0.001

Note. Abbreviations as above plus: PHQ-15: patient health questionnaire – 15 items.

Functional impairment (WSAS)

Fixed effects					
				95% CI	
Parameter	B	SE	*p*	Lower Bound	Upper Bound
Intercept	6.48	8.45	.447	-10.51	23.46
Assessment time point	-1.43	0.38	<.001	-2.19	-0.68
Baseline severity (WSAS)	0.73	0.08	<.001	0.56	0.89
Therapeutic alliance at session 2 (WAI-O)	-0.62	1.40	.660	-3.45	2.20
Interactions about changes noticed at session 2 (CICS)	*-0.85*	*0.24*	*0.001*	*-1.34*	*-0.37*

Random effects				
Parameter	Variance	SE	Z	*p*
Residual variance	24.53	3.42	7.17	<.001
Intercept variance	33.81	8.79	3.85	<.001

Note. Abbreviations as above plus: WSAS: work and social adjustment scale.

General health (VAS)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	53.05	17.76	.004	17.41	88.70
Assessment time point	0.62	0.78	.430	-0.93	2.16
Baseline health (VAS)	0.37	0.10	.001	0.16	0.57
Therapeutic alliance at session 2 (WAI-O)	-2.30	3.25	.483	-8.82	4.22
Interactions about changes noticed at session 2 (CICS)	*1.77*	*0.53*	*.002*	*0.70*	*2.84*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	108.65	14.72	7.38	<.001
Intercept variance	167.21	41.01	4.08	<.001

Note. Abbreviations as above plus: VAS: visual analogue scale.

A possible reason for the importance of noticing change interactions is explored below, using a case example and extended session extract.

Participant 6004[7] was a middle-aged man who had been on long-term sick leave due to health anxiety and associated social anxiety. He worried that he might choke or have a heart attack and die, particularly when in social settings, as his uncle had done. This fear was fuelled by anxiety symptoms which included a racing heart, dizziness and throat tightening. In the extract below he describes a change he has noticed. This extract illustrates the use of coded transcript data to build explanatory frameworks for outcome prediction, by helping to explain how and why noticing change interactions might be so

[7] Some elements of the case material are disguised to protect the participant's anonymity.

consistently associated with outcome. This extracted is one of several examples showing that where noticing change interactions occurred, they were often followed by focused, personalised action plans. That is, the specific changes already noticed are built upon for future therapeutic changes. As such, therapist guidance is no longer abstract, but bespoke to the changes already started by the specific patient:

Therapist: How did that go then going out for lunch with [your girlfriend] last week?

Patient: Yes, it wasn't too bad to be honest, I did feel slightly anxious at times but not anything- It didn't go into panic like symptoms so it was more feeling uncomfortable but I kind of just ploughed through it and did what I was doing and now after five minutes or so it subsided and I could carry on and enjoy myself. It was just getting over that initial- I think because I hadn't done it for so long, it has probably been six or seven months since we'd actually eaten out together.

Therapist: Really?

Patient: It was something before that we used to do on a weekly basis, we would always go out for food together, it was one thing we used to do quite a lot. It was quite strange actually being in a situation where I didn't want to be doing that but hopefully now that- Like I say with my three [days] off I'm trying to make the most of any time and push myself to do a bit more.

Therapist: That sounds like a massive hurdle to have got over to have gone out for lunch last week.

Patient: Yes, it was quite a big thing because, like I say, I hadn't felt anxious about things like that because I totally dismissed them as, "No, I'm not doing that," and that was kind of my initial reaction. Previously when we [said], "We've got this night out," or, "We're doing this," straight away I'd be looking for excuses why I couldn't go.

Therapist: Yes.

Patient: I think to a point I still do, I think I'm certain not 100% where I'm excited to go and do things and I think that's the difference. Where I'd like to be is it would be nice for her to say, "Right, we're going to do this in two weeks' time," and rather than feeling the dread to actually, "Oh, yes, that would be nice, I can look forward to that." I think that's where I am at the moment so I'm not at the extreme end with it where I'm running away from it straight away but I'm certainly not looking forward to anything at the moment.

This extract shows the patient moving from a description of a therapeutic change, to an extension of the change, to plans for the future, alongside articulation of a different perspective as a result. The patient identified that “It was quite strange” not to want to go out, which suggests that his perspective and beliefs about the situation may have already changed. He described “trying to make the most of any time” as part of his future plan. Perhaps most clearly he articulates a very personalised goal: “Where I’d like to be is it would be nice for her to say, ‘Right, we’re going to do this in two weeks’ time’ and [ ] to actually ‘Oh yes, that would be nice…” Although this occurred at only the second of ten CBT sessions, such clear personally relevant goals might have helped the therapist target therapeutic interventions in a way that could make the sessions have greater clinical impact than if the changes had not been described or elicited. This type of interaction might help therapist and patient to collaborate about plans and actions that would make the most meaningful difference to the patient, by building on early changes already noticed. Therefore, a potential explanation of the value held by noticing change interactions is that they sharpen the focus of future interventions and plans.

The greater the proportion of the second session that was spent describing problems and their context, the poorer the outcome in terms of generalised anxiety and general health. This means that there were trends for increased anxiety severity and decreased general health as the percentage of session two interactions spent describing problems increased (PCD; Table 6.10). This theme was the most coded of all CICS themes, with an average 48.5% of session 2 spoken words coded as problem and context description (range: 17.4-74.1%; SD: 14.4). Problem and context descriptions were not

deemed to be negative in and of themselves, but larger proportions of sessions spent describing problems was viewed as negative by drowning out opportunities for more active interactions:

> **Patient:** So that's been a bit hard work this week. Trying to sort of get new things into place with [my son] and get new things into place with me and [my partner] and of course it's caused arguments between us because we've not really agreed on -, we have very different ways of parenting so we have to try and find a bit of middle ground really!
>
> **Therapist:** OK, so kind of negotiating that's been like hard work by the sounds of it.
>
> [P 1051, PCD]

As with noticing change interactions, there were trends in the predicted directions for all outcomes, but only generalised anxiety and general health models reached significance when fully adjusted. Notably, this meant that trends were the reverse of positively scored themes: The percentage of problem description interactions increased with follow-up disorder severity (positively correlated) and decreased with follow-up general health and quality of life improvement.

Table 6.10. Multilevel model of problem and context description interactions at session two predicting follow-up general health and generalised anxiety.

Generalised anxiety (GAD-7)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.75	6.03	0.345	-6.37	17.86
Assessment time point	-1.01	0.24	0.000	-1.48	-0.54
Baseline severity (GAD-7)	0.30	0.10	0.005	0.09	0.51
Therapeutic alliance at session 2 (WAI-O)	-0.90	0.85	0.294	-2.60	0.80
Problem descriptions at session 2 (CICS)	*0.09*	*0.04*	*0.035*	*0.01*	*0.17*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	9.81	1.35	7.24	<.001
Intercept variance	11.88	3.10	3.83	<.001

Note. Abbreviations: GAD-7: generalised anxiety disorder – 7 items; B: unstandardized β coefficient; SE: standard error; *p*: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme; Z: Walt's z test for significance of variance.

General health (VAS)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	64.29	20.67	.003	22.78	105.80
Assessment time point	0.51	0.78	.514	-1.04	2.06
Baseline severity (VAS)	0.29	0.11	.010	0.07	0.50
Therapeutic alliance at session 2 (WAI-O)	0.06	3.32	.986	-6.61	6.73
Problem descriptions at session 2 (CICS)	*-0.32*	*0.15*	*.040*	*-0.62*	*-0.01*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	108.97	14.80	7.36	<.001
Intercept variance	190.88	45.97	4.15	<.001

Note. Abbreviations as above plus VAS: visual analogue scale.

Unexpectedly, the other theme which predicted multiple clinical outcomes and had intercepts showing trends in the expected direction for all outcomes was “other” interactions. That is, interactions where therapists were supporting patients to use remote technology:

> **Therapist**: If you click [the mouse] I think you should be back in control.
>
> **Patient**: [laughs] Yeah.
>
> **Therapist**: [laughs] Is it not doing it? OK, maybe not. Oh there you go.
>
> **Patient**: There we go.

[01093, OTH]

Alternatively there was general chat at the very beginning or end of a session:

> **Therapist**: I’m just looking because I can’t make next week so it will be two weeks on Tuesday at 11 o’clock.
>
> **Patient**: Have I screwed your brain up that much?
>
> **Therapist**: [Laughs]. Not at all. No, I’m looking for more, my brain is ready for more.
>
> **Patient**: You’re like me, in a way I’m like that, it’s the main reason I’ve got a computer, I like knowledge.

[01008, OTH]

The greater the proportion of the second session taken up with this type of interaction, the greater the improvement in symptoms of anxiety and depression over time (Table 6.9). These types of “other” interactions took up an average of 4.1% of spoken words at session two (range: 0.5-11.5%; SD: 2.5).

Table 6.11. Multilevel model of facilitative and general interactions at session two predicting follow-up depression and generalised anxiety.

Depression (PHQ-9)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	14.44	6.22	.024	1.96	26.93
Assessment time point	-0.52	0.23	.029	-0.98	-0.05
Baseline severity (PHQ-9)	0.42	0.10	<.001	0.21	0.63
Therapeutic alliance at session 2 (WAI-O)	-1.29	0.94	.176	-3.18	0.60
Facilitative interactions at session 2 (CICS)	*-0.68*	*0.26*	*.010*	*-1.20*	*-0.17*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	9.35	1.30	7.18	<.001
Intercept variance	15.78	3.92	4.03	<.001

Note. Abbreviations: PHQ-9: patient health questionnaire – 9 items; B: unstandardized β coefficient; SE: standard error; p: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme; Z: Walt's z test for significance of variance.

Generalised anxiety (GAD-7)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	16.08	5.62	.006	4.80	27.35
Assessment time point	-0.97	0.24	<.001	-1.44	-0.50
Baseline severity (GAD-7)	0.25	0.10	.018	0.04	0.45
Therapeutic alliance at session 2 (WAI-O)	-1.39	0.82	.098	-3.04	0.26
Facilitative interactions at session 2 (CICS)	*-0.62*	*0.22*	*.007*	*-1.07*	*-0.18*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	9.89	1.37	7.21	<.001
Intercept variance	10.73	2.92	3.68	<.001

Note. Abbreviations as above plus GAD-7: generalised anxiety disorder – 7 items.

Specific comment is made about the structuring and task-focus interaction theme, given the hypothesis made about its predictive value. Although there were consistent trends for structuring and task focused interactions, it was not significant in any fully adjusted models. First session structuring interactions did predict change in health anxiety until working alliance was added to the model, when it drop below significance (Table 6.10). Interestingly, a large proportion of session one was coded as structuring and task-focused by comparison to noticing change interactions and "other" interactions: 17.7% on average (range: 5.3-40.7%; SD: 6.9).

Table 6.12. Multilevel model for structuring and task focused interactions at session one as a predictor of health anxiety

Health anxiety (SHAI)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	14.53	8.48	.094	-2.57	31.63
Assessment time point	-1.55	0.28	<.001	-2.10	-1.00
Baseline severity (GAD7)	0.42	0.16	.014	0.09	0.75
Therapeutic alliance at session 1 (WAI-O)	-0.53	1.18	.657	-2.90	1.85
Structuring interactions at session 1 (CICS)	*-0.18*	*0.09*	*.052*	*-0.36*	*0.00*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	12.54	1.79	7.01	<.001
Intercept variance	22.47	5.82	3.86	<.001

Note. Abbreviations: SHAI: short health anxiety inventory; B: unstandardized β coefficient; SE: standard error; p: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme; Z: Walt's z test for significance of variance.

No other CICS themes met the above criteria of consistent predictive trends, multiple outcome predictions or prediction of the primary outcome. The

WAI-O did not predict any outcomes consistently from either first or second session ratings. The same was found whether CICS ratings were controlled in WAI-O models or not.

6.4 Discussion

Following the initial development in chapters 4 and 5, this chapter has presented a more thorough evaluation of the CICS' predictive validity in distal, long-term outcomes. This study indicates that interaction-level patient activation predicts multiple, long-term health outcomes from as early as the second session of CBT for health anxiety. However, this effect was heavily weighted onto specific types of interaction. Specifically, interactions about changes that had been noticed or where the absence of change had been noticed, were the most consistent predictor of all outcomes. This was the case despite these interactions taking up less than 5% of all speech at the second session. A possible explanation for this effect may lie in the observation that subsequent action planning was more personalised, so perhaps more meaningful and impactful for the patient. Interactions where problems were described predicted both generalised anxiety and general health over time. Problem description interactions also showed expected trends for all other outcomes. In an unanticipated finding, the percentage of the second session spent in facilitative or general conversation predicted improvement in depression and generalised anxiety. Again there were similar trends across all outcomes. The initial importance placed on structuring and task focused interactions did not translate into predictive validity in this study. There were expected trends for all outcomes from session one structuring ratings, but none were significant.

Taken together, this study suggests therapists should initiate and facilitate discussions about changes that may be noticed very early in psychological therapy. Conversely, therapists should be aware of making space for more than the description of problems in early sessions of psychological therapy, promoting more active discussion and interventions. Lastly, it seems that taking the time to facilitate the use technology early on when delivering psychological therapy remotely can have a beneficial impact on treatment outcome.

6.4.1 Strengths

This study highlights the importance and value of interaction themes early in CBT. The observational methods used help to overcome some of the inherent problems with self-report measures. This study also indicates the specific themes which may be of particular importance at that critical, early stage in therapy. The second session of CBT might be seen as too early in therapy, where little is occurring at the interaction level that might affect long-term outcomes. This study suggests that certain types of interaction make a positive outcome more likely than others. High rates of multimorbidity amongst the sample suggest that these findings have potential relevance across other conditions.

As this study controlled for baseline scores on all outcomes, it suggests that the findings are not an artefact of initial health and functioning. For example, one might expect that those who were less severely anxious or depressed prior to therapy might show higher patient activation at initial sessions. The study design suggests the findings are unlikely to be explained by this phenomenon. Furthermore, this study suggests that CICS ratings

measure something different to therapeutic alliance, given that all predictive models controlled for alliance ratings of the same sessions. Therefore, CICS ratings may be seen to have a value independent of therapeutic alliance. This means patient activation in early CBT interactions may be an important additional factor in relation to eventual outcome.

6.4.2 Limitations

It is notable that no interaction themes predicted outcome repeatedly from the first session. There were also several CICS themes that showed no predictive validity. The design of the study does not make it clear how these null findings should be received. It is possible that the non-predictive CICS themes may become more important later in therapy (as would likely be the case with therapeutic alliance), but some may be redundant. Longitudinal, session-by-session analysis would illuminate this issue. Despite the strategic attempts to account for type 1 errors in the results, there remains a risk of chance associations due to the multiple testing in this study. Therefore, results from problem description and facilitative interactions should be treated with some caution. However, noticing change interactions remain significant even at an alpha of 0.01 for five of the seven outcomes. This seems a robust and replicated finding. Methods have been developed for characterising the effect size for predictive factors in MLM (e.g. Raudenbush & Bryk, 2002). However, there is no agreed method and attempting to determine effect sizes in MLM can be inaccurate or misleading, so were not used in this study (Hox et al., 2017). As a result, it is difficult to determine the clinical significance of the results, because the extent of predictive validity is not quantified. Nonetheless, this study provides evidence that this area is worthy of additional future

investigation and can contribute to practical decisions about topics discussed early in psychological therapy. Effect size estimates are used in the following chapter to give an indication of effects across CICS themes from proximal outcomes. As this study did not include therapist level effects in the design, it is possible that the findings are simply characteristic of a small unrepresentative sample of therapists who may operate differently from the majority. That said, all therapists were routinely rated for CBT competence and evaluated as falling within the “competent” to “proficient” range on the CTS-R.

6.4.3 Conclusion

This chapter has presented the predictive validity of the CICS with distal outcomes, which are perhaps the most important. However, the value of the CICS is also illuminated through an understanding of more proximal outcomes. The next two chapters address the CICS’ predictive validity with proximal outcomes, using session-by-session general outcome ratings to evaluate session-level change from CICS scores. In particular, the following chapter focuses on the CICS’ ability to predict session-by-session outcomes and sudden gains using the same dataset as this chapter.

7. THE CICS' SENSITIVITY TO PROXIMAL OUTCOMES

7.1 Introduction

The previous chapter evaluated the CICS' predictive validity with distal outcomes: long-term change in health status over several months. This chapter addresses the CICS sensitivity to more immediate clinical changes. Overall, this chapter asks: Do relevant changes in interaction-level patient activation (as rated with the CICS) precede clinically important changes soon afterwards? The question is answered in two ways: Firstly, MLM is used to evaluate whether early CICS ratings predict weekly session-by-session changes in wellbeing (as opposed to assessments over several months). Secondly, time series data is used to evaluate whether there are significant changes in patient activation at the session directly prior to a sudden, substantial and stable improvement in wellbeing. Establishing temporal precedence is an essential component in determining causality. Demonstrating change in psychological processes that directly and immediately precede outcome change is an important means of evidencing potential mediation in CBT (Lorenzo-Luaces et al., 2015). This chapter will provide evidence of the interaction-types likely to influence outcome during CBT for health anxiety. This chapter will also demonstrate what type of interactions may play an important role in sessions directly before sudden improvements in outcome.

In order to comprehensively describe both components of the chapter, the relationship between early CICS ratings and session-by-session outcome will be described first (Study 1). This will be followed by a description of changes in patient activation that precede sudden, stable improvements in session-by-session rated outcome (Study 2).

7.2 Study 1: Using early CICS scores to predict session-by-session outcomes

Ilardi and Craighead (1994) initiated and emphasised the value of time series analysis in psychological therapy. They demonstrated that change in outcome varies significantly between individual patients and that there can be changes within therapy that are not captured by follow-up assessments. Similar findings have been replicated in different ways with different psychological therapies (Tang & DeRubeis, 1999a, 1999b; Tang et al., 2005; Tang, Luborsky, & Andrusyna, 2002). Although the picture is now more nuanced, it remains clear that *within* therapy changes can have important implications for *post* therapy outcome (Lorenzo-Luaces et al., 2015; Safran et al., 2011). As such, post therapy outcome alone is insufficient to fully understand changes that occur from psychological therapy. Furthermore, post-therapy outcomes alone tend to inflate the reported effectiveness of therapy, because patients who have poorer outcomes are more likely to drop out (Clark et al., 2009). One means of evaluating changes within psychological therapy involves the growing ROM method (Wampold, 2015).

The umbrella term, ROM, is used to describe methods in which an outcome assessment is given, typically at every session of psychological therapy, and resultant scores give therapists algorithm-based predictions of treatment outcome (Lambert, 2010). This aims to both alert therapists to patients who are at risk of having a poor outcome and promote patient-therapist collaboration to enhance effectiveness. With appropriate training and supervision, psychological therapists can tailor therapy to the outcome trajectory of individual patients. As such, ROM becomes more than an

assessment method and can be seen as an intervention in its own right. Current evidence from RCTs and subsequent meta-analyses suggest that ROM can have several beneficial effects for both treatment efficacy and efficiency (Delgadillo et al., 2018; Lambert et al., 2018; Shimokawa et al., 2010). The two best-evidenced ROM tools use general, overall assessments of mental health and wellbeing: The Outcome Questionnaire 45 items (OQ-45; Lambert, 2012) and the ORS (S. D. Miller et al., 2005). Both systems aim to give an overview outcome assessment which can be used across conditions, identifying the “mental health vital signs” (Lambert et al., 2018, p. 520).

Although the previous chapter has shown the way in which patient activation can influence long-term outcome through early session interactions, it is unclear whether similar patterns would be found from more immediate ROM outcomes. Concordance between proximal and distal outcome associations would strengthen the evidence of potential causality. It would also be a safeguard against chance associations, which could occur in early, exploratory, developmental studies such as those reported in these chapters.

The current study evaluates whether CICS rated patient activation at the first and second session predicts change in session-by-session rated ROM outcome over the course of psychological therapy. The same data set is used as in the previous study, so that a clearer relationship between CICS scores and outcome can either be established or rejected. Although this study remains exploratory, particular attention is paid to the replication of findings in the previous chapter:

- Is increased engagement in conversations about positive changes noticed at the second session predictive of greater session-by-session improvements through psychological therapy?
- Is increased time spent describing problems at session 2 predictive of reduced outcome-improvement over the course of psychological therapy?
- Is increased facilitative and general discussion (“other” interactions) predictive of increased session-by-session outcome improvement?
- Do the CICS’ outcome-predictive capabilities remain for session-by-session outcome when controlling for therapeutic alliance? This would indicate that CICS scores measure a construct which has significant elements that do not overlap with therapeutic alliance.

7.2.1 Study 1 method

7.2.1.1 Source data and design

The same dataset, methods and design were used as in the previous chapter except that session-by-session, time series ROM data were used for outcome predictions. The details are summarised here but described in more detail in chapter 6. Again an exploratory, observational design was applied to participant data from an RCT of remotely delivered CBT versus usual care for people with health anxiety who were using urgent care at a high rate (Morriss et al., 2019). Participants randomised to treatment (n = 79) were offered six to 12 sessions of CBT delivered remotely via video conferencing software or the telephone. Observer-rated data was used as in the previous study: The CICS was used to rate therapist-patient interactions in the first and second CBT session, based on the level of patient activation present. Observational

assessment of the therapeutic alliance (WAI-O; Horvath & Greenberg, 1989) was conducted for the same sessions, so analysis could control for the overlapping predictive value of the WAI-O and CICS. Audio or video recordings of the first and second CBT sessions were transcribed verbatim following participants' written informed consent.

Different from the previous study, ROM data (as opposed to independent outcome data) were then used to assess whether CICS ratings predicted ROM scores over time (up to 15 session-by-session ratings), controlling for baseline outcome score and WAI-O. Therapy sessions in the trials were generally delivered on a weekly basis with later sessions or booster sessions offered multiple weeks apart in line with patient preference. As patients attended eight sessions of CBT on average, many were likely to have completed therapy by the first follow-up assessment reported in the previous study. Therefore, ROM data gives an important insight into changes occurring prior to follow-up and over the course of therapy.

7.2.1.2 Inclusion and exclusion criteria

The same inclusion criteria was used as in the previous study, with one additional feature: participants were excluded if they had refused the use of ROM, but were not required to have completed any follow-up outcome assessments, as these were not necessary in the current study. Therefore, 49 of the 79 possible participants (62%) were included, having: (1) received CBT as part of the source RCT; (2) attended two or more CBT sessions (providing sufficient data for the analysis); (3) consented to treatment sessions being recorded and extracts anonymously reported; and (4) completed at least two ROM assessments. All those meeting the criteria were included.

7.2.1.3 Intervention

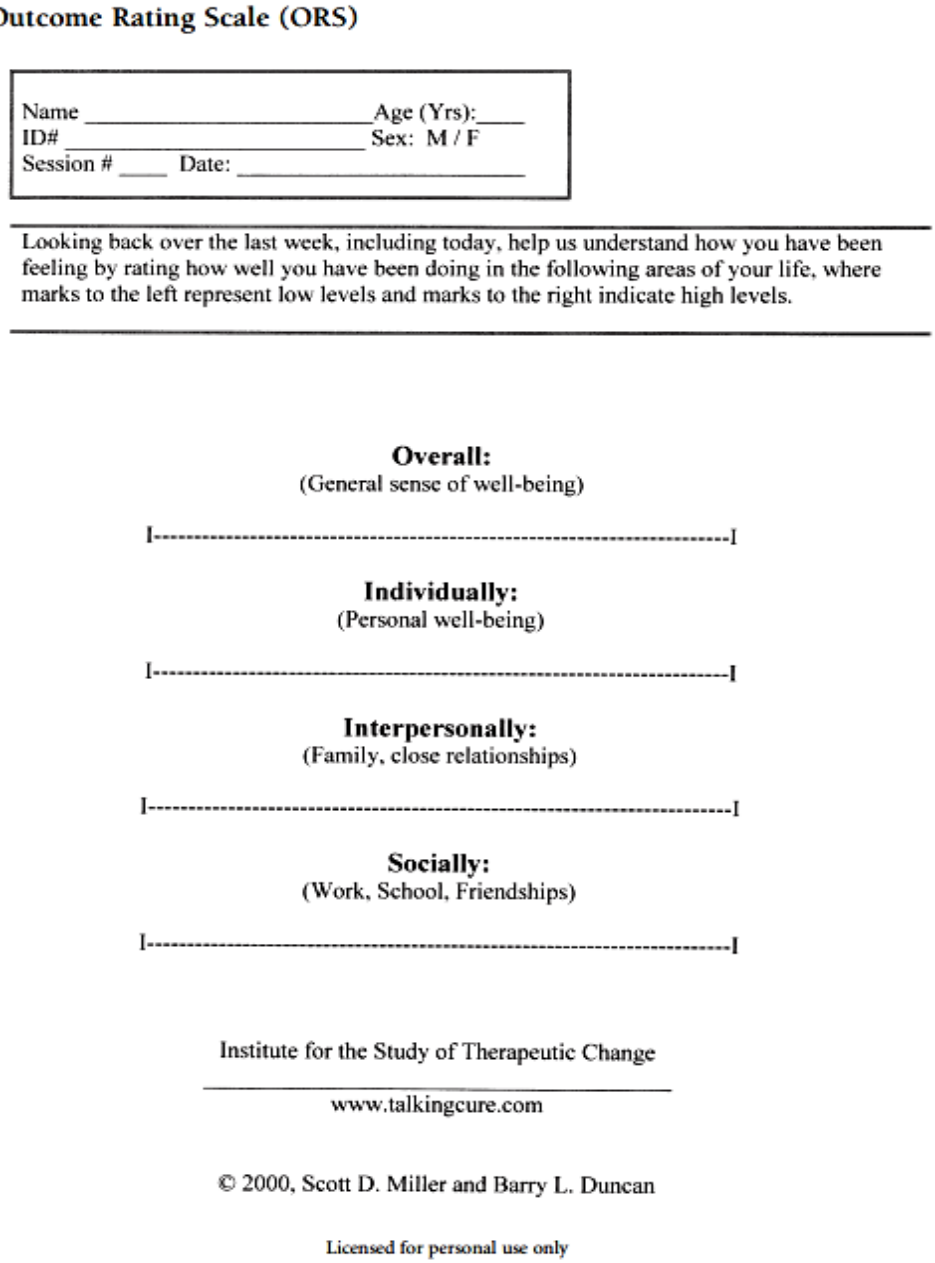
A team of four CBT therapists remotely delivered CBT adapted to health anxiety using an established treatment protocol (Morriss et al., 2019).

7.2.1.4 Routine outcome measure

The ORS was completed at each session of CBT and is the outcome measure in this study. The ORS is the central assessment for the Partners for Change Outcome Measurement System (PCOMS; S. D. Miller et al., 2005). It is one of the most widely used ROM systems, with approximately 45,000 therapists registered across 20 countries (Østergård, Randa, & Hougaard, 2018). It is designed for use in all healthcare but has predominantly been evidenced in psychological therapy populations (Lambert et al., 2018).

The ORS uses four self-reported, brief visual analogue scales, which aim to assess the central targets of psychological treatment: individual wellbeing, change in relationships, social functioning, and overall wellbeing (see Figure 6.5; S. D. Miller et al., 2003). In normal use of the ORS, each scale is rated 0 to 10 and the scores are totalled out of 40. Higher scores indicate improved wellbeing and scores below the clinical cut-off of 25 fall within the established clinical range (Prescott, Maeschalck, & Miller, 2017). The ORS has shown excellent internal consistency (α = .93); adequate test-retest reliability (*r*s = .49 - .66) and satisfactory convergent validity with the OQ-45 (*r*s = .56 - .69).

Figure 6.5. The print version of the ORS



Outcome Rating Scale (ORS)

Name _____ Age (Yrs):____
ID# _____ Sex: M / F
Session # ____ Date: _____

Looking back over the last week, including today, help us understand how you have been feeling by rating how well you have been doing in the following areas of your life, where marks to the left represent low levels and marks to the right indicate high levels.

Overall:
(General sense of well-being)

I----------I

Individually:
(Personal well-being)

I----------I

Interpersonally:
(Family, close relationships)

I----------I

Socially:
(Work, School, Friendships)

I----------I

Institute for the Study of Therapeutic Change

www.talkingcure.com

© 2000, Scott D. Miller and Barry L. Duncan

Licensed for personal use only

7.2.1.5 Observational process measures

The CICS categorises and rates patient activation for each in-session turn of speech. The WAI-O (Horvath & Greenberg, 1989) is the most commonly used measure of therapeutic alliance and an established benchmark for reliability and predictive validity**.**

7.2.1.6 Method of analysis

As patient characteristics and ratings are almost identical to the previous chapter, the reader is referred to these data (p. 191). Therefore, comparative characteristics and IRR are not reported in the results. Session 1 and 2 CICS and WAI-O ratings were associated with residualised post-treatment ORS scores using Spearman's rho.

Given that included participants attended a different number of sessions (range: 2-15), MLM was the most appropriate means of including all participants, accounting for missing data and controlling for within case variance. A statistically significant 44% of ORS covariance was explained by nesting of ORS scores within reporting participants. Therefore, a two level structure was explored in an unconditional model: session-by-session ORS score nested within patients. Following conventional model building guidelines (Raudenbush & Bryk, 2002), variables were introduced one-by-one using -2 log likelihood to assess goodness of fit. Fully adjusted models included random intercepts and slopes with diagonal covariance matrices. The following predictors were included: *assessment time point (adjusting for ordering of assessments); first session ORS score (adjusting for baseline severity); session 1 WAI-O (controlling for the overlap with therapeutic alliance); session 1 CICS theme (the target predictor variable).* The outcome was ORS score over time. Similar models were repeated for each CICS theme and with session 2 WAI-O and CICS theme scores. The six steps taken to develop these models are illustrated in Table 7.1, showing the development of a model to assess the predictive validity of first session interactions where patients

expressed (or were encouraged to express) an evaluation of themselves or therapy (EST).

A pseudo-R^2 measure was used to estimate the effect size of change in between-patient variance explained by CICS themes identified as significant predictors (Raudenbush & Bryk, 2002; Singer & Willett, 2003). Using the model developed in Table 7.1 as an illustration, the pseudo-R^2 was calculated as the percentage of: $((\sigma_5^2-\sigma_6^2)/ \sigma_1^2)$, where σ_1^2 is the unconditional model intercept variance (step1), σ_5^2 is intercept variance in the fully adjusted model *without* the target predictor (step 5) and σ_6^2 is the fully adjusted model's intercept variance (step 6). This calculation shows the percentage reduction in unexplained intercept variance (variance between participants) when CICS themes are added to fully adjusted models ("full" below) (Hox et al., 2017). Town, Abbass, Stride, and Bernier (2017) used this method to demonstrate the percentage by which "unexplained variation in change over time was reduced" (p. 20) during psychological therapy for depression. The same method was then replicated to explain variance in the rate of change between patients (slope variance):

$$Pseudo\ R^2 = \frac{(variance\ full\ without\ target)-(variance\ full\ with\ target)}{variance\ unconditional}$$

Both study 1 and 2 obtained ethical approval from the National Research Ethics Service, London-Riverside Committee (reference 14/LO/1102). Analysis was conducted using SPSS 24.

Table 7.1. Steps in the development of the MLM model assessing the predictive ability of session 1 self-evaluative interactions as a predictor of session-by-session ORS-rated wellbeing.

Step	Model	Deviance (-2LL)	Change in deviance, new dfs	Was parameter added statistically significant?	Residual variance	Intercept variance	Slope variance
1	Unconditional	3304	-	-	40.81	32.33	-
2	Fixed effects and random intercept for time	3202	102, 1	yes	31.92	38.16	-
3	Random slope for time	3177	25, 1	yes	27.55	39.88	0.27
4	Control variable 1: Fixed effect of baseline wellbeing (ORS)	3146	31, 1	yes	27.24	17.45	0.32
5	Control variable 2: Fixed effect of session 1 therapeutic alliance (WAI-O)	2705	441, 1	yes	27.59	19.75	0.27
6	Main predictor of interest: Fixed effect of session 1 evaluations of self or therapy (EST, CICS)	2698	7, 1	yes	27.66	15.63	0.25

Note. Abbreviations: -2LL: -2 log likelihood; dfs: degrees of freedom; ORS: outcome rating scale; WAI-O: working alliance inventory – observer rated.
[a]Statistical significance was assessed using chi-squared with the change in -2LL and the change in dfs.

7.2.2 Results

No CICS themes nor the WAI-O correlated with residualised post-treatment ORS scores from session 1 ratings (all r_Ss <.20). Only structuring (STF) and facilitative/general interactions (OTH) showed significant outcome correlations from session 2 ratings (STF: r_S = .33; OTH: r_S = .35).

Accounting for missing data, 40 patients were included in models using first session ratings and 48 patients were included in models evaluating second session ratings. All models reported met the previously described assumptions for MLM analyses.

Two of the three CICS themes that are the focus of confirmatory analysis showed replicated results in this study. Noticing change interactions at the second sessions were a significant predict of session-by-session wellbeing, replicating findings in the previous study (Table 7.2). The proportion of interactions in both the first and second session spent on problem and context description were again negatively correlated with ORS-rated wellbeing across time (Table 7.3). This replicates and extends previous findings – session 1 problem description interactions were not predictive of long-term outcomes in the earlier study, but are for session-by-session change in wellbeing. Facilitative interactions and general conversation were not predictive of session-by-session ORS outcome at either first or second session (Table 7.4).

Table 7.2. Multilevel model where session 2 interactions about noticed change (or the absence of change) predict session-by-session wellbeing (ORS).

Fixed effects					
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	6.93	5.89	.245	-4.91	18.78
Assessment time point	0.81	0.11	<.001	0.59	1.04
Baseline wellbeing (ORS)	0.75	0.11	<.001	0.53	0.98
Therapeutic alliance at session 2 (WAI-O)	-0.17	1.01	.869	-2.20	1.87
Interactions about changes noticed at session 2 (CICS)	*0.41*	*0.16*	*.015*	*0.08*	*0.74*

Random effects				
Parameter	Variance	SE	Z	*p*
Residual variance	27.63	1.96	14.07	<.001
Intercept variance	14.73	4.59	3.21	.001
Slope variance	0.28	0.10	2.85	<.001

Note. Abbreviations: ORS: outcome rating scale; B: unstandardized β coefficient; SE: standard error; p: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme; Z: Walt's z test for significance of variance.

Table 7.3. Multilevel model where the proportion of session 1 interactions taken up with problem description predicts session-by-session wellbeing (ORS)

Fixed effects					
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	23.83	9.37	.015	4.89	42.77
Assessment time point	0.89	0.12	<.001	0.66	1.13
Baseline wellbeing (ORS)	0.75	0.12	<.001	0.51	0.99
Therapeutic alliance at session 1 (WAI-O)	-1.22	1.27	.341	-3.79	1.34
Problem or context description at session 1 (CICS)	*-0.16*	*0.05*	*.006*	*-0.27*	*-0.05*

Random effects				
Parameter	Variance	SE	Z	*p*
Residual variance	27.59	2.10	13.16	<.001
Intercept variance	15.01	5.01	2.99	0.003
Slope variance	0.26	0.10	2.59	0.010

Note. Abbreviations as above.

Table 7.4. Multilevel model where the proportion of session 2 interactions taken up with problem description predicts session-by-session wellbeing (ORS)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	13.92	6.52	0.038	0.80	27.04
Assessment time point	0.81	0.11	<.001	0.58	1.03
Baseline wellbeing (ORS)	0.73	0.11	<.001	0.51	0.95
Therapeutic alliance at session 2 (WAI-O)	0.02	0.99	0.984	-1.97	2.01
Problem or context description at session 2 (CICS)	*-0.13*	*0.05*	*0.008*	*-0.23*	*-0.04*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	27.64	1.96	14.07	<.001
Intercept variance	14.22	4.48	3.18	0.001
Slope variance	0.28	0.10	2.86	<.001

Note. Abbreviations as above.

Two other themes showed predictive validity to proximal, routinely assessed outcome. Firstly, interactions in which evaluations of therapy, the therapist or the patient were made (EST). These interactions predicted session-by-session outcome from both first and second session ratings (Tables 7.5 and 7.6). This is noteworthy, as positive evaluation of therapy or the parties within therapy took up only 0.5% of all spoken words. Taken together, this means that even a small increase in interactions where therapy was evaluated positively would contribute to improved wellbeing over the course of therapy:

> **Therapist:** Is there anything you'd want us to do differently next time we meet?

> **Patient:** No because I really am enjoying what you're doing and it is sort of, I like to see the visual stuff because that helps, so yeah it's really good. Really good, yeah.
>
> [P 01051]

The same would be true of increases in self-supporting statements made by patients at session 1 or 2:

> **Patient:** Well I nearly walked out [ ] I thought that would be something the old me would've done so I wanted to stick with it and not give up and just take what I could out of the class [ ] so I was really pleased with myself that I stayed and learnt about [yoga] (laughs).
>
> **Therapist:** Yes.
>
> **Patient:** about being enough and about never saying I can't, and always trying.
>
> [P 03002]

On the other hand, the absence of positive evaluations or actively negative evaluations of the self or therapy in early sessions (taking up only 0.1% of spoken words) contributed to poorer outcome over sessions:

> **Patient:** Yes that's what I do, when I go on a plane [ ] I literally virtually have a breakdown [ ] you know which is stupid I just I hate myself for being weak and you know I need to be strong about these things and it's just it's just silly I know it's silly [ ].
>
> **Therapist:** Yeah, yeah, can I just ask then why do you that?
>
> [P 01071]

Structuring and task-focused interactions at the second session predicted ORS-rated wellbeing over time, but not from first session ratings (Table 7.7).

Table 7.5. Multilevel model where evaluations of self or therapy at session 1 predict session-by-session wellbeing (ORS)

Fixed effects

Parameter	B	SE	*p*	95% CI	
				Lower Bound	Upper Bound
Intercept	12.00	7.19	.103	-2.54	26.55
Assessment time point	0.90	0.12	<.001	0.66	1.13
Baseline wellbeing (ORS)	0.71	0.12	<.001	0.47	0.95
Therapeutic alliance at session 1 (WAI-O)	-0.81	1.24	.518	-3.31	1.69
Evaluations of self or therapy at session 1 (CICS)	*2.04*	*0.76*	*.011*	*0.50*	*3.58*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	27.66	2.10	13.14	<.001
Intercept variance	15.63	5.15	3.03	.002
Slope variance	0.25	0.10	2.57	.010

Note. Abbreviations: ORS: outcome rating scale; B: unstandardized β coefficient; SE: standard error; p: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme; Z: Walt's z test for significance of variance.

Table 7.6. Multilevel model where evaluations of self or therapy at session 2 predict session-by-session wellbeing (ORS)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	9.48	6.23	0.135	-3.05	22.00
Assessment time point	0.82	0.11	<.001	0.59	1.05
Baseline wellbeing (ORS)	0.80	0.12	<.001	0.57	1.03
Therapeutic alliance at session 2 (WAI-O)	-0.71	1.11	0.527	-2.95	1.53
Evaluations of self or therapy at session 2 (CICS)	*0.68*	*0.32*	*0.039*	*0.04*	*1.32*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	27.50	1.95	14.12	<.001
Intercept variance	15.39	4.77	3.23	0.001
Slope variance	0.31	0.11	2.94	0.003

Note. Abbreviations as above.

Table 7.7. Multilevel model where structuring and task focus interactions at session 2 predict session-by-session wellbeing (ORS)

Fixed effects

Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.68	5.90	.340	-6.19	17.56
Assessment time point	0.81	0.11	<.001	0.59	1.03
Baseline wellbeing (ORS)	0.74	0.11	<.001	0.52	0.97
Therapeutic alliance at session 2 (WAI-O)	-0.26	1.03	.805	-2.33	1.82
Structuring interactions at session 2 (CICS)	*0.20*	*0.09*	*.024*	*0.03*	*0.38*

Random effects

Parameter	Variance	SE	Z	*p*
Residual variance	27.78	1.99	13.97	<.001
Intercept variance	15.09	4.66	3.24	.001
Slope variance	0.26	0.10	2.71	.010

Note. Abbreviations as above.

No other CICS themes showed predictive validity from associations with session-by-session ORS scores. Neither session 1 nor session 2 WAI-O ratings showed predictive validity with the ORS over the course of therapy, irrespective of whether CICS themes were included as covariates or not.

Attempts to calculate effect sizes from MLM analyses are done with caution, as estimates may not be stable or accurate; in particular negative pseudo-R^2 values are not interpretable (Hox et al., 2017). Nonetheless, pseudo-R^2 values are presented in Table 7.8 and 7.9 to give a quantifiable estimate of effect size. This represents the percentage change in between-patient outcome variance (intercept variance) unique to significant CICS themes when controlling for baseline severity, therapeutic alliance and change over time (Table 7.8). Change in slope variance is also presented (Table 7.9). This quantifies how variation in the rate of change is uniquely explained by CICS themes, by estimating the reduction in slope variance when CICS themes are added to fully-adjusted model. Unexplained variation in outcome over time was reduced by 13 and 15% from adding session 1 evaluation of self/therapy and problem description interactions respectively. Unexplained variation in the rate of change was reduced by 1 and 4% respectively for session 1 ratings of the same themes. Significant CICS outcome predictors at session 2 reduced unexplained variation in outcome over time by 7-10%. Session 2 predictors also reduced unexplained variation in the rate of change by 6-28%. Evaluations of self or therapy at the second session showed a negative value for change in slope variation, which cannot be interpreted (Raudenbush & Bryk, 2002). All reductions in unexplained variance controlled

for any change in variance associated with baseline severity, therapeutic alliance and change over time.

Table 7.8. Reduction in unexplained outcome variance between patients over time by CICS theme

CICS theme	% Reduction in unexplained intercept variance	Fully adjusted model variance with CICS theme	Fully adjusted model variance without CICS theme	Variance in unconditional model[a]
Session 1 Problem or context description	15%	15.01	19.75	32.33
Session 1 Evaluations of self or therapy	13%	15.63	19.75	32.33
Session 2 Problem or context description	10%	14.22	17.57	32.33
Session 2 Noticing change or otherwise	9%	14.73	17.57	32.33
Session 2 Structuring and task focus	8%	15.09	17.57	32.33
Session 2 Evaluations of self or therapy	7%	15.39	17.57	32.33

Note. Abbreviation: CICS: Consultation interaction coding scheme.
[a]Unconditional models contained random intercepts only with no covariates.

Table 7.9. Reduction in unexplained rate of change variance by CICS theme

CICS theme	% Reduction in unexplained slope variance	Variance in fully adjusted model with CICS theme	Variance in fully adjusted model with CICS theme	Variance in unconditional model[a]
Session 2 Noticing change or otherwise	28%	0.22	0.30	0.30
Session 2 Structuring and task focus	13%	0.26	0.30	0.30
Session 2 Problem or context description	6%	0.28	0.30	0.30
Session 1 Evaluations of self or therapy	4%	0.26	0.27	0.30
Session 1 Problem or context description	1%	0.26	0.27	0.30
Session 2 Evaluations of self or therapy	-3%	0.31	0.30	0.30

Note. Abbreviation: CICS: Consultation interaction coding scheme.
[a]Unconditional models contained random intercepts and random slopes for time with no covariates.

7.2.3 Discussion

This study partly replicates the findings from the previous chapter, showing a predictive relationship between early-rated CICS scores and long-term outcomes. In this study similar relationships were found with outcomes in closer proximity to the predictive events. Session 2 interactions where changes were noticed again predicted outcome, but this time wellbeing ratings over the course of therapy. Problem description interactions were negatively associated with session-by-session wellbeing. This replicated and extended existing findings: Although there were trends for problem description to be negatively associated with all distal outcomes at session 1 in the previous study, no associations were significant. This study suggests that there may be an important relationship at session 2 *and 1* worth further investigation.

Facilitative and general interactions, which were surprisingly associated with outcome in the previous study were not predictive of proximal outcome in this study. This may indicate that previous findings were chance associations or there may indeed be a relationship to distal anxiety and depression outcomes that are not demonstrated in this study of proximal wellbeing. The generic model of psychotherapy suggests that session outcomes should related to post-treatment outcomes (Orlinsky & Howard, 1987). However, this does not hold in the case of facilitative interactions during CBT for health anxiety.

The association found between structuring interactions and outcome suggests that further investigation of this type of interaction is warranted. This finding is supported by results from the developmental study and the trends for

association with all outcomes in the previous study, despite none reaching significance.

Given the exceptionally small proportion of sessions rated as evaluations of self or therapy, it seems important that ratings from both the first and second session were associated with outcome. In the previous study there were consistent trends for outcome association with all long-term measures from both first and second session ratings, but none were significant. Evaluations of self or therapy were the least coded interaction theme and take a very small proportion of all interactions. However, this type of interaction was a significant predictor of session-level therapeutic outcome from first and second session ratings. This finding suggests that even brief comments about the value of therapy can be a significant process marker in initial therapy interactions. Furthermore, it suggests therapists should attend to patients' evaluations of themselves made early on, even brief throw away comments given the infrequency of such occurrences. Small amounts of this type of interaction can influence therapy's trajectory in either direction. As such, interactions where therapy or the parties within therapy are evaluated could be an important and potent process marker.

7.2.3.1 Strengths and limitations

When the results described are combined with those reported in the previous chapter a more comprehensive view of interaction-level patient activation can be taken. This study shows how outcomes much nearer to the evaluated events (sessions 1 and 2) are also predicted by specified changes in interaction-style and type. As the predictors and outcomes are closer in time and there are many more outcome ratings (up to 15), a dependable

assessment of outcome trajectory can be estimated across treatment and a more likely mediating effect can be inferred.

However, as with the previous chapter, this study can be seen as being on the borderline of an appropriate sample size for MLM analyses (Maas & Hox, 2005; Raudenbush & Bryk, 2002). Therefore, the findings should be treated with some degree of caution. In particular, the method used to give an estimate of the percentage reduction in unexplained variance may not be dependable. This study also used multiple testing, which increases the likelihood of type 1 errors. Yet, this study was to some degree confirmatory, using findings from the previous chapter as a guide. Therefore, the differential focus on proximal and distal outcomes can give an indication of how dependable the findings are. In this case, the replication of results for problem description interactions and noticing change interactions increases the reliability of these findings. However, the failure to replicate findings for facilitative/general interactions suggests that the previous results should be treated with greater caution.

Overall, a consolidation and confirmation of previous findings is given through this study's focus on proximal outcomes. This study confirms problem description and noticing change interactions as important outcome predictors. By contrast, this study also indicates that facilitative and general interactions should be treated more tentatively as outcome predictors. Furthermore, current results suggest that structuring and task-focus interactions should not be rejected as potentially important to outcome prediction. Lastly, this chapter introduces evaluations of self or therapy as a potentially important therapeutic outcome predictor worthy of ongoing investigation.

7.3 Study 2: Changes in patient activation prior to sudden gains

It is unclear from current research whether motivational factors such as patient activation play a significant role in sudden gains: clinically important, rapid, sustained improvements. If this were the case, there are several advantages to being able to assess patient activation from observable, in-session processes rather than self-report questionnaires as is usual practice. This study assesses whether increases in interaction-rated patient activation precede sudden gains, to help understand the types of interactions that may be particularly important in promoting sudden gains.

The evidence presented in previous chapters suggests that patient activation can be reliably rated at an interaction-level within psychological therapy and patient activation in interactions is associated with clinical outcomes. The potential value of predicting outcomes on the basis of patient activation in interactions is limited by poor understanding of whether changes in patient activation predict clinically important session-by-session changes in psychological therapy. Sudden gains are an important type of clinical change occurring between two sessions.

Sudden gains are defined as clinical improvements which are rapid (occurring between two adjacent psychological therapy sessions); large in absolute and relative terms; sustained in a way that would be unlikely to occur by chance fluctuations, and typically occur early in the treatment process (Tang & DeRubeis, 1999b). Sudden gains are seen as clinically important phenomena, because those experiencing sudden gains report greater clinical improvements than those who do not and their improvements are more likely to be sustained long-term (Aderka, Nickerson, Bøe, & Hofmann, 2012; Tang,

DeRubeis, Hollon, Amsterdam, & Shelton, 2007). Given that sudden gains usually occur prior to the delivery of most therapeutic content (i.e. before the completion of activities deemed to facilitate clinical improvement) other factors have been explored to help explain why such dramatic improvements occur before core treatment elements have been delivered.

Cognitive changes are identified as a possible predictor of sudden gains during CBT for depression, given that more cognitive changes occur directly prior to a sudden gain than at control sessions (Tang et al., 2005). Cognitive changes were defined as a patient's explicit identification of a belief, errors in beliefs or arriving at new beliefs. Many view cognitive change as the mediating process through which CBT operates (Lorenzo-Luaces et al., 2015). Therefore, this finding has been taken as evidence supporting cognitive mediation (Tang et al., 2005). As therapeutic alliance did not improve until after the sudden gains in the same study, it was suggested that cognitive changes may be of greater causal importance than therapeutic alliance (Tang & DeRubeis, 1999b). This type of research has emphasised the idea of "critical sessions": The long-standing proposal that some psychological therapy sessions have a significantly greater beneficial effect than others (Elliott, 1989). The session directly prior to a sudden gain has been called the critical session, because of the important changes occurring at this session that do not occur in preceding sessions (Tang et al., 2005).

However, with the exception of cognitive changes, few other process factors have been investigated in a similar fashion for sudden gains. This is puzzling when a number of mechanisms have been identified as predictors of early symptom change in related research on therapeutic improvements early

in psychological therapy. For example, Poulin et al. (2018) identified that interaction-level motivation within early consultations predicted outcome in combined MI and CBT for anxiety. In agreement with the previous chapters, this suggests that outcomes can be predicted from early, in-session motivational factors, but it cannot say whether important session-level changes are predicted. The same is true of interaction-level patient activation and sudden gains: We can infer from the evidence presented so far that early patient activation predicts outcome, but we cannot say the specific effects of patient activation on clinically important session-level changes, such as sudden gains. This study aims to address this research gap.

As sudden gains are beneficial to treatment outcome, identifying the types of interaction that may lead to such improvements might help therapists increase their likelihood and improve understanding of sudden gains as a phenomenon. This study aimed to assess:

(1) Whether patient activation is greater at sessions directly prior to sudden gains (critical sessions) as compared to control sessions from the same patient.
(2) Whether patient activation is greater at sessions directly prior to sudden gains as compared to control sessions from matched patients who did not experience sudden gains.
(3) Which specific interaction types are most important prior to sudden gains.

7.3.1 Method

7.3.1.1 Design

This study adhered to the methods used to assess whether cognitive changes increased prior to sudden gains, but applied adapted criteria (Tang & DeRubeis, 1999b; Tang et al., 2005). In this case patient activation was assessed using the same methodology: comparing the patient activation levels in sessions directly before a sudden gain (session *N,* also known as the critical session) with the closest control session occurring before the gain (session *N* – 1). An additional exploratory assessment compared the critical session from the sudden gainers with sessions from non-sudden gainers at a similar stage in therapy. Non-sudden-gainers were matched on the basis of age and gender. This paired comparison aimed to assess whether there were meaningful *between-case* differences at similar time points in therapy on the basis of patient activation expressed in session interactions. In both assessments it was hypothesised that there would be greater patient activation at critical sessions than either within or between case comparator sessions.

An observational design was applied to a secondary analysis of data from the same CBT for health anxiety RCT used in the previous chapter (Morriss et al., 2019). Study participants were drawn from those who received CBT through the trial (n = 22). This study used ROM data collected by therapists at each session using the ORS and CBT session recordings and transcripts to complete the processes described.

7.3.1.2 Participants

Eleven of those identified as experiencing a sudden gain during CBT were included in the current study having met the following inclusion criteria:

5. Patients consented to their sessions being recorded and extracts anonymously reported;
6. There were available ROM measurements for each session using the ORS;
7. Session recordings were available for both sessions *N* and *N* – 1.

7.3.1.3 Identifying sudden gainers

Participants experiencing sudden gains during CBT in the trial were identified from ORS data using an adaptation of Tang and DeRubeis' (1999b) criteria. A sudden gain was deemed to occur if there was a large increase in ORS score between consecutive sessions (between session N directly prior to the gain and session N +1) if the following criteria were met:

- The gain was large in absolute terms – there was an increase in score of at least 5 points, the established reliable change index for the ORS (Prescott et al., 2017).
- The gain was large in relative terms – the N + 1 gain was at least a 25% improvement on session N.
- The gain was sustained beyond random fluctuations – the mean ORS score of the three sessions prior to the gain were at least a reliable change (5 points) lower than the mean ORS score for the three sessions following the sudden gain. This criterion was adapted to include sudden gains at the very beginning or end of treatment given that treatment in the current study was much briefer than in the study from which the criteria originate (16-20 versus 6-12 sessions). Indeed, only a sudden gain at session 3 could be identified if the minimum number of sessions was attended. This is particularly problematic when sudden gains identified at the first session have been found to positively

influence outcome in their own right (Busch, Kanter, Landes, & Kohlenberg, 2006). Therefore, fewer than three sessions prior to the gain were allowed, as long as the gain was sustained for the three ensuing sessions. Conversely, fewer than three sessions were allowed after a gain assuming that the pre-gain score was stable for the previous three sessions. This aimed to include very early or late gains which would otherwise be missed.

Once participants experiencing sudden gains were identified, audio/video recordings and transcripts of the critical session (*N*) and the session directly prior to that (*N* – 1) were used to rate the therapeutic alliance and patient activation at each session. This was carried out using the WAI-O (Horvath & Greenberg, 1989) and the CICS respectively.

7.3.1.4 Identifying matched non-sudden gainers

Each of the eleven sudden gainers included were matched to the non-sudden gainer in the remaining treatment sample who was the nearest in age of same gender. The critical sessions from the sudden gainers were compared with session 2 from the non-sudden gainers (the median for critical sessions amongst sudden gainers) on the basis of patient activation and therapeutic alliance using session recordings and transcripts.

7.3.1.5 Assessments

As already described, the ORS is a ROM instrument assessing patient wellbeing using four visual analogue scales. The ORS was completed at each session to monitor treatment progress and scores were used to identify sudden gains.

Therapeutic alliance was assessed for each included session using the WAI-O (Horvath & Greenberg, 1989). The CICS was used to rate patient activation at the interaction level within different categories of interaction (Table 7.10 gives examples of level descriptors for information discussion interactions).

The WAI-O and the CICS were both completed by two independent observers, having reviewed session transcripts with recordings for each included session. Inter-rater reliability was calculated and average scores for the two assessors were used in the final analysis.

Table 7.10. Patient activation level descriptors and exemplars for the Information Discussion CICS theme.

Level	Level Descriptor[a]	Exemplar Quote
+2	*Patient initiated request or cue* for specific information, relevant to current therapeutic endeavour. Encouraged, endorsed or developed by the therapist.	*P:* How do you find out which habits are going under the radar, how do I honestly identify them? [ ] *T:* We identify what these things are, what the kind of habits of responding to worry are and we gradually reduce them [P 03002]
+1	Therapist initiated information discussion which the patient acknowledges and *develops further, summarises or seeks clarification.*	*T:* in terms of sleep, if you aren't getting seven hours, then that's probably, you could definitely do with more but there's variability [ ] how is that kind of landing with [ ]? *P:* Yeah, I mean quite, sleep, yeah, like unless I have commitments I probably, I do like to sleep as much as I can. The exercise one, I can get halfway there probably. [01071]
0	Therapist initiated information-giving - *no development* OR *development suggests unresolved misunderstanding* OR therapist acknowledges patient request for information, but does not develop.	*T:* And when we think that we oh I've got a gut feeling about they're very, very often completely wrong they're just because we just feel a certain way. *P:* Yeah. [06001]
-1	Therapist gives information – *patient displays ambivalence* in response OR gentle dismissal OR Information given does not seem relevant or applicable and this is clear from their responses.	*T:* Those things that are going on in your mind, they aren't facts, [ ] they aren't necessarily true, they're just thoughts, [ ] *P:* Yeah, I know what you mean but, like, sometimes if I am out, sometimes they start happening while I'm out. [01017]
-2	Information given that *elicits a negative response*, such as active rejection or mocking.	*T:* So if you stop checking your body you are likely to feel less anxious about your health. *P:* I really think you are completely wrong [hypothesised]

Note. [a]Italics add emphasis to the key component of the level descriptor; Abbreviations: *P*: Patient; *T*: Therapist.

7.3.1.6 Statistical analysis

Inter-rater reliability of WAI-O and CICS raters was assessed using two-way, random effects, average-measures, consistency ICCs (Shrout & Fleiss, 1979). The comparison of critical sessions (N) with control sessions (N – 1) was carried out using Wilcoxon Signed-Ranks test for these matched pairs of sessions. The same was used for the matched, between-case comparison of sudden gainers' critical sessions and non-sudden gainers control sessions. The clinical characteristics of the sudden gainers and their non-sudden gainer pairs were also compared using Wilcoxon Signed-Ranks test for continuous variables. As participants were matched by gender, chi-squared tests were not needed to compare any other categorical variables. Analysis was carried out using SPSS 24.

7.3.2 Results

7.3.2.1 Sample characteristics

Sudden gainers were majority female with ages ranging from 20 to 53 years old. There were no statistically significant differences in any demographic or clinical characteristics (Table 7.11). However, the percentage improvement in ORS score was close to a statistically significant difference in favour of sudden gainers ($Z = -1.955$, $p = .051$). The mean and median difference in age between sudden gainers and their matched pair was four years, with a range of 0 to 10 years difference.

Table 7.11. Sudden and non-sudden gainer sample characteristics.

	Sudden gainers n = 11	**Non-sudden gainers n = 11**
Mean age (min, max)	33 (20, 53)	36 (21, 51)
Female (%)	10 (91%)	10 (91%)
No sessions (IQR)	12 (11,15)	11 (6,12)
First session ORS (IQR)	20.3 (13.3, 24.5)	21.5 (20.4, 24.4)
Last session ORS (IQR)	33.6 (25.9, 36.9)	29 (22.9, 31.7)
% improvement in ORS (IQR)	62 (34, 91)	25 (0, 38)

Note: IQR, inter-quartile range; ORS, outcome rating scale.

7.3.2.2 Inter-rater reliability

All CICS themes fell within the good to excellent range of IRR (ICCs: .61 to .88; Cicchetti, 1994). Only Structuring and task focusing (STF) interactions achieved ICCs below .70 (.61). Excellent IRR was also achieved for WAI-O ratings (ICC: .81; Cicchetti, 1994). Overall, this indicated a satisfactory level of IRR for dependable results, although results for the STF theme were treated with caution.

7.3.2.3 Critical sessions versus control sessions

There were significantly higher scores for Information Discussion (IDI) interactions during critical sessions as compared to control sessions ($Z = -2.03$, $p = .043$). This meant that there was greater patient activation during interactions information-based discussions directly prior to a sudden gain, as compared to similar control sessions.

Therefore patients were more actively receptive or engaged with information giving interactions:

> **Therapist:** the best way to think about sleep is like [washing the brain]. Because if you [ ] don't wash thoroughly, over several days, the grime's going to start to build up. [ ] I find the washing idea really useful to…
>
> **Patient:** I'm noting that down, I like this.
>
> [P 01015, IDI +1]

By contrast there were fewer occasions of either minimal responses where it is unclear how information is being received or ambivalent responses:

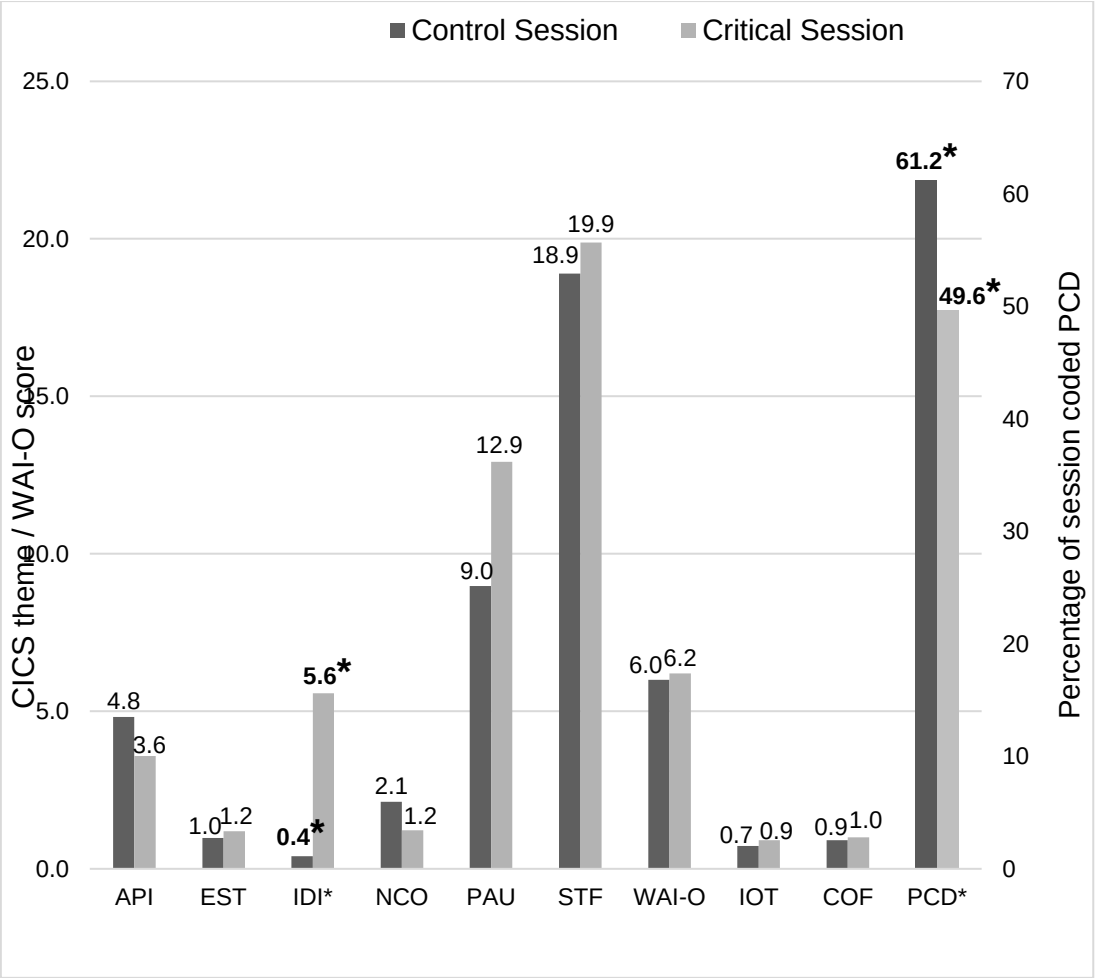
> **Therapist:** But, I mean, it's whether, because I mean pacing and goalsetting is all about setting goals that you think a) are a priority that need to be done but b) that you think that you're capable of doing.
>
> **Patient:** Yeah.
>
> [P 01024, IDI, 0]

A significantly lower percentage of critical sessions were taken up with the description of problems and their context ($Z = -2.05$, $p = .041$). This means that there were fewer interactions where problems were described without observable elements of patient activation:

> **Patient:** That's why I'm quite anxious in a meeting, like wondering what they're going to think of me because of the way I'm acting, yeah? Are they going to think 'he's completely nuts'? Are they going to think 'oh he's not trustworthy'? Do you know what I mean? Things like that. But I can't help it [ ]
>
> **Therapist:** Yeah. And so at this point here, before the meeting, you're probably a bit anxious anyway, just about meeting.
>
> [P 01008, PCD]

There were no significant differences between the critical sessions and control session in any other CICS themes or WAI-O scores (Figure 7.1).

Figure 7.1. Comparative scores for CICS interaction themes and the WAI-O at critical sessions (N) and control sessions (N – 1)

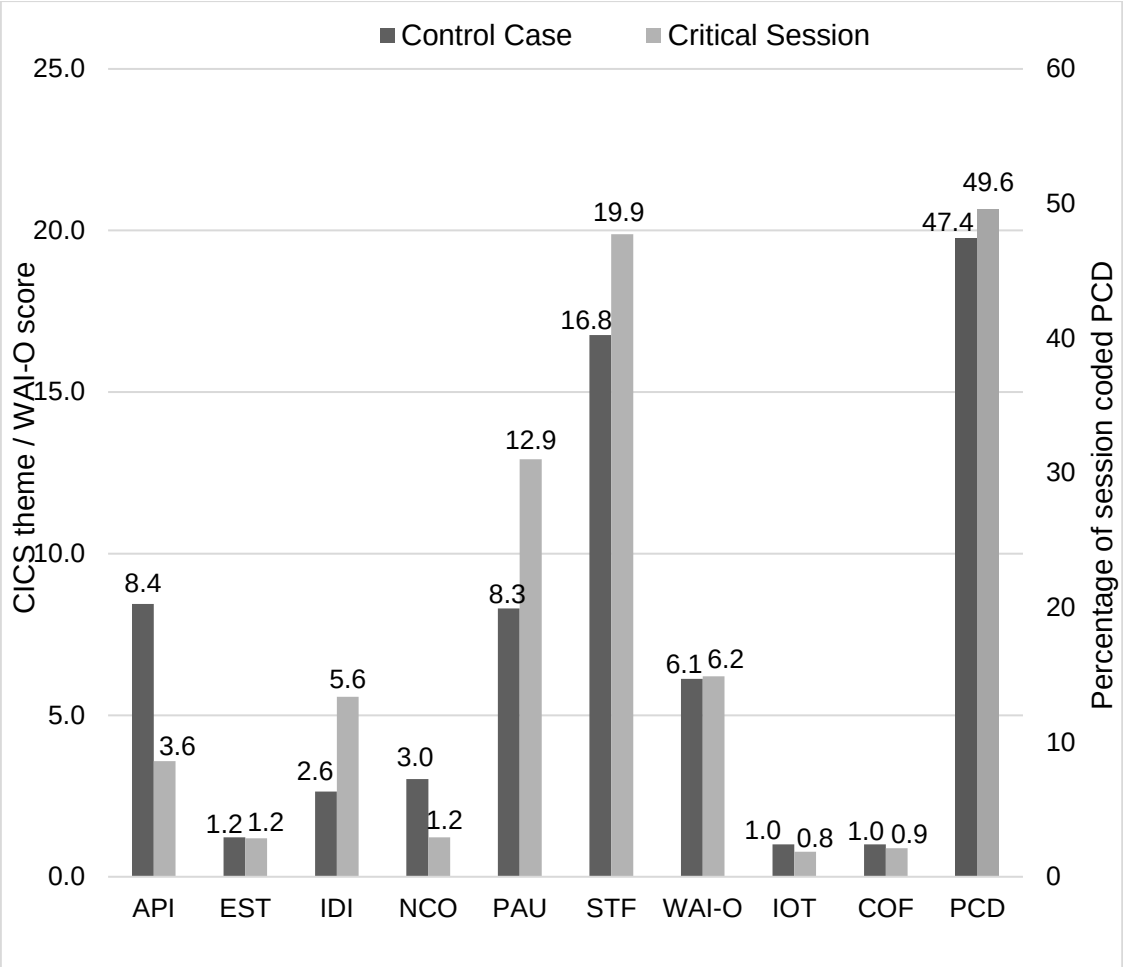


*Note. * p< .05.* Abbreviations: API: Action planning and idea generation; EST: evaluation of self or therapy; IDI: information discussion; NCO: noticing change or otherwise; PAU: problem analysis and understanding; STF: structuring and task focusing; WAI-O: observer rated working alliancing inventory; IOT: Integration of themes; COF: collaborative flow; PCD: problem or context description.

7.3.2.4 Critical sessions versus control cases

There were no significant differences in patient activation on any CICS themes or on the WAI-O when comparing critical sessions of sudden gainers with control sessions from non-sudden gainers. In a non-significant trend, information discussions still had more than double the average score at critical sessions compared to case control session (Figure 7.2).

Figure 7.2. Comparative scores for CICS interaction themes and the WAI-O at critical sessions for sudden gainers and control sessions for matched non-sudden gainers



Note. Abbreviations: API: Action planning and idea generation; EST: evaluation of self or therapy; IDI: information discussion; NCO: noticing change or otherwise; PAU: problem analysis and understanding; STF: structuring and task focusing; WAI-O: observer rated working alliancing inventory; IOT: Integration of themes; COF: collaborative flow; PCD: problem or context description.

7.3.3 Discussion

This study suggests that significant increases in patient activation occur directly prior to sudden gains in CBT for health anxiety, as compared to control sessions. However, these increases are focused on particular interaction themes: positive information discussion – where there are verbal signs that the patient is receptive to and actively engaged with information being given, usually involving psychoeducation. There was also a significant reduction in

the proportion of the session spent *describing* problems. This may relate to trends of increased patient activation in other types of interaction. For example, as problem *description* reduced, there were trends of increased active involvement in problem *analysis*; positive evaluations of self or therapy and structuring increased in a similar non-significant trend. These findings suggest that moving beyond problem description may play a role in the development of sudden gains. The results presented also highlight the value of making psychoeducation a two way process, as large increases in patients' efforts to apply and personalise psycho-educative information preceded sudden gains.

There were also trends for greater patient activation in information discussions amongst those who experienced sudden gains compared to control sessions from patients who did not experience sudden gains. However, none of the between-case tests showed significant differences, suggesting that there are limits to the value of patient activation at critical sessions between cases. Alternatively, greater variability in CICS scores between cases could have been caused by individual differences in interaction-style between different therapists and patients. Combined with a small patient sample, significant findings could have been masked by low statistical power.

7.3.3.1 Strengths and limitations

This study applied the within-case paired comparison design originally used to assess cognitive changes prior to sudden gains (Tang & DeRubeis, 1999b). As such, it replicates an established method for assessing meaningful changes prior to sudden gains. Using the session directly prior to the critical session as the comparison is advantageous in being very close to the critical

session in time and sequential order. Furthermore, both are prior to the sudden gain. The addition of turn-by-turn ratings of session interactions is an enhancement on previous methods of evaluating critical sessions in sudden gains. This is a more rigorous method than the overall session ratings taken by Tang and DeRubeis (1999b), because it reduces the potential for rater biases that might be emphasised by single, overall session ratings. Consistently high IRR across CICS themes with the exception of STF (.61) mean the ratings reported can be viewed as dependable measures. The between-case comparison adds an initial understanding of how the contents of a critical session may compare across cases. Focusing on cases matched by age and gender may have helped overcome some interaction-style differences that are more difficult to account for in varied samples.

Although this study gives broad support for the importance of patient activation in sudden gains, it is unclear how to interpret the CICS themes where null results were observed. In particular noticing change interactions were highlighted in the previous chapter and in Study 1 as an important predictor of outcome, but this is not borne out in the current findings. The sample size may have been underpowered to detect significant differences. As such, this study can be seen as offering only initial indications which require future confirmation. The weakest component of this study, from an experimental perspective, was the between case analysis. This was particularly problematic as paired cases' control sessions were not matched in order. Therefore, compared control and critical sessions did not occur at the same point in therapy, despite occurring at a similar point overall. Although session 2 was the most common point for sudden gains to occur they were

spread across therapy up to the 11th session. Therefore, a comparison between the second session and the 11th session in such a small sample may not be appropriate. Furthermore, control cases could have been matched on variables other than age and gender, perhaps factors relating to outcome. Together, these criticisms suggest that greater control over between case analyses would be required for more dependable results. As sudden gains have been previously associated with cognitive changes in CBT, another weakness of the design is the absence of concurrent assessment of cognitive change to benchmark CICS results against.

In summary, this study posits that facilitating active patient engagement during information-giving interactions may be an important means of maximising therapeutic benefits. This is particularly pertinent when psychoeducation is often central to early session content in CBT and other therapies. The reductions in problem description and increases in other patient activation processes again suggest moving beyond problem description is important for therapeutic benefit. Specifically, this study indicates that reduction in problem description and more active patient engagement in other therapeutic interactions may form part of the conditions that promote sudden gains. However, establishing temporal precedence does not infer causality. The reduction in problem description and increased engagement with psychoeducation could both be effects of changes that have already occurred, rather than the cause. Nonetheless, when taken in the context of Study 1's findings this study supports the possibility of increased patient activation contributing to sudden gains. Therefore, therapists can be encouraged to move beyond problem description to interactions where patients can make

more active attempts to operate upon identified issues. Current results say this shift can take place at an early stage to beneficial effect. Overall, this study supports the possibility that interaction-level patient activation may be an important contributing factor in the development of sudden gains.

7.4 Overall discussion

These two studies show a predictive relationship between interaction-level patient activation and important session-level changes in outcome. As with the whole of this thesis, the level of analysis allows a detailed and nuanced picture to be drawn. In the context of other studies reported, the most consistent finding supported here is that reduced problem description interactions predict improved outcomes. This has been found across all the assessments made in the thesis. This chapter suggests that reduced problem description predicts greater improvement in wellbeing over the course of CBT for health anxiety. There is also a significant drop in problem descriptions directly prior to a sudden gain, suggesting that reduced problem description (combined with increased patient activation in other areas) may be a predictive indicator of ensuing sudden gains. Study 1 allows us to illustrate the characteristics of this effect: Results suggest that approximately a 6% reduction in problem description at session 1 would lead to a 1 point improvement in the ORS over the course of therapy. Therefore, it would not take enormous changes to the interactive composition of session 1 to potentially make significant clinical outcome improvements.

This chapter also highlights that two interaction themes have predictive value to proximal, session-by-session outcome which had not previously been identified as significant predictors of outcome: Information discussions showed

significant increases prior to sudden gains and positive evaluations of self or therapy increased at session 1 and 2 as outcome trajectories improved. Confidence was tempered in relation to facilitative and general interactions as an outcome predictor. The findings from long-term outcomes were not replicated in proximal outcomes. Whereas, the case for structuring and task focus interactions was reopened; it was found to be a significant predictor of session-by-session outcome from session 2 ratings, as it was from early session in developmental work. Again, the predictive validity of some CICS themes remained when controlling for therapeutic alliance, baseline severity and change over time. This confirms that interaction-level patient activation may offer an important additional dimension to process-outcome research to therapeutic alliance.

However, two interaction themes have shown no predictive validity in either this study or any previously conducted: Action planning and idea generation and Problem analysis and understanding. As evaluations from first and second sessions are a very challenging test for a new instrument, it may be premature to reject these interaction themes. They could become potentially important interaction-types later in therapy. This point is emphasised by findings from therapeutic alliance assessments from session 1 and 2. Therapeutic alliance ratings have consistently shown no predictive validity in any assessments. However, given the weight of existing evidence in favour of therapeutic alliance as an outcome predictor, particularly using the WAI-O, one might expect that predictive validity would emerge from ratings of later therapy sessions. Important to this discussion is the point that the vast majority of existing studies showing a relationship between therapeutic alliance

and outcome were taken from later points in therapy than the second session (Flückiger et al., 2018). Also, many studies were not focused on CBT as the treatment. Much alliance-outcome research has focused on treatments where the therapeutic relationship is seen as the main vehicle for change (e.g. psychodynamic psychotherapy; Luborsky et al., 1988).

This pattern in findings between alliance and outcome may help explain why neither of the global interaction themes were again not found to be important outcome predictors. Although there is likely overlap between therapeutic alliance and all CICS themes, the global themes are perhaps the most closely related. The global themes characterise the overall pattern of interaction-style across a session in a similar way to the WAI-O. As with therapeutic alliance, detailed longitudinal analysis would help clarify the predictive validity of global themes across the course of psychological therapy.

The relationship between CICS themes and distal, then proximal outcomes have been shown in the previous chapter and this chapter respectively. However, the relationship between the outcomes themselves has not been investigated thus far. If there is little relationship between proximal outcomes (ORS scores across therapy) and distal outcomes (a range of standardised, longer-term, independently assessed health outcomes), then the CICS' predictive validity is undermined. Such a finding would suggest that therapy ROM outcomes have little bearing on patients' long-term health, thus potentially nullifying the findings from this chapter. Alternatively, if there are strong relations between proximal and distal outcomes it strengthens the value of this chapter's results. Such findings would mean that predictive validity demonstrated with outcomes close to the target events have a significant

bearing on other important outcomes much further away in time. This would also be important for the long-term significance of sudden gains. If sudden gains identified on the ORS do not have any clinical significance long-term on other outcomes, it undermines the value of the analysis presented. Conversely, if ORS sudden gainers do have better long-term outcomes then the changes in patient activation observed at critical session could have an impact on long-term health.

Given the importance of the relationship between proximal and distal outcomes in evaluating the CICS predictive validity, the next chapter investigates the association between proximal and distal outcomes. Chapter 8 considers whether proximal therapeutic outcomes from the same sample predict long-term outcomes over time. The next chapter also examines the significance of sudden gains identified on the ORS to long-term health outcomes post-treatment. This will aid the interpretation of findings from chapters' 6 and 7, by linking the outcomes predicted.

8. ESTABLISHING THE VALUE OF PROXIMAL OUTCOMES

8.1 Introduction

Having presented the predictive validity of the CICS to proximal and distal outcomes, this chapter investigates the relationship between the session-by-session ROM outcomes used and the numerous long-term health outcomes. The primary purpose of this chapter is to evidence the strength of the relationship between the outcomes used in the two previous chapters, giving an indication of the dependability that can be placed on the CICS predictive validity. This chapter also answers an important research question in its own right: Although studies investigating ROM and the ORS have garnered much support for session-by-session outcome measurement, there are questions over the validity of ROM to long-term independent health outcomes. This chapter will also give a relatively rare assessment of ROM's predictive validity in long-term, independently assessed health outcomes.

The Partners for Change Outcome Monitoring System (PCOMS; including the ORS as the main outcome) has been shown to have beneficial effects on outcome as compared to usual care and reduces the number of patients showing no improvement at the end of treatment (Lambert et al., 2018). However, it does not reduce the number of patients dropping out or deteriorating. Furthermore, Østergård et al. (2018) found that PCOMS only had a beneficial treatment effect when outcome was measured with the ORS, but not when independent measures were used. The same was found in three studies using independent outcome measures in the nine-study sample investigated by Lambert, Whipple, and Kleinstäuber (in-press). This illustrates

a central unaddressed issue within ROM research: The unclear relationship between ROM outcomes and other important independent outcomes.

8.1.1 Unaddressed issues in ROM

The case is presented specifically for PCOMS, but in many RCTs assessing the effectiveness of ROM generally, there have been criticisms for using the ROM measure as the primary outcome (Wampold, 2015). This means that the ORS is frequently used both as the ROM intervention and the outcome of the ROM intervention, meaning one cannot tell the independent value of ROM interventions. Even when independent assessments are used, they are typically general in nature with few studies including assessment of quality of life, functioning, or disorder-specific outcomes (Kendrick et al., 2014). Furthermore, ROM RCTs and meta-analyses have almost all been directly or indirectly conducted by researchers with an allegiance to the method investigated, which tends to inflate effectiveness (Luborsky et al., 1999). This is supported by evidence that RCTs carried out more independently have had less positive results (e.g. Rise, Eriksen, Grimstad, & Steinsbekk, 2016). Another element commonly missing from ROM research is independently-assessed, longer-term outcome follow-up (Kendrick et al., 2014). As ROM studies often use outcome data collected by therapists during treatment, there are fewer studies where outcome data are collected by independent or blinded assessors at fixed intervals after treatment (Østergård et al., 2018). The absence of independent longer-term follow-up or blinding is another factor likely to inflate the treatment effect, due to both assessor bias and the potential for treatment effects to diminish after completion (Wood et al., 2008).

8.1.2 Comorbidity and ROM

From this type of evidence, it is difficult to assess the independent predictive ability of ROM measures on treatment outcomes for specific disorders, because of the lack of relevant assessments and because studies are typically conducted by those with an allegiance. This situation is especially problematic in mental health problems that commonly co-occur with physical health problems, as key outcomes would not be assessed by ROM measures. Comorbid physical and mental health problems lead to poorer health outcomes for patients and higher costs for services (Barnett et al., 2012; Naylor et al., 2016). Demonstrating that a commonly-used ROM tool is predictive of more specific, independently-assessed physical and mental health outcomes would expand our understanding of the value such ROM assessments hold. However, it is unclear whether the most widely-used ROM measures (such as the ORS) can effectively assess the types of outcome that are important where comorbid physical and mental health problems are present. Therefore, patients who appear to be improving on ROM measures may not improve in important unassessed areas to which the clinician is blind. Alternatively, the tailored care which is integrated within ROM systems may make ROM meaningful even for problems with physical and mental health sequelae, such as health anxiety.

8.1.3 Health anxiety and comorbidities

Cognitive behavioural therapy has proven effective for mental health problems that are frequently comorbid with chronic physical health problems; health anxiety is one of the most common examples (e.g. P. Tyrer et al., 2017). Symptoms of health anxiety are characterised by an interplay between physical and psychological difficulties. Given this interaction, if health anxiety

were to improve, one might expect concurrent improvement in general health too (Morriss et al., 2019) . Yet this type of change may not be picked up in routine practice using ROM. Conversely, ROM may give a false indication of improvement when core symptoms remain. Therefore, it is unclear whether ROM tools can make a meaningful contribution to psychological therapy outcome assessment of specific disorders, such as health anxiety, which often involve comorbid physical and mental health problems. Independent assessment of a range of outcomes over time would give a clearer understanding of the value ROM may hold for specific disorders, especially comorbid physical and mental health disorders.

8.1.4 Sudden gains and ROM

A similar situation is apparent with sudden gains in ROM: Sudden gains have been identified from ROM data, but it is unclear whether they translate into improvements on other outcomes (Lambert, 2010). Sudden gains are defined as large, sustained, clinically meaningful improvements in symptoms, typically occurring early in psychological treatment (Tang & DeRubeis, 1999b). Those experiencing sudden gains often have better long-term outcomes than those without (Aderka et al., 2012). This has been most clearly evidenced in CBT for depression (Tang et al., 2005; Tang et al., 2007). Beneficial effects of sudden gains have also been evidenced in health anxiety, but eligibility criteria restricted the likelihood of comorbidity and no outcome assessments were taken for other disorders (Hedman, Lekander, et al., 2014). Little crossover research has been conducted between sudden gains in disorder-specific outcomes and those occurring on ROM tools.Therefore, it is unclear whether sudden gains on ROM tools have any impact on long-term disorder-specific

outcomes. Moreover, sudden gains have not been studied in populations with high rates of multimorbidity. Thus, it is unclear whether the phenomenon still has clinical significance where multiple problems are present or what outcome measures are best for assessing sudden gains in this context.

This study assessed the predictive validity of a ROM assessment (the ORS) on long-term health anxiety severity measured independently up to 12 months after CBT for high service utilisers with health anxiety (Morriss et al., 2019) . The study also assessed the predictive value of ROM on other anticipated effects of health anxiety: overall health, functional impairment, quality of life, somatic symptoms, generalised anxiety, and depression. Lastly, this study assessed whether sudden gains in ROM have a significant impact on long-term trajectories of health anxiety and the other outcomes.

8.2 Method

8.2.1 Design and participants

The predictive value of the ORS for a range of independently-assessed health outcomes over time was assessed using an observational design. The study applied a secondary analysis to data from the same RCT used in the previous two chapters, evaluating the effectiveness of remotely delivered CBT for high service utilisers with health anxiety compared to TAU (Morriss et al., 2019). All 79 participants receiving CBT were offered weekly CBT adapted to health anxiety over video conferencing or the telephone. All those attending sessions were offered the use of ROM by their therapists with the ORS completed at each session. In parallel, a series of physical and mental health outcome assessments were taken at baseline, 3-, 6-, 9-, and 12-month follow-up by independent researchers blinded to treatment group.

Fifty-six participants (71%) in the treatment arm attended two or more sessions of CBT and completed the ORS. All were included in the first stage of this study, to assess whether the ORS score at the final session predicted follow-up health outcomes. The predictive validity of the ORS was directly compared to 6-month follow-up assessment of health anxiety as a predictor of all secondary outcomes. The 57 participants who had responded to treatment at follow-up were included in a sensitivity assessment: Those who responded to treatment and experienced a sudden gain on the ORS were compared to the remaining treatment responders to see if sudden gains in routinely-monitored outcomes conferred a long-term clinical benefit.

8.2.2 Instruments

8.2.2.1 Routine Outcome Monitoring

The ORS is a commonly used ROM outcome measure that uses four self-reported, brief visual analogue scales to assess wellbeing (S. D. Miller et al., 2003). A detailed description of the ORS is given in the previous chapter.

8.2.2.2 Independent Assessments

Due to high rates of comorbidity amongst those with health anxiety, a range of outcomes were assessed independently (greater detail is offered in previous chapters):

1. SHAI (Salkovskis et al., 2002).
2. GAD-7 (Spitzer et al., 2006).
3. PHQ-15 (Kroenke et al., 2002).
4. PHQ-9 (Kroenke et al., 2001).
5. WSAS (Mundt et al., 2002).
6. EQ5D-5L (Herdman et al., 2011).

7. VAS (Herdman et al., 2011).

The SCID (First et al., 1998) was also administered at baseline. Long-term physical health conditions were recorded from baseline patient interviews.

The approach used mitigates common methods variance in some ways when associating the ORS with other outcomes. In most outcome prediction studies, both the predictor and outcome measurements are completed with the same assessors, which is likely to inflate associations (Campbell & Fiske, 1959), whereas in the current study, the ORS uses a visual analogue scale completed with therapists to compare with questionnaire-based measures (with the exception of the VAS) completed with independent researchers. This reduces the likelihood of inflated associations and type-I errors.

8.2.3 Procedure

8.2.3.1 Remote CBT

As described in previous chapters, a team of four CBT therapists remotely delivered CBT adapted to health anxiety using an established treatment protocol (Morriss et al., 2019).

8.2.3.2 Administration of the ORS

The ORS was completed by patients with their therapist at every session of CBT, unless patients refused or did not attend. Therapists used the ORS to tailor treatment by focusing on key problems of importance to the patient in each domain and setting goals associated with points on the scales. For example, if a patient highlighted social avoidance as a key problem that was reducing the score on the social functioning scale to 3, the therapist might establish what would be different if they were functioning ‘at the level of a 4’. They could then work towards what the patient describes as a short-term goal.

The ORS was completed on software which gave a computerised traffic light system (www.myoutcomesukapp.com), alerting therapists to patients who were at risk of non-improvement or deterioration, from algorithm predicted outcome trajectories (S. D. Miller et al., 2005). If patients were identified as at risk of a poor outcome, therapists initiated conversations to help identify unaddressed problems and collaboratively adjust the course of therapy. All trial therapists used ORS data to identify patients on their caseload at risk of poor outcome. They then attended monthly supervision focused on addressing issues with these cases in particular, using observation of recordings from previous sessions to guide decision-making.

8.2.4 Method of analysis

The baseline characteristics of included and excluded patients were compared using Mann-Whitney U tests for continuous variables and chi-squared for categorical variables. Pearson's correlation coefficients were calculated for associations among baseline outcome scores and Spearman's rho was used to calculate correlations among post-treatment scores (last session score for the ORS and 6-month follow-up for all other measures).

8.2.4.1 Predictive modelling

Due to the repeated measurement design, MLM was used to assess whether ORS total scores at the last session separately predicted each health outcome individually. Data were structured in a two-level model with repeated follow-up assessment time-point ("*time*" hereafter) at level 1, nested within patients at level 2. Unconditional models were used to assess clustering within patients and linearity of trends before adding explanatory variables. Model fit was explored using -2 log likelihood and parameter estimation was made using

maximum likelihood. Each parameter was added to the model in a stepped manner. As each parameter was added, change in -2 log likelihood and change in degrees of freedom were used to calculate whether a significant reduction in model deviance was caused using chi-squared. Table 8.1 illustrates the methods used for model development where health anxiety is the outcome.

Table 8.1. Illustration of MLM development for last session ORS predicting health anxiety over time.

Step	Model	Deviance (-2LL)	Change in deviance, new dfs	Was parameter added statistically significant?[a]	Residual variance	Intercept variance	Slope variance
1	Unconditional	1271	-	-	16.53	34.39	-
2	Fixed effects and random intercept for time	1237	34, 1	yes	12.92	35.95	-
3	Random slope for time (removed)	1235	2, 1	no	11.93	31.35	0.65
4	Control variable 1: Fixed effect of baseline SHAI	1010	227, 1	yes	13.59	20.83	-
5	Control variable 2: Fixed effect of 1st session ORS	1004	6, 1	yes	13.64	17.55	-
6	Control variable 3: Number of morbidities	1003	1, 1	no	13.65	17.26	-
7	Main predictor: Fixed effect of last session ORS	998	5, 1	yes	13.59	15.48	-

Note. Abbreviations: -2LL: -2 log likelihood; dfs: degrees of freedom; ORS: outcome rating scale; WAI-O: working alliance inventory – observer rated.
[a]Statistical significance was assessed using chi-squared for the change in -2LL and change in dfs.

Fully adjusted models included the following fixed predictor variables for each outcome: *time; baseline outcome score* (controlling for baseline outcome severity); *first session ORS score* (controlling for initial ORS severity); *total number of mental and physical health morbidities* (controlling for comorbidity); and *last session ORS score* (the main predictor of interest). Random intercepts for *time* were included in all models, but random slopes were not. This was either due to non-significance in fully adjusted models or non-significant change in -2 log likelihood when slope was added (using chi-squared). Below is an illustration of the equation tested using health anxiety (SHAI):

$$y_{ij} = \beta_{0i} + \beta_1 \text{time}_{ij} + \beta_2 \text{Baseline SHAI}_{ij} + \beta_3 \text{First session ORS}_{ij} + \beta_4 \text{Total disorders}_{ij} + \beta_5 \text{Last session ORS}_{ij} + e_{ij}$$

Where y_{ij} is the SHAI score for the $_j$th participant at the $_i$th follow-up time point. As such j is participants (level 2 unit) and i is follow-up time order (level 1 within level 2 unit). β_{0i} indicates that estimates of the intercept for both the overall model and variability of level 1 intercepts were calculated. β_1 through β_5 were fixed level 2 effects, adjusting for differences between participants. e_{ij} is the residual or unexplained variance in the model.

The same methods were used to build models with 6-month health anxiety (SHAI) as the focal predictor-of-interest for each secondary outcome in turn. This is illustrated below, where y_{ij} is the generalised anxiety (GAD-7) score for the $_j$th participant at the $_i$th follow-up time:

$y_{ij} = \beta_{0i} + \beta_1 time_{ij} + \beta_2$ Baseline GAD-$7_{ij} + \beta_3$ Baseline $SHAI_{ij} + \beta_4$Total morbidities$_{ij} + \beta_5$ 6-month $SHAI_{ij} + e_{ij}$

Fully adjusted models included the same fixed predictor variables for each outcome with SHAI in exchange for ORS: *time; baseline outcome score*; *baseline SHAI*; *total morbidities*; *SHAI at 6 months* (the target predictor). Random intercepts, but not slopes, were again included in all models.

Parallel models were constructed to compare the predictive validity of the ORS versus SHAI – facilitating evaluation of the predictive ability of the ORS against that of the primary treatment target and outcome variable of the trial (SHAI). This comparison was conducted using a pseudo-R^2, which gives a measure of effect size in MLM, enabling a direct comparison between the ORS and SHAI (Raudenbush & Bryk, 2002). Using the model developed in Table 8.1 as an illustration, the pseudo-R^2 was calculated as the percentage of: $((\sigma_6^2-\sigma_7^2)/ (\sigma_1^2))$, where σ_1^2 is the unconditional model intercept variance (step 1), σ_6^2 is intercept variance in the fully adjusted model *without* the focal predictor (step 6) and σ_7^2 is the fully adjusted model's intercept variance (step 7). This shows the percentage reduction in unexplained variance between participants when post treatment SHAI/last-session ORS are added to fully adjusted models ("full" below) (Hox et al., 2017):

$$Pseudo\ R^2 = \frac{(variance\ full\ without\ target)-(variance\ full\ with\ target)}{variance\ unconditional}$$

Analysis was conducted using SPSS 24.

8.2.5 Sudden gains modelling

The same adaptation of Tang and DeRubeis' (1999b) definition of sudden gains was applied to ORS scores as in Chapter 7. This assessed for

large intersessional gains between session N and N + 1 if the following criteria were met:

- Sudden gains should be large in absolute terms – specifically, the N + 1 gain needed to be at least 5 points, the established reliable change index on the ORS.
- Sudden gains should be large in relative terms – specifically, the N + 1 gain needed to be at least a 25% improvement on session N.
- Sudden gains should be sustained beyond random fluctuations – specifically, the mean ORS score of the three sessions prior to the gain should be at least a reliable change (5 points) lower than the mean ORS score for the three sessions following the sudden gain. Fewer than three sessions prior to the gain were allowed as long as the gain was sustained for the three ensuing sessions. Conversely, fewer than three sessions were allowed after a gain assuming that the pre-gain score was stable for the previous three sessions.

In order to assess whether these adapted criteria were of clinical value, results were compared to those achieved using the more strict criteria from the original method: To be identified as a sudden gain, the mean ORS score of the three session scores prior to the gain were required to be significantly lower than the three sessions after the gain using an independent t-test, with an alpha of .05. Two sessions before and after were used if there were no more than two sessions before or after the gain, otherwise three were required.

Treatment responders who experienced sudden gains were compared to treatment responders who did not experience sudden gains. Treatment responders were defined as those experiencing ≥ 4 point reduction in health

anxiety score at follow-up, as this is deemed the minimum clinically important change index on the SHAI (Clark & Oates, 2014). Again MLM was used to account for the hierarchical structure of level 1 *time* clustered within level 2 *patients*. Similar methods of model development as previously described were used. Random intercepts and fixed slopes for *time* were included in all models. Four fixed predictors were entered as covariates in each model: *time*, *baseline outcome score, total morbidities* and *sudden gains* (a binary variable assessing the difference in growth curves over time of those identified as sudden gainers versus remaining treatment responders). As well as independently-assessed outcomes over 12-month follow-up, a model was developed which used session-by-session ORS scores as the ordered *time* parameter. This model evaluated whether sudden gainers had significantly improved ORS trajectories during treatment sessions.

Ethical approval was obtained from the National Research Ethics Service, London-Riverside Committee (reference 14/LO/1102).

8.3 Results

8.3.1 Sample characteristics

There was a greater proportion of females included in the assessment of the ORS' predictive validity than amongst those without available data (χ^2 = 3.95, p = .047). No other demographic or clinical characteristics were significantly different at baseline (Table 8.2).

Table 8.2. Demographics and clinical characteristics of included and excluded participants

Demographics	**Included Sample N=56**	**Excluded Sample N=23**
Females	44 (79%)	13 (57%)
Mean age (SD)	37 (17)	40 (19)
Ethnicity		
White British	42 (75%)	17 (74%)
Unemployed	6 (11%)	6 (26%)
Clinical Characteristics		
Baseline wellbeing - ORS (SD)	20.5 (6.5)	-
Baseline health anxiety - SHAI (SD)	27.2 (5.2)	27.7 (6.1)
Baseline depression – PHQ-9 (SD)	13.4 (6.4)	13.1 (6.4)
Baseline anxiety – GAD-7 (SD)	13.3 (5.4)	12.9 (5.6)
Baseline somatic symptoms – PHQ-15 (SD)	13.9 (5.1)	14.0 (6.5)
Baseline functional impairment – WSAS (SD)	17.0 (11.1)	17.6 (11.8)
Baseline quality of life – EQ5D-5L (SD)	.624 (.245)	0.572 (.353)
Baseline general health – VAS (SD)	54.0 (21.5)	55.1 (23.5)
Mean number of SCID Diagnoses **(range)**	6 (0-16)	6 (0-13)
Generalised anxiety disorder	38 (68%)	15 (65%)
Hypochondriasis	35 (63%)	12 (52%)
Current depressive episode	34 (61%)	15 (65%)
Somatoform disorder	32 (57%)	13 (57%)
Panic disorder	32 (57%)	11 (48%)
Mean chronic physical health problems **(range)**	1 (0-6)	1 (0-3)
Chronic pain conditions	17 (30%)	4 (17%)
Bowel disorders (IBS, colitis, diverticulitis)	8 (14%)	1 (4%)
Chronic fatigue	6 (11%)	0
Functional neurological disorders	5 (9%)	0
Diabetes	3 (5%)	1 (2%)

Note. SD: standard deviation; ORS: outcome rating scale; SHAI Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale; SCID: structured clinical interview for DSM IV disorders.

8.3.2 Outcome Correlations

Significant weak-to-moderate correlations were observed between first-session ORS scores and baseline scores for other outcome measures in the expected directions, except for a lack of correlation with baseline health anxiety (absolute *r*s = .055 – .502; Table 8.3). Weak-to-moderate correlation was found between baseline health anxiety (SHAI) and generalised anxiety (GAD-7), but no correlations were observed with any other outcomes. No correlations between last-session ORS and other post-therapy outcome measures were of more than moderate magnitude (absolute r_Ss = .310 – .446; Table 8.4). A similar pattern of significant weak-to-moderate correlations were observed between 6-month follow-up health anxiety and other post-treatment outcomes.

Table 8.3. Pearson's correlations between baseline outcome scores and wellbeing at the first treatment session.

	ORS	SHAI	EQ5D	GAD-7	WSAS	VAS	PHQ-9	PHQ-15
Wellbeing (ORS)	1	-.055	.496**	-.400**	-.336*	.293*	-.502**	-.340*
Health anxiety (SHAI)		1	.043	.389**	.044	-.161	.22	.096
Quality of Life (EQ5D)			1	-.336*	-.523**	.570**	-.488**	-.401**
Generalised anxiety (GAD-7)				1	.253	-.479**	.634**	.314*
Functional impairment (WSAS)					1	-.411**	.654**	.521**
General health (VAS)						1	-.430**	-.304*
Depression (PHQ-9)							1	.632**
Somatic symptoms (PHQ-15)								1

Note. *p < .05 **p < .01. n = 56. Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

Table 8.4. Spearman's correlations between 6-month follow-up outcome scores and wellbeing at the last treatment session.

	ORS	SHAI	EQ5D	GAD-7	WSAS	VAS	PHQ-9	PHQ-15
Wellbeing (ORS)	1	-.435**	.388*	-.365*	-.446**	.354*	-.433**	-.310*
n	*56*	*46*	*37*	*45*	*45*	*46*	*45*	*46*
Health anxiety (SHAI)		1	-.449**	.520**	.286	-.340*	.482**	.333*
n		*46*	*37*	*45*	*45*	*46*	*45*	*46*
Quality of Life (EQ5D)			1	-.351*	-.786**	.591**	-.721**	-.762**
n			*37*	*36*	*36*	*37*	*36*	*37*
Generalised anxiety (GAD-7)				1	.364*	-.445**	.739**	.278
n				*45*	*45*	*45*	*45*	*45*
Functional impairment (WSAS)					1	-.493**	.660**	.587**
n					*45*	*45*	*45*	*45*
General health (VAS)						1	-.666**	-.556**
n						*46*	*45*	*46*
Depression (PHQ-9)							1	.648**
n							*45*	*45*
Somatic symptoms (PHQ-15)								1
n								*46*

Note. *p < .05 **p < .01. Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

8.3.3 Predictive validity of the ORS

The assumption of normality was not met for the data analysed. Therefore, the analysis described below was rerun with a bootstrap of 1,000 samples to give a more accurate confidence interval estimate and greater assurance of the reliability in reported findings. These analyses confirmed the same results as described and are presented in Appendix O.

In unconditional models for all outcomes significant covariance was explained by clustering within patients, indicating that a two-level hierarchical structure was appropriate to account for this effect. Last-session ORS scores significantly predicted health anxiety over time. This meant that improvement on the ORS (higher scores) predicted lower severity of health anxiety (lower scores). The same was true of depression, generalised anxiety, somatic symptoms, and functional impairment. Significant predictive relationships were found between last-session ORS scores and quality of life. Therefore, improvement on the ORS predicted improvement in quality of life and the same was true of general health. All models controlled for baseline ORS, baseline outcome score, number of comorbidities, and changes over time (Table 8.5).

Table 8.5. Predictive validity of patient wellbeing at the last treatment session on follow-up outcomes over 3, 6, 9 and 12-month follow-ups, controlling for baseline severity and changes over time.

Health anxiety (SHAI)

Fixed effects						Random effects				
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	22.53	5.01	<.001	12.47	32.59	Residual variance	13.59	1.75	7.75	<.001
Assessment time point	-1.53	0.26	<.001	-2.05	-1.01	Intercept variance	15.48	3.93	3.94	<.001
Baseline health anxiety (SHAI)	0.22	0.13	.107	-0.05	0.49					
session 1 wellbeing (ORS)	-0.08	0.10	.391	-0.28	0.11					
Total morbidities	0.44	0.20	.028	0.05	0.83					
Last session wellbeing (ORS)	*-0.34*	*0.09*	*<.001*	*-0.51*	*-0.16*					

Somatic symptoms (PHQ-15)

Fixed effects						Random effects				
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	13.21	3.55	.001	6.06	20.36	Residual variance	7.84	1.05	7.44	<.001
Assessment time point	-0.76	0.20	<.001	-1.16	-0.36	Intercept variance	9.72	2.65	3.67	<.001
Baseline somatic symptoms (PHQ-15)	0.52	0.10	<.001	0.31	0.73					
session 1 wellbeing (ORS)	-0.11	0.08	.184	-0.27	0.05					
Total morbidities	0.004	0.14	.978	-0.28	0.29					
Last session wellbeing (ORS)	*-0.19*	*0.07*	*.008*	*-0.33*	*-0.05*					

Generalised anxiety (GAD-7)

Fixed effects						Random effects				
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	16.27	2.74	<.001	10.75	21.78	Residual variance	10.63	1.43	7.45	<.001
Assessment time point	-0.90	0.24	<.001	-1.36	-0.43	Intercept variance	5.12	1.82	2.82	.005
Baseline generalised anxiety (GAD-7)	0.15	0.09	.087	-0.02	0.33					
session 1 wellbeing (ORS)	-0.11	0.07	.129	-0.26	0.03					
Total morbidities	0.30	0.12	.019	0.05	0.55					
Last session wellbeing (ORS)	*-0.33*	*0.06*	*<.001*	*-0.44*	*-0.21*					

Depression (PHQ-9)

Fixed effects						Random effects				
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	17.83	3.89	<.001	10.00	25.65	Residual variance	10.43	1.41	7.41	<.001
Assessment time point	-0.44	0.24	.067	-0.90	0.03	Intercept variance	10.81	3.00	3.61	<.001
Baseline depression (PHQ-9)	0.32	0.10	.002	0.12	0.52					
session 1 wellbeing (ORS)	-0.13	0.10	.190	-0.32	0.07					
Total morbidities	0.001	0.15	.994	-0.31	0.31					
Last session wellbeing (ORS)	*-0.34*	*0.07*	*<.001*	*-0.49*	*-0.19*					

General health (VAS)

Fixed effects						Random effects				
				95% CI						
Parameter	B	SE	*p*	Lower Bound	Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	13.21	3.55	.001	6.06	20.36	Residual variance	7.84	1.05	7.44	<.001
Assessment time point	-0.76	0.20	<.001	-1.16	-0.36	Intercept variance	9.72	2.65	3.67	<.001
Baseline general health (VAS)	0.52	0.10	<.001	0.31	0.73					
Session 1 wellbeing (ORS)	-0.11	0.08	.184	-0.27	0.05					
Total morbidities	0.00	0.14	.978	-0.28	0.29					
Last session wellbeing (ORS)	*-0.19*	*0.07*	*.008*	*-0.33*	*-0.05*					

Quality of life (EQ-5D)

Fixed effects						Random effects				
				95% CI						
Parameter	B	SE	*p*	Lower Bound	Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	-0.018	0.135	.896	-0.289	0.253	Residual variance	0.015	0.002	7.27	<.001
Assessment time point	0.019	0.009	.040	0.001	0.036	Intercept variance	0.025	0.006	4.10	<.001
Baseline quality of life (EQ-5D)	0.643	0.120	<.001	0.401	0.885					
session 1 wellbeing (ORS)	-0.002	0.004	.633	-0.011	0.007					
Total morbidities	0.002	0.007	.817	-0.012	0.016					
Last session wellbeing (ORS)	*0.008*	*0.003*	*.021*	*0.001*	*0.015*					

Functional impairment (WSAS)

Fixed effects						Random effects				
Parameter	B	SE	*p*	95% CI Lower Bound	95% CI Upper Bound	Parameter	Variance	SE	Z	*p*
Intercept	17.30	6.43	0.010	4.38	30.23	Residual variance	30.65	4.14	7.40	<.001
Assessment time point	-1.01	0.41	0.014	-1.82	-0.20	Intercept variance	32.82	9.06	3.62	<.001
Baseline functional impairment (WSAS)	0.59	0.09	<.001	0.40	0.77					
Session 1 wellbeing (ORS)	-0.06	0.15	0.716	-0.36	0.25					
Total morbidities	0.02	0.27	0.926	-0.51	0.56					
Last session wellbeing (ORS)	*-0.46*	*0.13*	*0.001*	*-0.73*	*-0.20*					

Note. Abbreviations: B: unstandardised β coefficient; SE: standard error; p: significance at an alpha of .05; CI: confidence interval;; Z: Walt's z test for significance of variance; ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

8.3.4 Comparison between the ORS and SHAI

The same models were run with the SHAI at 6 months as the target predictor, which demonstrated that it was also a significant predictor of all secondary outcomes over time when controlling for baseline severity. The percentage reduction in unexplained between-participant variance uniquely accounted for by the ORS and SHAI were compared for each model, to give an estimate of comparative effect size (Table 8.6). This analysis indicated that a similar proportion of outcome variance was explained by the ORS and the SHAI across outcomes (average 19% and 18% respectively). There were trends of larger predictive effects where there were large pre-post treatment effect sizes (with health anxiety, generalised anxiety and depression) compared to other outcomes showing smaller treatment effects.

Table 8.6. Pre-post treatment effect sizes and predictive effect sizes for post-treatment wellbeing and health anxiety.

Outcome	Pre-post treatment effect size (Cohen's *d*)[a]	% Reduction in unexplained variance unique to last session ORS	% Reduction in unexplained variance unique to 6-month SHAI
Health anxiety (SHAI)	1.9	30	-
Generalised anxiety (GAD-7)	1.4	35	29
Wellbeing (ORS)	1.3	-	11
Depression (PHQ-9)	0.8	24	27
General health (VAS)	0.7	15	19
Somatic Symptoms (PHQ-15)	0.6	11	16
Functional impairment (WSAS)	0.6	9	10
Quality of life (EQ-5D)	0.2	10	15

Note. [a]Cohen's d effect sizes descriptors: 0.2: small; 0.5: Medium ≥0.8: Large. Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

8.3.5 Sudden gains

Thirty-two (56%) of the 57 treatment responders (72% of total treatment group) were identified as sudden gainers using the adapted criteria. Session 2 was the median session for sudden gains to occur (IQR 1–3). The average sudden gain led to a 53% increase in ORS score and accounted for 80% of the average sudden gainer's pre- to post-therapy improvement (Figure 8.1). Multilevel modelling was used to establish whether there was a significant difference in outcomes over time for sudden gainers versus non-sudden

gaining treatment responders. This analysis showed that sudden gainers had significantly greater improvement in wellbeing over the course of therapy (ORS; *B* = 4.38*, SE B* = 1.36, *p* = .002). Sudden gainers also reported significantly greater improvements in functional impairment (WSAS; *B* = -4.38*, SE B* = 1.84, *p* = .021) and symptoms of depression (PHQ-9; *B* = -2.54*, SE B* = 1.20, *p* = .039) over the 12-month follow-up period. There were trends for greater improvements across all outcomes, though no others reached significance. Figure 8.2 shows the percentage greater improvement achieved by sudden gainers as compared to other treatment responders using baseline scores as the percentage denominator. All models controlled for baseline ORS score, baseline outcome score, number of comorbidities and repeated measurement effects.

Using the strict original criteria for classifying sudden gains, 9 (16%) sudden gainers were identified and no differences were observed on any outcome (all *p*s ≥ .320). The allowance for gains at the earliest and latest points in therapy accounted for the vast majority of those not identified by the strict criteria. In the analysis using the adapted criteria, 22 sudden gains were identified before the third session and two further sudden gains at the final session.

Figure 8.1. The average sudden gain. The averaged ORS scores are shown for the sessions around the pre-gain session (N).

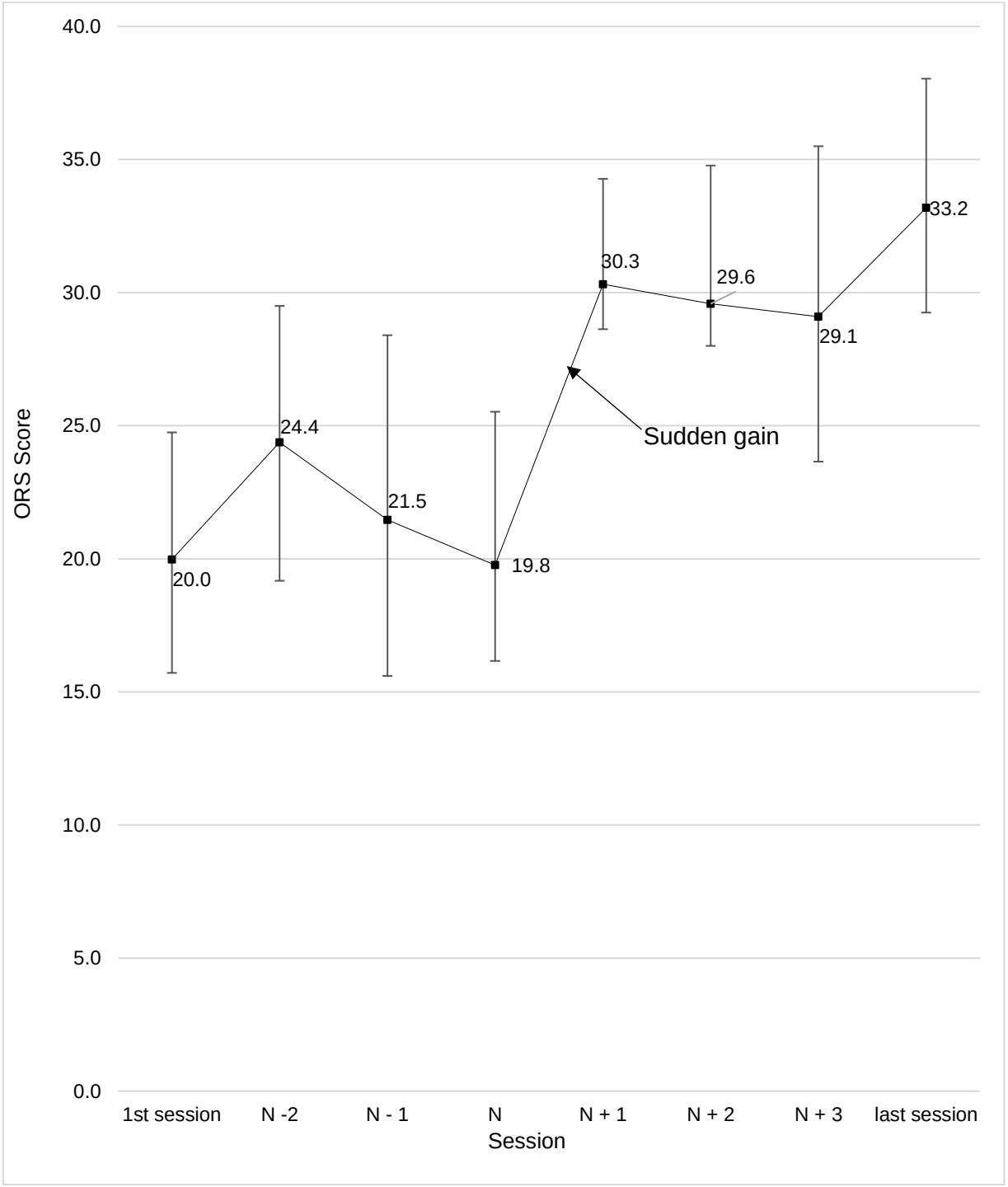
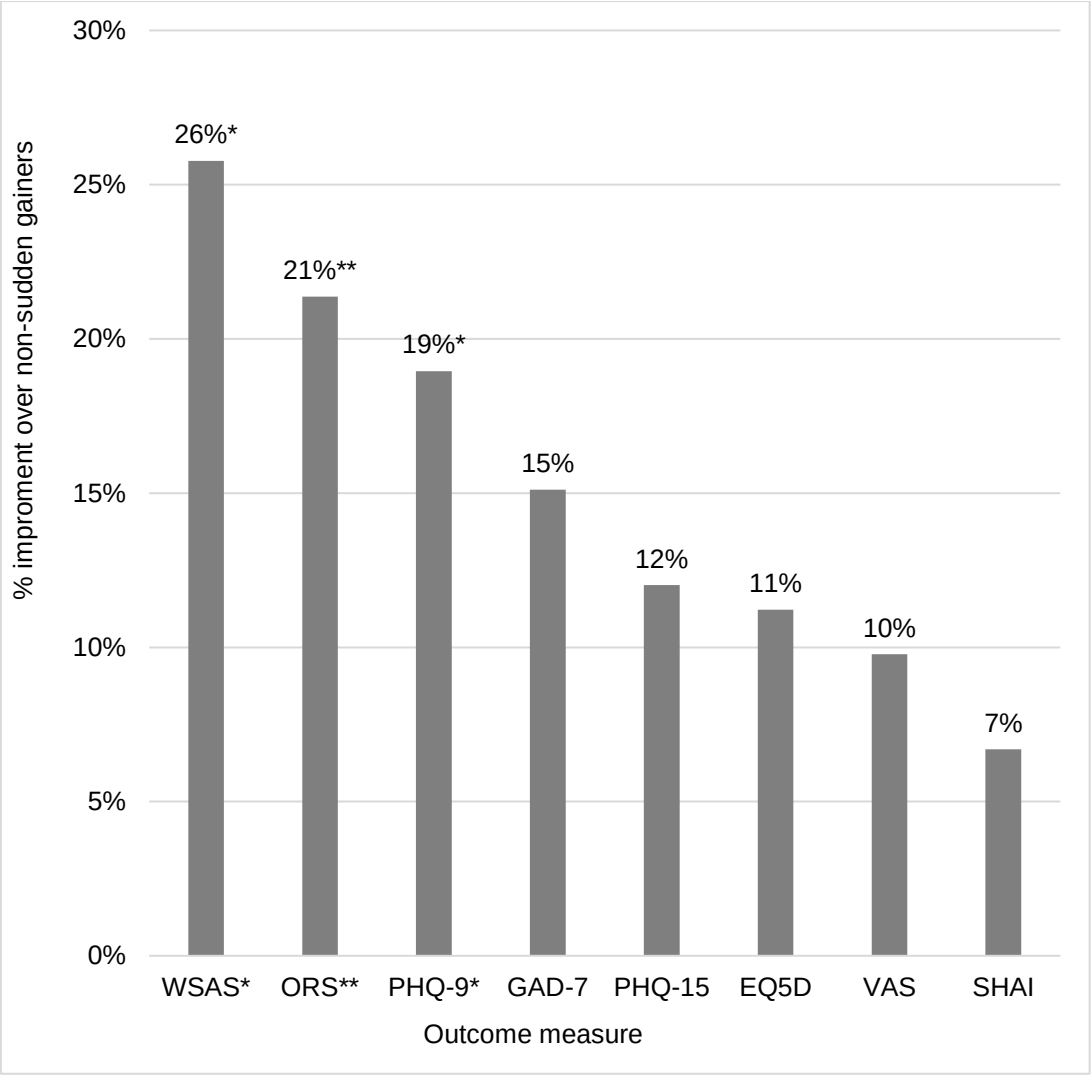


Figure 8.2. Percentage greater outcome improvement achieved by treatment responders with sudden gains versus non-sudden gaining treatment responders.



Note. *p < .05 **p < .01. Abbreviations: ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; EQ5D-5L: Euroqol 5 levels; VAS: visual analogue scale.

8.4 Discussion

8.4.1 Predictive validity of the ORS

This study used longitudinal, independent, single-blinded assessments to evaluate the predictive validity of ORS ratings on a range of health outcomes from CBT for health anxiety. Findings suggest that ORS scores at the last treatment session predict a range of longer-term physical and mental health outcomes, irrespective of how many sessions were attended. Furthermore, the MLM design means that this relationship is less likely to be due to covariance across time, an artefact of baseline problem severity, or related to the number of comorbidities experienced by patients. This is a significant expansion in the use of the ORS as a common ROM outcome measure, because ORS outcome studies are often clinician-assessed with short follow-up periods. The assessments used in previous ORS studies have also been limited in what they can say about the relevance of ROM for specific disorders or its impact on general health (rather than just mental health) and quality of life. In the context of this thesis, the results indicate that the relationships found between the CICS and the ORS could have important implications for longer-term outcomes, even if not identified as predictors of long-term outcome. The large gap in time between the sessions rated and the follow-up outcomes assessed, means that important relationships could have been eroded by intermediary processes. But this chapter identifies that ROM outcomes could be indicative of long-term outcomes.

This study indicates that those ending CBT for health anxiety with higher ORS scores relative to their baseline score are more likely to sustain long-term improvement in a range of health domains. Conversely, those who

end treatment with lower ORS scores are unlikely to spontaneously improve further down the line. Therefore, therapists and other members of relevant patients' care-teams should be alert to the need for alternative interventions if psychotherapy offered was either not accepted or not effective. Consequently, ORS scores at the last treatment session may offer a broad, brief assessment of long-term health and quality of life after treatment. Given that the source data came from remote delivery of CBT via video conferencing or the telephone, this suggests that ROM may be a helpful predictor of outcome when using this modality.

Comparable post-treatment predictive abilities were achieved for the ORS and health anxiety assessment. As health anxiety was both the primary outcome and the main treatment target for CBT interventions, this suggests that the more general ORS can have a cross-cutting predictive effect that is similar to more specific measures.

8.4.2 Predictive validity of sudden gains

Amongst treatment responders, sudden gains on the ORS were associated with greater improvements compared with non-sudden gainers. There were significantly greater improvements in ORS-rated wellbeing across therapy sessions and significant improvements in depression and functioning across 12-month follow-up. These findings add to the extensive literature emphasising the importance of sudden gains for long-term benefits of psychological therapy (Aderka et al., 2012). This study also suggests that session-by-session ORS changes are somewhat sensitive to specific clinical changes in symptomatology – given that rapid, large, sustained ORS

improvements correspond with greater long-term improvement in clinically important domains.

The inclusion of sudden gains from the first session using the ORS is an important adaptation for brief psychological treatments such as CBT for health anxiety, as compared to longer-treatments upon which established criteria have been based. It carries the risk of including sudden gains that are more likely to have occurred due to random fluctuations in score. However, in the current sample, the adapted criteria were more clinically meaningful than criteria closer to the original. This may be because fewer sudden gainers were identified using the original criteria than has been typical in previous studies. The ORS could be more susceptible to random fluctuations in score, which would make it less likely that a statistically significant pre-post gain difference would be identified. However, the inclusion of very early and very late sudden gains seemed to be the most influential factor in identifying the majority of sudden gainers excluded by the strict criteria. The initial sessions of psychological treatment are often considered to be primarily focused on assessment only, but this study and other sudden gains research suggest that clinically significant changes can occur very early (Busch et al., 2006). Within the treatment offered in the source RCT, only ‘problem identification and prioritisation’ is intended to be covered in the first sessions, yet this contributed to significant improvements for some participants. This may be to some degree unique to this study, as most patients were unfamiliar with the concept of health anxiety at referral and so psychoeducation around this issue might have been an important therapeutic step.

No significant difference was identified between sudden gainers and the remainder of the sample in severity of health anxiety, generalised anxiety, general health, somatic symptoms, or quality of life – despite trends of greater improvement. Further research is required to establish the meaning of these results. It is possible that ORS sudden gains can only predict some specific outcomes but not others amongst patients with health anxiety. It is also possible that statistical power may have been lacking in this sample, given that there were trends towards greater improvements amongst sudden gainers on all outcomes and the original RCT sample-size was not powered for this purpose. Taken together, the current results suggest that sudden gains on the ORS may indicate improvements in some domains, but, as previously identified, supplementing the ORS with disorder-specific assessment is likely to help understand changes of particular interest (Nugter et al., 2017).

A key strength of this study is the account it takes of multiple changes over time, as – through applying an MLM design – it does not rely on a specific measurement at a particular time-point. As such, the study accounts for trends during and after psychological therapy across health domains. This approach has rarely been used in sudden gains research.

8.4.3 Limitations

Previous ROM studies have often been carried out in usual care settings. This adds two elements missing from the current study: Larger, mixed patient samples and large therapist samples. In particular, the limited variance amongst the four therapists in this study meant that the outcome variance often present in larger therapist samples was absent. However, the high rates of multimorbidity, including several mental health problems, suggest that

findings may be relevant to other patient groups – including mixed usual care populations. This study aimed to assess the value of ROM within a patient population where a variety of physical and mental health changes would be important – it did not therefore aim to be comprehensive, rather it aimed to be illustrative of the value ROM might have for outcomes amongst people with difficulties which are hard to monitor using single outcomes. The analysis aimed to account for comorbidities by included the total number of disorders into models, but this was only a significant predictor of outcome in health anxiety and generalised anxiety. This may have been because the total number of morbidities was too crude an assessment of the impact multimorbidity can have.

The criteria used to identify sudden gains have been subject to debate, with no clear agreement on the standards to be used in different clinical populations. The Tang and DeRubeis (1999b) criteria are the most commonly used and the adaptation to include very early gains aimed to fit with the general finding that very early changes can be significant. The proportion of patients identified as experiencing sudden gains in this study is broadly similar to those found in previous studies and the impact is comparable (Aderka et al., 2012). However, the null findings on the primary outcome (health anxiety) and some other measures suggests that there are important limitations to the value of ROM sudden gains identified in the way used. This is particularly important when reducing health anxiety was the primary focus and purpose of the therapy in question. Yet, sudden gains on the ORS did not translate into differential gains in health anxiety.

Common methods variance was addressed by the study design in some ways, but the repeated completion of self-report measures could still have inflated outcome associations. This means that the relationship between ORS scores and follow-up outcomes could have been over-emphasised by the fact that similar types of assessment were completed by the same person at different time points.

In addition, this study does not give evidence for the effectiveness of ROM per se. It cannot establish whether use of ROM improved the therapy offered or not. However, it does inform the clinical meaning of ROM amongst patients with health anxiety using rigorous methodology.

8.4.4 Future research and implications

Future investigations of ROM systems should include independent, blinded, multiple assessments during and for a sustained period after psychological therapy. The primary outcome should be assessed session-by-session alongside any ROM system to give a clearer understanding of their covariance over treatment. This would also inform how frequently sudden gains on ROM systems co-occur with sudden gains on the primary outcome. It would also aid understanding of the differential effect of general and disorder-specific ROM.

Overall, this study suggests that the ORS score at the final session of CBT for health anxiety can give a broad indication of long-term improvement or otherwise across a range of health domains. An assessment like this may help to address some of the confusion that can occur when working with comorbid physical and mental health problems. However, there are limits to ROM's usefulness. This study also emphasises the value of disorder-specific

feedback alongside general outcome feedback. For example, during treatment a therapist may assume improvement in health anxiety when there is a large stable improvement on the ORS, but the findings presented suggest this assumption could be incorrect.

8.4.5 Conclusion

This chapter has underlined the value of the CICS findings reported in previous chapters, emphasising the relationship between the short-term and long-term outcomes used. In particular, this chapter suggests that the predictive ability of the ORS used as a proximal outcome in the previous chapter would be very similar to that of the main treatment target and outcome: health anxiety. Therefore, the relationship between ORS scores and the CICS explored previously can be seen as a more trustworthy, general indicator of the CICS relationship to health-status in several domains. This chapter also gives weight to the value of the CICS changes shown directly before sudden gains on the ORS. Results from the current chapter suggest this type of change is likely to be clinically meaningful to important health domains. Although the study in this chapter has an independent value in its own right, within the context of this thesis it strengthen the dependability of the CICS clinical utility as presented in earlier chapters.

The following final chapter of the thesis summarises all the findings reported and discusses the relative strengths and weaknesses. The findings are placed in the context of current research, then implications for future research and clinical practice are discussed.

9. DISCUSSION

9.1 Introduction

This chapter will integrate and summarise the findings from the whole thesis. Findings will then be evaluated with respect to the relative merits and short-comings of the design used. The overall fit of the findings with existing research will be discussed. Then, having weighed up the issues arising, implications for future research are presented. Finally, a conclusion is offered, which aims to clarify what can be taken from the work as a whole.

9.2 Summary of findings

This thesis aimed to address a gap in understanding around the way patient activation operates at the level of spoken interaction. It is already evidenced that patient activation predicts a range of health outcomes and healthcare costs, but it was unclear how patient activation was presented, built or eroded within consultation discussions. As psychological therapy is largely focused on spoken interactions and patients with multimorbidity are most likely to be clinically affected by patient activation, psychological therapy sessions for patients likely to experience multimorbidity were used in an exploration of interaction-level patient activation.

Using mixed-methods research, a measure (the CICS) was developed which assesses interaction-level patient activation. Importantly, the development was grounded in rich qualitative data from direct psychological therapy interactions. This meant that the CICS was derived from several iterations of interpretation from a team of researchers, informed by quality standards in qualitative research (Elliott et al., 1999). When coupled with the

quantitative methods used, the development was also theory-informed and outcome-orientated. The development of the CICS aimed to uphold the integration of qualitative and quantitative methods often missing in mixed methods research (McKim, 2017).

The mixed methods approach meant that a tool was devised which was understandable to new users who were naïve to the entire process. As a measure of the CICS clarity, good-to-excellent IRR was achieved with four coders, including the two newly trained raters. Convergent validity was also established with therapeutic alliance as assessed by the WAI-O and therapist competence as assessed by the CTS-R. There were also initial indications of an association between CICS scores and important clinical outcomes, which warranted further study across the remainder of the thesis. However, predictive validity was strongly weighted towards specific themes of interaction; a pattern which continued through the whole thesis.

The rest of the thesis focused on the predictive validity and clinical utility of the CICS. Firstly, a comprehensive evaluation was conducted of the CICS predictive validity with long-term outcomes from CBT for health anxiety amongst high services utilisers with high rates of comorbidity. This study confirmed the initial promise of predictive validity, with general facilitative interactions and problem descriptions at the second session predicting multiple outcomes over the ensuing year. However, the most persuasive evidence was found with interactions where positive changes were noticed at the second session. Patient engagement in these interactions predicted the trajectories of all seven, diverse outcomes over one year.

In partial confirmation of findings from long-term outcomes, an investigation was conducted to explore the CICS predictive validity with session-by-session outcomes during therapy. This study confirmed some existing findings and raised useful questions in other areas. Results confirmed that reducing problem descriptions early in CBT may be beneficial to session-by-session outcome. Increased patient activation within noticing change interactions was also confirmed as beneficial to outcome. However, divergent results in other interaction types suggested that more evidence is needed before firm conclusions can be drawn. Specifically, there were discrepant results with general, facilitative interactions and evaluations of self or therapy. The stronger case is presented for evaluations of self or therapy. As these interactions take such a small proportion of total utterances, the long-term outcome study may have been underpowered to find significant associations. Nonetheless, there was a predictive relationship between positive self/therapy evaluations and greater improvement on session-by-session outcomes with trends for the same in long-term outcomes. Divergent findings also suggested that further research is required for structuring and task-focused interactions.

The study of short-term outcomes highlighted that significant changes in patient activation occur directly prior to sudden gains. This suggests that significant increases in patient activation may be partly responsible for rapid, clinically significant improvements in outcome. Again, significant reductions in problem descriptions were identified as part of the process, as was increased patient activation during information-sharing discussions. Current evidence cannot discern whether patient engagement during information giving is specifically important prior to sudden gains or if this was just the particular way

in which patient activation increased with the small sample used. There was no significant increase in noticing positive changes directly prior to a sudden gain, but this may be because important changes had not yet happened. Perhaps if the post-gain session had also been sampled an increase in changes noticed would have ensued.

Lastly, the predictive ability of the short-term ROM outcome used (the ORS) was investigated to give an indication of the long-term and transferable value of the CICS predictive abilities. This study demonstrated that the ORS predicted all long-term outcomes over the 12-month follow-up period. Furthermore, sudden gains on the ORS were identified as an important sign of greater long-term improvements in important clinical domains. In light of this evidence, findings from the CICS-outcome relationships can be taken with more confidence.

In all the assessments carried out, the identified interaction-level CICS themes outperformed the CICS global themes and standardised assessment of therapeutic alliance. This suggests that interaction-level ratings may have an added benefit in early-session analyses to global assessments.

Overall, this thesis suggests that interaction-level patient activation is an area worthy of further detailed study. In particular, the results suggest that more immediate and long-term outcomes can be predicted from early interactions in brief psychological therapy, specifically CBT for health anxiety. The methods used also help to define particular topics and types of interaction that are more or less predictive of outcome during early interactions. Two consistent findings suggest that reducing the proportion of early sessions spent describing problems could improve short and long-term outcomes, as

well as promoting sudden gains. Second, actively engaging patients in discussion about early positive changes predicts greater improvement in short and long-term outcomes. The other interaction themes showing promise for clinical utility require further research to discern whether they have consistent predictive value across situations, in specific situations or if findings were due to chance associations.

In the CICS, this thesis offers researchers a new tool with which to investigate the potential of patient activation, but in a more detailed manner than was previously possible. Should large-scale future evaluation support the current findings, the CICS may offers clinicians a means of evaluating and developing their clinical effectiveness, particularly early in therapy with anxious patients who have comorbidities.

9.3 Relationship to existing literature

Current findings corroborate evidence from MI that clinical outcomes can be predicted from motivational factors in psychological therapy interactions (Magill, Apodaca, et al., 2018). This study also supports evidence from MI skills assessments, that interactions can be reliably rated by trained coders and interaction-level patient engagement correlates with therapeutic alliance (Boardman, Catley, Grobe, Little, & Ahluwalia, 2006; Moyers et al., 2016). The CICS adds to existing literature the ability to reliably rate at each turn of speech, specify the topics of interaction which may be predictive of outcome and then quantify the combined effect of therapist efforts and patient responses to predict outcome. MI assessments have attempted to address this issue by adding ratings of therapist responsiveness to patient language, but

these are only rated on global scales not at the more sensitive interaction-level.

A key example of the link to MI theory is the significant increase in patient activation during psychoeducation directly prior to sudden gains. This finding fits closely with findings in MI: More active patient engagement during health-related information giving is more likely to lead to associated behaviour change (Butler et al., 2012; W. R. Miller & Rollnick, 2012). Following this logic, one might expect that a large increase in patient activation during therapeutic information-giving might lead to important therapeutic behaviour change. Therefore, existing MI theory supports a potential causal mechanism, between the increased engagement in information-giving and subsequent sudden gains. However, the pattern of results obtained was not entirely consistent with MI theory.

The complete absence of outcome-association with the CICS' action planning and idea generation theme is a surprise in relation to existing MI research. This theme is the closest to representing the MI concepts of CT and CCT. In the CICS, action planning positively rated interactions where patients actively expressed motivation to make therapeutic changes and negatively rated reluctance to make changes. This is a very similar formulation to CT and CCT, which have been found to predict outcome from early sessions of MI (Magill, Apodaca, et al., 2018; Magill et al., 2014) and in CBT for anxiety specifically (Aviram et al., 2016; Poulin et al., 2018). However, there may have been important, nuanced differences in the way that CT and CCT were defined in relation to the CICS action planning theme. For example, the patient-spoken sentence: "I can change, because I am very capable" would be coded as

“reason” CT in MI, but would be coded as an evaluation of self on the CICS. Overall, action-planning findings were discrepant from existing MI research. This is partly because the CICS did not aim to replicate MI-skill assessments, although CICS development was informed by them.

It is also somewhat surprising that patient engagement in problem analysis and understanding showed no association with outcome, as this interaction theme is closely related to psychological formulation or conceptualisation. Collaborative case formulation is central to CBT practice and methods (Kuyken, Padesky, & Dudley, 2009). There is much consensus that formulation is an essential element of CBT, which helps explain patients’ presentations and actively involving patients in the process enhances its value (Division of Clinical Psychology, 2011; Eells, 2007; Kuyken, Padesky, & Dudley, 2008). Problem analysis interactions appeared to evaluate this type of discussion, without successful links to outcome. It is possible that sampled sessions were too early in therapy for comprehensive case formulation to occur. An alternative explanation may lie in the paucity of evidence supporting case formulation as an influence on outcome itself (Chadwick, Williams, & Mackenzie, 2003; Kuyken, Fothergill, Musa, & Chadwick, 2005). Despite being a recommended feature of CBT, there is little evidence that case formulation, or active engagement of patients in the process, improves outcome (Kuyken, 2006). Indeed, some have argued that neither reliability nor outcome prediction are relevant goals for case formulation (Johnstone, 2018). It is therefore perhaps less surprising that problem analysis interactions did not achieve significant outcome prediction, particularly from sessions sampled very early in therapy. A competing hypothesis is that processes prior to CBT starting

minimised the impact of problem analysis discussions in early sessions. Study recruitment procedures meant that each patient had a discussion about health anxiety with their referring clinician; then had been screened for health anxiety by a researcher, and then had the results of this assessment explained. Given this background and the fact that most patients were unaware of health anxiety prior to study involvement, some of the formulation process would likely have occurred prior to the first session of CBT. Therefore, problem analysis may have played a smaller role in its effect on outcome from early session assessments.

The two most credible findings (the predictive validity of noticing early changes and early problem descriptions) are both foreshadowed in solution focused brief therapy (SFBT; De Shazer & Coulter, 2012). Within SFBT therapists are not expected to linger on problem descriptions but to move rapidly to an understanding of the patient's goals, aims and dreams. Similarly, CBT for health anxiety in people with different but chronic medical problems may not need to focus on describing complex relationships between problems, but focus on patients' life goals that have been prevented by health anxiety. Within the SFBT method, extended problem description would be seen as unnecessarily looking towards problems when solutions are the necessary requirement for improvement in the problem (De Shazer, 1988). The idea of limiting problem description is also supported by mood induction research, which suggests that reporting sad autobiographical stories induces lower mood (Jallais & Gilet, 2010). Furthermore, mood induction for negative mood states is more potent than for positive mood states (Westermann, Spies, Stahl, & Hesse, 1996). Taken together this suggests that the reduction of problem

description interactions may make room for interactions more geared towards generating solutions and possibly raised mood. It also suggests that extended periods of problem description might lower patient mood and perhaps give a less helpful experience of early therapy sessions or reduce expectations of benefit.

Discussing changes noticed very early in therapy is also a feature of SFBT. In SFBT a second session would be expected to start with the therapist explicitly asking about positive changes that may have occurred (De Shazer & Coulter, 2012). Even in the first session of SFBT a therapist would ask about occasions when the problem has been managed differently or better. This is a prominent feature, because it aims to help articulate personally meaningful solutions that may be developed during sessions, right from the very start of therapy. Describing positive changes may also act as a positive mood induction process, wherein the mutual enjoyment of describing a positive change enhances mood and expectations of positive change. In the broader context of therapeutic approaches more generally, Frank and Frank (1993) suggested that early successes, however small, were a feature of successful intervention. This again supports the idea that changes patients notice early could be a useful indicator of eventual success.

The CICS' measure of the changes noticed early in therapy could also have acted as an in-session ROM process: Rather than a standardised questionnaire completed between sessions, the changes noticed by patients give a within-session measure of the benefits they are experiencing from therapy at that point. The CICS findings also fit with ROM research, which has identified that most therapeutic improvement occurs early in therapy, often

within the first five sessions (Lambert, 2010). This suggests that noticing changes at the second session might be indicative of the type of ROM improvements expected for patients who will gain greater benefit from therapy.

Noticing change interactions could also form a useful counterbalance to resistance and other negative motivational factors identified as the best predictors in the systematic review chapter. Current evidence suggests that negative motivational interaction factors (particularly resistance) affect outcome more than positive factors (Westra, 2014). Furthermore, the influence of positive motivational factors disappears once negative factors are accounted for in predictive models (Magill, Apodaca, et al., 2018; Magill et al., 2014). Noticing change interactions could form the mirrored process to resistance: An interaction process also taking up a very small proportion of in-session utterances, but accounting for a significant proportion of outcome variance.

Evaluations of self or therapy take up even fewer in-session utterances than noticing change interactions, but were also identified as a predictor of outcomes achieved during therapy. This finding fits with theories underpinning compassion-focused therapy (Gilbert, 2005). Self-compassion is defined as “treating oneself with kindness recognising one’s shared humanity and being mindful when considering negative aspects of oneself” (Neff & Vonk, 2009; p. 23). Self-compassion of this kind has demonstrable links to better psychological health (Neff, Rude, & Kirkpatrick, 2007). Conversely, self-criticism and shame have a toxic effect on psychological health, particularly in recurrent depression (Gilbert, 2005; Segal, Williams, & Teasdale, 2013). Self-compassion and self-criticism can be seen as at opposite ends of the rating

scale used for evaluations of self. It may be that self-evaluative interactions acted as an in-session proxy for self-compassion early in CBT. It is also possible that evaluations of therapy may have acted as a proxy for patient expectations of therapeutic benefit and/or the perceived credibility of treatment. Both these constructs predict outcome from early interactions (Constantino, Coyne, Boswell, Iles, & Vîslă, 2018; Constantino, Vîslă, Coyne, & Boswell, 2018). For example, a patient expressing that a session was surprisingly helpful could be an in-session indication of positive outcome expectations, alongside improved perception of treatment credibility.

The unexpected findings related to facilitative interactions and their predictive validity with generalised anxiety and depression, may be explained by the “real relationship”. The real relationship is the relationship occurring between a therapist and patient, which is characterised by more informal, personal relations of one individual to another and separate from therapeutic processes and functions (Gelso, 2011). In a recent meta-analysis, the real relationship was found to hold a moderate association with outcome (Gelso, Kivlighan Jr, & Markin, 2018). The CICS-rated proportion of the session spent on these facilitative interactions may have been a substitute measure of the real relationship. From this perspective, more time spent in general interactions indicated a stronger real relationship. It is also noteworthy that uncertainty about technology can put healthcare patients off using digital interventions (Hardiker & Grant, 2011). Therefore, interactions which facilitated competent use of technology in CBT might have enhanced the beneficial effects produced. However, the inconsistent results mean that this interpretation must be held with caution.

Similarly cautionary findings were reported for structuring and task-focus interactions. There was some evidence of their value to short-term outcomes and in developmental work, but not long-term outcomes. These results fit with existing findings that client facilitation of “negotiating/structuring” in early sessions of CBT for depression predicted outcome (Strunk, Brotman, & DeRubeis, 2010). The current study extends these findings in suggesting that structuring and task negotiation remains important when rated at a turn-of-speech level, as well as at global, session-level ratings. However, the remaining uncertainty around long-term outcomes require further study to resolve.

In several ways CICS themes offer micro-measures of established outcome predictors usually assessed with questionnaires and only received once between sessions, the CICS allows ongoing feedback with every exchange within the session.

9.4 Strengths

All of the findings in this thesis emerged from and are grounded in the contents of specific psychological therapy interactions. This means that it is always possible to make clear references to the raw data from which inferences are drawn (as was demonstrated throughout the thesis). This differs from much psychological therapy process research, which often offers results a degree removed from the actual clinical consultation. For example, psychological therapy data are often derived from questionnaires about a patient or therapist’s view of therapy or enquiring about a construct deemed to be important to the psychological therapy process. These kind of data require a type of back-translation to inform therapists of what they should actually do

in their consultations. Even where the contents of psychological therapy are used as the source data in psychological therapy process measures (such as with the WAI-O), consultation data is translated into global scores for the construct of interest, which cannot be related to specific therapy content. By contrast, the CICS' focus on interaction-level data offers a greater degree of transparency and verifiability.

The CICS also *emerged* from observations of specific psychological therapy interactions. A measure of dependability is added by the development of a measure built on direct observation and analysis of therapy contents. This has meant that any conceptual considerations arising during the development had to be grounded in specific psychological therapy contents, rather than just relying on the researcher's ideological hypotheses. For example, no claim could be made for structuring and task focus as a theme, despite existing evidence to support it, without multiple exemplar quotes being scrutinised and agreed by thematic analysts, new coders and project supervisors. Therefore, enduring, repeated and iterative reference to actual therapy interactions kept the development and ensuing analyses grounded in the focus of the work: psychological therapy interaction patterns and their influence on outcome. In addition, the work was protected from the sway of the researcher's a priori ideas.

The iterative focus on specific interactions opens up many possibilities for future developments, which would not be possible if the content was only conceptual. The iterative focus on observed therapy interactions means coding and scoring can be adapted for different populations, session-types and therapy types in line with outcome prediction. For example, it is possible that

the facilitative and general interactions, which emerged as a predictor of multiple long-term outcomes, might only be relevant to outcomes of remotely delivered psychological therapy. This could be derived from examination of the interactions coded at this theme, which identify facilitation of remote technology use as a central component of facilitative interactions. Consequently, facilitative interactions may play less of a role in face-to-face therapy. The methodology used allows for this type of nuance to be derived and developed. It also gives opportunity for focused expansion of the interaction-level understanding described in this thesis. This might include greater understanding of the relationship between themes, the contextual factors contributing to theme presentation or explanations of why or how certain themes predict outcome.

The MLM design accounts for trends occurring across both therapy and an extensive follow-up period. This means that clinical changes happening across the course of therapy can be accounted for in the analysis. Changes occurring across the 12 months following the sessions evaluated are also included in this method of analysis. Therefore, the variance in outcome attributable to individual patient trends over time were partitioned off when assessing the predictive validity of the CICS. As a result, more reliable estimations can be made for the effects explained by the predictors of interest. The MLM method also means that models need not be dependent on a single outcome time-point (e.g. 12-month follow-up). Thus, MLM takes better account of change trends over a period of time and are a more reliable estimate of longer-term predictive patterns. Furthermore, missing data is better addressed through MLM. For example, if a single outcome point were selected, outcome

trajectories would be excluded for patients who had missing data at that point, but not others. If the 3-month follow-up time point had been used 44 participants could have been included, but only 37 if the 12-month follow-up. When combining the two 50 participants were included.

9.5 Limitations

Some characteristics of the patient samples limit the generalisability of findings. Despite the use of MLM to account for missing data across time, there was a significant amount of missing data, which means that the effects described are less precise for the sample as a whole. Although there were minimal differences between included and excluded samples, there may be clinically important differences between those with partially complete data and those with complete data. Results would be weighted towards those with complete data. The same is possible for patients who had recordings available at second sessions but not the first. For example, it may be that more confident patients were comfortable with session recordings from the very start whereas more tentative patients only built up confidence by the second session. This could mean that a more engaged sub-sample of patients are over-represented in this study. Thus the findings are less generalisable.

A predominantly female sample was used and there are important identified gender differences in interaction-style (Beebe, Beebe, Redmond, Geerinck, & Salem-Wiseman, 2015). These gender-based interaction-style differences could have meant differing results dependent on gender. For example, there were two male therapists and two female therapists providing treatment in the RCT. Interaction-style and type may have been qualitatively different where therapists and patients were of the same gender to where they

were different. Male patients may have been more willing to express vulnerability with female therapists than with male therapists, as one possibility.

Stratification of results by age might also have illuminated important nuances. Although the mean age within the RCT was lower than in previous trials of CBT for health anxiety, age ranged from 18 to 79. Again, interactions may have been qualitatively different amongst patients of different ages. For example, older patients may have been less familiar with the digital medium and so been more cautious in what was revealed. Differences in interaction-style attributable to ethnicity were not accounted for by the CICS, nor was the interaction between therapist and patient ethnicity. Giving patients choice on the ethnicity of their therapist has been shown to enhance outcome, so the effect of ethnicity may not be trivial (Soto, Smith, Griner, Domenech Rodríguez, & Bernal, 2018).

Patients in the validation studies received CBT either by telephone or videoconferencing. However, the analysis does not differentiate between the two approaches when there may be important differences in CICS observations. Furthermore, the delivery modality (telephone or video conferencing) was based on patient choice. Thus, possible differences in CICS scores between the two media could be related to patient characteristics. For example, patients who were more anxious may have opted for telephone so that a less personal interaction was required. However, no clinical differences were identified from comparison of patients who received telephone and video conferencing CBT in the main trial. Given the limited sample size, age, gender,

ethnicity and CBT delivery medium could not be added as covariates in MLMs, but would help to answer this question in future research.

Another key sampling issue is that the CICS was developed using sessions from face-to-face CBT, but predictive validity was evaluated amongst patients receiving remotely delivered CBT. The themes of interaction may not, therefore, directly transfer across different modalities in future. Furthermore, both samples had high rates of multimorbidity. Hence, the divergences in findings reported between studies may relate to differences in patients' presenting problems, as the population was clinically heterogeneous. Different clinical problems may have required different therapeutic approaches and techniques, which could in turn lead to differences in the types of interaction themes present in early sessions. In sum, these limitations restrict the general applicability of the findings reported. Replication in a larger sample would allow for clearer conclusions to be drawn.

The mixed method design used for development has limitations that could have filtered through to the later studies. Although a number of quality assurance measures were used to account for thematic analysts' pre-existing beliefs, it remains likely that unacknowledged assumptions influenced the framing and definition of CICS themes. This could explain why action planning and problem analysis findings were at odds with expectations from the literature. For example, patient activation theory was used to inform theme development, but psychological formulation theory was not. So the framing of interactions on this topic (in the problem analysis theme) may not have made the best use of existing literature. Therefore, important theoretical absences could have prevented stronger results in terms of outcome prediction. In

relation to sampling in the developmental study, only 11 of the 32 patients offered CBT in the development study agreed to having their sessions recorded. This might indicate that a more compliant sub-group of patients was represented in development, which could have affected the type of CICS themes developed. The development sample and the validation sample may also have differed in their previous experiences of mental health services. Few patients in the validation sample were seeking mental health treatment, but several of the patient from the development sample reported negative experiences of using mental health services during qualitative interviews (M. Stubley 2013, personal communication, 30 April). As a result, the development sample may have been more reluctant or guarded in their approach to CBT sessions. This again might have translated into less representative CICS themes.

In spite of MLM's benefits, the number of CICS themes tested means that multiple comparisons took place in each study. This increases the likelihood of chance associations and type 1 errors. Therefore, it is possible that some positive associations reported may be due to chance. Therefore, themes where consistent, repeated associations remained across studies and outcomes have been reported separately from less replicated result.

The absence of sufficient therapists for therapist-level effects to be accounted for in the analysis is regrettable, given that this is an important and promising area of research to enhance clinical improvement. Therapist effects are a particularly important feature amongst patients with more severe or complex problems (Barkham et al., 2017). This means that the current analysis cannot say the extent to which differences in interaction-style and type are due

to the characteristics of specific therapists. For example, some therapists in the small sample may have had a style which particularly utilises patient activation to evoke therapeutic change compared to the average CBT therapist. In this case, the role of patient activation could be over-emphasised as an outcome predictor when it might not reflect the change processes used by CBT therapists more generally. Therefore, it could be prematurely assumed that the thesis' findings are generalisable to all CBT therapists, when the therapists involved may not be typical of CBT therapists in interaction-style. Larger samples of therapists and patients in future research would be the best way to overcome this issue. Nonetheless, therapist competence was regularly assessed throughout the main RCT and similar levels of competence were reported across therapists (within the competent to proficient range). This suggests there was a degree of similarity between therapists in terms of CBT competence.

There are potential validity issues related to the construct of patient activation itself. Patient activation was developed to address healthcare needs in general, with more of a focus on physical health problems and associated healthcare consultations. Although adapting the concept of patient activation to psychological therapy interactions has benefits, it may not be wholly compatible with CBT for health anxiety - a very specific type of healthcare treatment for specific problems. By contrast to the CICS, MI-based assessment of observed motivation was developed for a specific psychological therapy with a particular set of clinical methods in mind. Therefore, the coherence of CICS ratings may be limited until clearer relationships are formed with existing psychological theory and therapeutic methods.

The generalisability of CICS results may also be limited by the atypical use of recordings within the main RCT. Patients were given session recordings to review and recordings of non-responding patients were reviewed in supervision. These procedures could have led to a degree of responsiveness that would not be typical in CBT. For example, therapists on reviewing a session may have intuitively noted unhelpful interaction-styles and adjusted them at following sessions. Similarly, an unusually high level of responsiveness might have been observed amongst patients having reviewed their own sessions and reconsidered how to make the most of future sessions.

Lastly, the thesis has emphasised the value in detailed analysis of in-session communications. However it is, in essence, focused at the level of *spoken* communication. Nothing can be said about the role of unspoken in-session communication, such as facial expression or body language. In particular, there is now a large body of evidence demonstrating that patients and therapists synchronise vocal pitch, bodily movements and even internal body states during sessions (Koole & Tschacher, 2016; Ramseyer & Tschacher, 2008, 2011; Reich, Berman, Dale, & Levitt, 2014). The current thesis is unable to account for these in-session processes or offer any understanding of how they relate to patient activation. Nevertheless, integration of the effect linguistic and physiological data may have on patient activation could be a fruitful path in future research.

9.6 Implications for future research

Current findings, though important, were developed in exploratory studies with small patient samples. Therefore, similar studies need to be replicated amongst a larger sample of patients and therapists, whilst

maintaining the same patient group ideally receiving remote CBT for health anxiety. A focused replication of this kind would give results that would be more dependable, whilst accounting for therapist effects. This type of study would allow for sample size estimates to be made a priori. In addition, important demographic and clinical covariates could be included to offer a better controlled estimate of the CICS' contribution to outcome. Subsequent research could then use similar designs in studies where treatment is offered face-to-face to further enhance the generalisability of findings.

Using the findings from these studies, research could then investigate generalisations across patient groups. Large-scale, longitudinal studies where the CICS is rated across whole episodes of therapy (rather than just the first and second sessions) would give a better understanding of interaction patterns over time. Then patterns of change in the CICS across themes and across therapy could be mapped more easily. Also a contextualised understanding of each theme's relative merits could be developed, with more nuanced guidance possible for therapists. This longitudinal understanding of patient activation in therapeutic interactions would be integral to developing and tailoring relevant interventions. For example, if patient activation is very stable across the intervention period it might suggest that either psychological intervention is not appropriate to help change patient activation or that the intervention used is not effective for addressing patient activation.

The CBT for health anxiety sessions used to validate the CICS is a brief form of CBT and very brief by comparison to some other forms of therapy. Future studies could investigate the predictive validity of The CICS in different forms of CBT and then other forms of therapy. In particular, the CICS'

relationship to MI should be clarified, both in terms of MI interaction assessment tools and MI as a therapy. All future research involving the CICS should, therefore, include other tools commonly used for rating motivational factors from in-session interactions. This would mean comparing CICS predictive validity with results from the MI Skill Code (Glynn & Moyers, 2009) and possibly the client resistance code (Chamberlain et al., 1984). Comparative analysis would highlight the relative strengths and weaknesses of these similar measures. For further elaboration of convergent and divergent properties, a comparison of predictive validity in CBT versus MI would clarify the specific value of each measure to particular treatment modalities.

The type of large-scale, longitudinal research recommended could help make quick progress with developing the use of patient activation in clinical practice. However, the lengthy coding process is a barrier to the kind of work described. Yet, recent technological innovations mean that transcribing can be automated (e.g. https://trint.com/). Detailed in-session processes can also be automatically coded using machine-learning to a reliability similar to human coders, as has been achieved in MI (Tanana, Hallgren, Imel, Atkins, & Srikumar, 2016; Xiao et al., 2016; Xiao, Imel, Georgiou, Atkins, & Narayanan, 2015). Therefore, achieving automated processes for CICS coding would be a high priority for future research. This would accelerate the processes of mapping and refining an understanding of the role CICS themes play in outcome prediction. It would mean that large datasets could be analysed quickly and passively without extended human coding. Automating coding processes would also mean that currently unknown correlates with specific themes might be identified. For example, noticing change interaction may be

associated with particular speech characteristics such as tone and prosody. Identifying sophisticated changes of this type could help therapists be more attuned to subtle feedback offered by patients in their interactions. Large, longitudinal datasets of CICS coded data could also mean that algorithm-based outcome trajectories could be passively derived (i.e. without patient or therapist needing to do anything extra like complete questionnaires). The interaction processes themselves could act as micro-outcomes, which, if reliably related to other outcomes, could be used as clinical outcomes in themselves.

Future research could also establish a standardised, qualitative meaning to CICS scores. The CICS scores used in the analysis do not yet have any external validity. This means that there is not currently any benchmarked meaning to particular scores with the exception of problem description. Problem description interactions can be reported as a percentage of the session where this type of interaction occurred. However, for themes with interactions rated on the +2 to -2 scale, total scores ranged from +200 (for a session entirely led by the patient on one interaction theme – highly unlikely to ever occur) to -200 (a session where therapist and patient are in direct opposition throughout the session using only one interaction theme – again highly unlikely to ever happen), but these boundaries are almost meaningless to the scores which are more typically achieved. Therefore, future research should aim to establish benchmarks for the meaning of different scores at differing stages of the therapeutic process. One such method would be to align CICS scores with the four levels of patient activation identified in the PAM. This would offer both external validity and meaningful integration with well-

established means of assessing patient activation. Establishing convergent validity with the PAM would also clarify to what degree the CICS and PAM agree on measurement of patient activation.

The CICS global themes received no support during the studies in this thesis. They also do not hold to the value placed on interaction-level ratings and are perhaps best viewed as brief measures of therapeutic alliance. Therefore, it might be prudent to find alternative ways of assessing integration of themes and collaborative flow using turn-by-turn data in future research. For example, integration of themes could be calculated from turn-by-turn ratings by giving a positive score if three or more interaction themes are positively coded in a single session for more than 5% of interactions. This would give a mathematical calculation of score from interaction-level ratings rather than relying on the global assessments used in this study. Overall, reconceptualisation of global themes is required in future research.

In the long-term, to evaluate the CICS clinical value (should it show promise in larger-scale replication) an RCT could compare the clinical outcomes of therapists prepared with CICS-informed training and ongoing supervision versus therapists receiving more traditional training and supervision models. Inclusion of feedback from global ratings of interaction-style (such as the WAIO) for the control group would help establish whether any additive clinical value is offered from CICS-based therapist feedback.

9.7 Implications for clinical practice

The CICS has the potential to contribute to the development of measurable, outcome-oriented, interaction-level feedback to enhance deliberate practice methods in CBT. This could support the development of

CBT therapists' expertise and through this contribute to improved clinical effectiveness (S. D. Miller, Hubble, Chow, & Seidel, 2015). Reliable turn-by-turn CICS coding provides the potential to offer greater specificity in guidance for individual practitioners: Identifying the interactional skills most important for outcome. Existing research has not been able to show what types of deliberate practice are best, perhaps, in part, because the specificity offered by a measure like the CICS has been lacking (Chow et al., 2015).

Should the CICS show further validity in future independent samples, then it could have the following clinical application: A therapist may identify a patient as at risk of poor outcome from ROM data. Using the CICS to code the previous session, patterns of interaction where patient activation is particularly low can be identified and adjusted in ensuing sessions. For example, if structuring interactions showed particularly low patient activation, more efforts could be made to facilitate greater patient choice in how sessions are structured at the following session.

The CICS could also be used to support therapists in development of clinical decision-making expertise. Patterns of patient activation could be reviewed across cases and interaction-types. Areas where there is consistently low patient activation across cases can be identified to help therapists recognise their contributing characteristic habits. For example, a therapist may observe that there are few positive interactions about noticing change amongst their patients. On reviewing example interactions, the therapist may find that they routinely ask only about what has gone wrong between sessions and do not ask about positive changes that may have occurred. Given that asking about positive inter-sessional changes is an identified clinical intervention (De

Shazer & Coulter, 2012) this could be incorporated into their clinical methods through deliberate practice and feedback.

The CICS has potential implications for selecting, assessing and training new therapists. Primarily, the CICS could mean that estimates of clinical effectiveness are possible from very few cases, using in-session micro-skills as a detailed proxy. Typically 25-30 cases should be completed before estimates of clinical effectiveness can be reliably made (Goodyear & Rousmaniere, 2017; Lambert, 2010; Prescott et al., 2017). This number could be reduced given the extent of in–session data that could be rated from comparatively few cases. It is also likely that some interactive habits will be consistent across cases that might be affecting outcomes. These could be identified very early in the training process and addresses through training, deliberate practice and supervision. The CICS could also help to identify patterns of interaction that trainee therapists might not be aware of that have a detrimental impact on outcome.

Long-term, the CICS could be useful for clinical interventions beyond CBT for health anxiety and may be relevant to other therapeutic approaches. The CICS could also provide a complementary method for assessing patient activation to global questionnaire-based assessment. By contrast to questionnaire data, the CICS could offer specific information on how clinical interaction-style could be used to improve patient activation in healthcare consultations.

9.8 Conclusion

This thesis proposes that there is a potentially unique clinical value in assessing early CBT interactions on the basis of patient activation. The studies

reported suggest that where very large amounts of early sessions are taken up with problem description, poorer short and long-term outcomes are more likely to ensue in CBT for health anxiety. Conversely, greater patient involvement in noticing very early positive changes led to greater improvement across all measured outcomes in the short and long-term. Possible predictive effects were identified in interactions where patients reported positive evaluations of themselves or therapy and where patients were more involved in structuring therapy sessions. There were also indications that greater patient activation in information giving interactions preceded sudden gains. However, these findings alongside the overall study findings require larger-scale replication within a similar population using a similar therapy. The design and results presented in this thesis show the CICS to be a promising tool with potential for clinical utility if supported by more robust future replication.

10. REFERENCES

Aderka, I., Nickerson, A., Bøe, H. J., & Hofmann, S. G. (2012). Sudden gains during psychological treatments of anxiety and depression: A meta-analysis. *Journal of consulting and clinical psychology, 80*(1), 93.

Ahmed, M., Westra, H. A., & Constantino, M. (2012). Early therapy interpersonal process differentiating clients high and low in outcome expectations. *Psychotherapy research, 22*(6), 731-745.

Allen, L., Woolfolk, R., Escobar, J., Gara, M., & Hamer, R. (2006). Cognitive-behavioral therapy for somatization disorder: a randomized controlled trial. *Arch Intern Med., 166*, 1512 - 1518.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th Ed.). Washington, DC: American Psychiatric Association.

Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of consulting and clinical psychology, 71*(5), 862.

Anderson, T., Crowley, M. E. J., Himawan, L., Holmberg, J. K., & Uhlin, B. D. (2015). Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome. *Psychotherapy Research*, 1-19.

Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of consulting and clinical psychology, 84*(1), 57.

Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*(7), 755-768.

Andrews, G., Basu, A., Cuijpers, P., Craske, M., McEvoy, P., English, C., & Newby, J. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. *Journal of Anxiety Disorders, 55*, 70-78.

Antony, M. M., Ledley, D. R., & Heimberg, R. G. (2005). *Improving outcomes and preventing relapse in cognitive-behavioural therapy.* New York, NY: Guilford Press.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3.

Arkowitz, H., Westra, H. A., Miller, W. R., & Rollnick, S. (2008). *Motivational interviewing in the treatment of psychological problems*. New York, NY: Guilford Press.

Armstrong, N., Tarrant, C., Martin, G., Manktelow, B., Brewster, L., & Chew, S. (2017). *Independent evaluation of the feasibility of using the Patient Activation Measure in the NHS in England-Final report*. London: The Health Foundation.

Aviram, A., & Westra, H. A. (2011). The impact of motivational interviewing on resistance in cognitive behavioural therapy for generalized anxiety disorder. *Psychotherapy research, 21*(6), 698-708.

Aviram, A., Westra, H. A., Constantino, M. J., & Antony, M. M. (2016). Responsive management of early resistance in cognitive-behavioral therapy for generalized anxiety disorder. *Journal of consulting and clinical psychology, 84*(9), 783-794. doi:http://dx.doi.org/10.1037/ccp0000100

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258-297). New Jersey: Wiley.

Barker, I., Steventon, A., Williamson, R., & Deeny, S. R. (2018). Self-management capability in patients with long-term conditions is associated with reduced healthcare utilisation across a whole health economy: cross-sectional analysis of electronic health records. *British Medical Journal Quality and Safety, 27*(12), 989-999.

Barkham, M., Lutz, W., Lambert, M. J., & Saxon, D. (2017). Therapist effects, effective therapists, and the law of variability. In L. Castonguay & C. Hill (Eds.), *How and why are some therapists better than others? Understanding therapist effects* (pp. 13-36). Washington, DC: American Psychological Association.

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet, 380*(9836), 37-43.

Barnfield, T. V., Mathieson, F. M., & Beaumont, G. R. (2007). Assessing the development of competence during postgraduate cognitive-behavioral therapy training. *Journal of Cognitive Psychotherapy, 21*(2), 140-147.

Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological monographs: General and applied, 76*(43), 1.

Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2014). The impact of guidance on Internet-based mental health interventions—A systematic review. *internet Interventions, 1*(4), 205-215.

Beck, A. T., & Dozois, D. J. (2011). Cognitive therapy: current status and future directions. *Annual review of medicine, 62*, 397-409.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology, 56*(6), 893.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond*. New York, NY: Guilford press.

Beebe, S. A., Beebe, S. J., Redmond, M. V., Geerinck, T., & Salem-Wiseman, L. (2015). *Interpersonal communication*. Toronto, Canada: Pearson.

Berge, L. I., Skogen, J. C., Sulo, G., Igland, J., Wilhelmsen, I., Vollset, S. E., . . . Knudsen, A. K. (2016). Health anxiety and risk of ischaemic heart disease: a prospective cohort study linking the Hordaland Health Study (HUSK) with the Cardiovascular Diseases in Norway (CVDNOR) project. *BMJ open, 6*(11), e012914.

Bermingham, S. L., Cohen, A., Hague, J., & Parsonage, M. (2010). The cost of somatisation among the working-age population in England for the year 2008–2009. *Mental Health in Family Medicine, 7*(2), 71.

Beutler, L. E., Harwood, T. M., Michelson, A., Song, X., & Holman, J. (2011). Resistance/reactance level. *Journal of Clinical Psychology, 67*(2), 133-142.

Blackburn, I., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): psychometric properties. *Behavioural and cognitive psychotherapy, 29*(04), 431-446.

Blakemore, A., Hann, M., Howells, K., Panagioti, M., Sidaway, M., Reeves, D., & Bower, P. (2016). Patient activation in older people with long-term conditions and multimorbidity: correlates and change in a cohort study in the United Kingdom. *BMC health services research, 16*(1), 582.

Boardman, T., Catley, D., Grobe, J. E., Little, T. D., & Ahluwalia, J. S. (2006). Using motivational interviewing with smokers: Do therapist behaviors relate to engagement and therapeutic alliance? *Journal of substance abuse treatment, 31*(4), 329-339.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., . . . Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior therapy, 42*(4), 676-688.

Borkovec, T., & Roemer, L. (1995). Perceived functions of worry among generalized anxiety disorder subjects: distraction from more emotionally distressing topics? *Journal of behavior therapy and experimental psychiatry, 26*(1), 25-30.

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy research, 25*(1), 6-19.

Bouwmeester, W., Zuithoff, N. P., Mallett, S., Geerlings, M. I., Vergouwe, Y., Steyerberg, E. W., . . . Moons, K. G. (2012). Reporting and methods in clinical prediction research: a systematic review. *PLoS medicine, 9*(5), e1001221.

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour research and therapy, 68*, 19-26.

Braun, S., Gregor, B., & Tran, U. S. (2013). Comparing bona fide psychotherapies of depression in adults with two meta-analytical approaches. *PloS one, 8*(6), e68135.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Handbook of Research Methods in Psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological.* (pp. 843-860). Washington, DC: American Psychological Association.

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.

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour research and therapy, 30*(1), 33-37.

Busch, A. M., Kanter, J. W., Landes, S. J., & Kohlenberg, R. J. (2006). Sudden gains and outcome: A broader temporal analysis of cognitive therapy for depression. *Behavior therapy, 37*(1), 61-68.

Butler, C., Simpson, S. A., Dunstan, F., Rollnick, S., Cohen, D., Gillespie, D., . . . Evans, J. (2012). Effectiveness of multifaceted educational programme to reduce antibiotic dispensing in primary care: practice based randomised controlled trial. *BMJ, 344*, d8173.

Button, M. L., Westra, H. A., Hara, K. M., & Aviram, A. (2015). Disentangling the impact of resistance and ambivalence on therapy outcomes in cognitive behavioural therapy for generalized anxiety disorder. *Cognitive behaviour therapy, 44*(1), 44-53.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81.

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cognitive behaviour therapy, 47*(1), 1-18.

Chadwick, P., Williams, C., & Mackenzie, J. (2003). Impact of case formulation in cognitive behaviour therapy for psychosis. *Behaviour research and therapy, 41*(6), 671-680.

Chamberlain, P., Patterson, G., Reid, J., Kavanagh, K., & Forgatch, M. (1984). Observation of client resistance. *Behavior therapy, 15*(2), 144-155.

Charlesworth, A. (2018). Running to Stand Still – Why £20.5bn is a Lot But Not Enough to Do Everything. Retrieved from https://www.health.org.uk/blogs/running-to-stand-still-%E2%80%93-why-%C2%A3205bn-is-a-lot-but-not-enough-to-do-everything

Charmaz, K. (2014). *Constructing Grounded Theory* (2 ed.). London, UK: Sage.

Chow, D., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy, 52*(3), 337.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment, 6*(4), 284.

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour research and therapy, 47*(11), 910-920.

Clark, D. M., & Oates, M. (2014). *Improving Access to Psychological Therapies: Measuring improvement and recovery Adult Services. Version 2*. London: NHS England.

Clark, D. M., Salkovskis, P. M., Hackmann, A., Wells, A., Fennell, M., Ludgate, J., . . . Gelder, M. (1998). Two psychological treatments for

hypochondriasis. A randomised controlled trial. *The British Journal of Psychiatry, 173*(3), 218-225. doi:10.1192/bjp.173.3.218

Collins, G. S., Omar, O., Shanyinde, M., & Yu, L.-M. (2013). A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *Journal of clinical epidemiology, 66*(3), 268-277.

Constantino, M., Coyne, A. E., Boswell, J. F., Iles, B. R., & Vîslă, A. (2018). A meta-analysis of the association between patients' early perception of treatment credibility and their posttreatment outcomes. *Psychotherapy, 55*(4), 486.

Constantino, M., Romano, F. M., Coyne, A. E., Westra, H. A., & Antony, M. M. (2017). Client interpersonal impacts as mediators of long-term outcome in cognitive-behavioral therapy integrated with motivational interviewing for generalized anxiety disorder. *Psychotherapy research, 28*(6), 861-872.

Constantino, M., Vîslă, A., Coyne, A. E., & Boswell, J. F. (2018). A meta-analysis of the association between patients' early treatment outcome expectation and their posttreatment outcomes. *Psychotherapy, 55*(4), 473.

Coolican, H. (2019). *Research methods and statistics in psychology* (7th ed.). New York, NY: Routledge.

Cooper, K., Gregory, J. D., Walker, I., Lambe, S., & Salkovskis, P. M. (2017). Cognitive behaviour therapy for health anxiety: a systematic review and meta-analysis. *Behavioural and cognitive psychotherapy, 45*(2), 110-123.

Coyne, A. E., Constantino, M., Westra, H. A., & Antony, M. M. (2018). Competing Indirect Effects in a Comparative Psychotherapy Trial for Generalized Anxiety Disorder. *Psychotherapy.* doi:http://dx.doi.org/10.1037/pst0000163

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: a technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science, 16*(1), 42-51.

Crits-Christoph, P., Connolly Gibbons, M., & Mukherjee, D. (2013). Psychotherapy Process-Outcome Research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6 ed., pp. 298-340). Somerset: Wiley.

Crits-Christoph, P., & Gallop, R. (2006). Therapist effects in the national institute of mental health treatment of depression collaborative research program and other psychotherapy studies. *Psychotherapy research, 16*(02), 178-181.

Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clinical psychology review, 34*(2), 130-140.

Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2013). The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World psychiatry, 12*(2), 137-148.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative

outcome studies. *Journal of consulting and clinical psychology, 76*(6), 909.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge, UK: Cambridge University Press.

De Nadai, A. S., & Karver, M. S. (2013). Limited motivation, patient-therapist mismatch, and the therapeutic alliance. In E. A. Storch & D. McKay (Eds.), *Handbook of Treating Variants and Complications in Anxiety Disorders*. New York, NY: Springer.

De Ridder, D., Geenen, R., Kuijer, R., & van Middendorp, H. (2008). Psychological adjustment to chronic disease. *The Lancet, 372*(9634), 246-255.

De Shazer, S. (1988). *Clues: Investigating solutions in brief therapy.* New York, NY: Norton & Co.

De Shazer, S., & Coulter, M. (2012). *More than miracles: The state of the art of solution-focused brief therapy*. New York, NY: Routledge.

Debray, T. P., Damen, J. A., Snell, K. I., Ensor, J., Hooft, L., Reitsma, J. B., . . . Moons, K. G. (2017). A guide to systematic review and meta-analysis of prediction model performance. *BMJ, 356*, i6460.

Deen, D., Lu, W.-H., Rothstein, D., Santana, L., & Gold, M. R. (2011). Asking questions: the effect of a brief intervention in community health centers on patient activation. *Patient education and counseling, 84*(2), 257-260.

Deeny, S., Thorlby, R., & Steventon, A. (2018). Reducing emergency admissions: Unlocking the potential of people to better manage their long-term conditions. Retrieved from https://www.health.org.uk/sites/default/files/Reducing-Emergency-Admissions-long-term-conditions-briefing.pdf

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., . . . Nevin, J. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry, 5*(7), 564-572.

Derksen, F., Bensing, J., & Lagro-Janssen, A. (2013). Effectiveness of empathy in general practice: a systematic review. *British Journal of General Practice, 63*(606), e76-e84.

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2006). A Conceptual and Methodological Analysis of the Nonspecifics Argument. *Clinical Psychology: Science and Practice, 12*(2), 174-183. doi:10.1093/clipsy.bpi022

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., . . . Brown, L. L. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of general psychiatry, 62*(4), 409-416.

Dickens, G., & Picchioni, M. (2012). A systematic review of the terms used to refer to people who use mental health services: User perspectives. *International Journal of Social Psychiatry, 58*(2), 115-122.

Diefenbach, T. (2009). Are case studies more than sophisticated storytelling?: Methodological problems of qualitative empirical research mainly based on semi-structured interviews. *Quality & Quantity, 43*(6), 875.

Division of Clinical Psychology. (2011). *Good practice guidelines on the use of psychological formulation*. Leicester, England: British Psychological Society.

Dozois, D. J. A., Westra, H. A., Collins, K. A., Fung, T. S., & Garry, J. K. (2004). Stages of change in anxiety: Psychometric properties of the University of Rhode Island Change Assessment (URICA) scale. *Behaviour research and therapy, 42*(6), 711-729.

Duarte, A., Walker, J., Walker, S., Richardson, G., Hansen, C. H., Martin, P., . . . Sharpe, M. (2015). Cost-effectiveness of integrated collaborative care for comorbid major depression in patients with cancer. *Journal of Psychosomatic research, 79*(6), 465-470.

Duncan, B., & Reese, R. J. (2015). The Partners for Change Outcome Management System (PCOMS) revisiting the client's frame of reference. *Psychotherapy, 52*(4), 391.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment implications for health, education, and the workplace. *Psychological science in the public interest, 5*(3), 69-106.

Eastin, M. S., & Guinsler, N. M. (2006). Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. *CyberPsychology & Behavior, 9*(4), 494-498.

Eells, T. D. (2007). *Handbook of Psychotherapy Case Formulation* (2nd ed.). New York, NY: Guilford Press.

Elliott, R. (1989). Comprehensive Process Analysis: Understanding the change process in significant therapy events. In M. Packer & R. Addison (Eds.), *Entering the circle: Hermeneutic investigations in psychology* (pp. 165-184): State University of New York, NY.

Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British journal of clinical psychology, 38*(3), 215-229.

Ericsson, K. A. (2009). Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments. In K. A. Ericsson (Ed.), *Enhancing the development of professional performance: Implications from the study of deliberate practice*. New York, NY: Cambridge University Press.

EuroQol Group. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy, 16*, 199-208.

Fairclough, N., Jessop, B., & Sayer, A. (2002). Critical realism and semiosis. *Alethia, 5*(1), 2-10.

Falkenström, F., Ekeblad, A., & Holmqvist, R. (2016). Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *Journal of consulting and clinical psychology, 84*(8), 738.

Falkenström, F., Hatcher, R. L., Skjulsvik, T., Larsson, M. H., & Holmqvist, R. (2015). Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. *Psychological assessment, 27*(1), 169.

Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of consulting and clinical psychology, 67*(4), 578.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). London, UK: Sage.

Fineberg, N. A., Haddad, P. M., Carpenter, L., Gannon, B., Sharpe, R., Young, A. H., . . . Nutt, D. J. (2013). The size, burden and cost of disorders of the brain in the UK. *Journal of Psychopharmacology, 27*(9), 761-770.

Fink, P., Ørnbøl, E., & Christensen, K. S. (2010). The outcome of health anxiety in primary care. A two-year follow-up study on health care costs and self-rated health. *PloS one, 5*(3), e9873.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1998). *Structured Clinical Interview for DSM-IV Axis I Disorders: Patient Edition (February 1996 Final), SCID-I/P*: Biometrics Research Department, New York State Psychiatric Institute.

Flückiger, C., Del Re, A., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy relationships that work, 3*.

Frank, J. D., & Frank, J. B. (1993). *Persuasion and healing: A comparative study of psychotherapy* (3rd ed.). London: John Hopkins University Press.

Garcia, J., & Gustavson, A. (1997). The Science of Self-Report. Retrieved from https://www.psychologicalscience.org/observer/the-science-of-self-report

Gask, L., Dowrick, C., Salmon, P., Peters, S., & Morriss, R. (2011). Reattribution reconsidered: narrative review and reflections on an educational intervention for medically unexplained symptoms in primary care settings. *Journal of Psychosomatic research, 71*(5), 325-334.

Gelso, C. (2011). *The real relationship in psychotherapy: The hidden foundation of change.* Washington, DC: American Psychological Association Press.

Gelso, C., Kivlighan Jr, D. M., & Markin, R. D. (2018). The real relationship and its role in psychotherapy outcome: A meta-analysis. *Psychotherapy, 55*(4), 434.

Georgaca, E. (2001). Voices of the self in psychotherapy: A qualitative analysis. *British Journal of Medical Psychology, 74*(2), 223-236.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism, 10*(2), 486.

Gilbert, P. (2005). Compassion and cruelty: A biopsychosocial approach. In P. Gilbert (Ed.), *Compassion: Conceptualisations, research and use in psychotherapy* (pp. 9-74). Hove, UK: Routledge.

Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour research and therapy, 50*(12), 747-752.

Glover, G., Webb, M., & Evison, E. (2010). *Improving Access to Psychological Therapies: A Review of the Progress Made by Sites in the First Roll-Out Year.* Durham, UK: North East Public Health Observatory.

Glynn, L. H., & Moyers, T. (2009). Manual for the motivational interviewing skill code (MISC), version 1.1: Addendum to MISC 1.0. *Unpublished manuscript. Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico. Retrieved March 24, 2019 from Center on*

Alcoholism, Substance Abuse, and Addictions Web site: http://casaa.unm.edu/download/MISC_1.1_Manual.pdf.

Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W. T., Whipple, J. L., . . . Wampold, B. E. (2016). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy, 53*(3), 367.

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of counseling psychology, 63*(1), 1.

Goldsmith, L. P., Dunn, G., Bentall, R. P., Lewis, S. W., & Wearden, A. J. (2015). Therapist effects and the impact of early therapeutic alliance on symptomatic outcome in chronic fatigue syndrome. *PloS one, 10*(12), e0144623.

Goodyear, R., & Rousmaniere, T. (2017). Helping therapists to each day become a little better than they were the day before: The expertise-development model of supervision and consultation. In T. Rousmaniere, R. Goodyear, S. D. Miller, & B. E. Wampold (Eds.), *The cycle of excellence: Using deliberate practice to improve supervision and training.* Chichester, UK: John Wiley and Sons Ltd.

Grant, N., Hotopf, M., Breen, G., Cleare, A., Grey, N., Hepgul, N., . . . Wingrove, J. (2014). Predicting outcome following psychological therapy in IAPT (PROMPT): a naturalistic project protocol. *BMC psychiatry, 14*(1), 170.

Greene, J., & Hibbard, J. (2012). Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. *Journal of general internal medicine, 27*(5), 520-526.

Greene, J., Hibbard, J., Sacks, R., Overton, V., & Parrotta, C. D. (2015). When patient activation levels change, health outcomes and costs change, too. *Health Affairs, 34*(3), 431-437.

Greeven, A., van Balkom, A., Visser, S., Merkelbach, J., van Rood, Y., van Dyck, R., . . . Spinhoven, P. (2007). Cognitive behaviour therapy and paroxetine in the treatment of hypochondriasis: a randomized controlled trial. *Am J Psychiatry, 164*, 91 - 99.

Grimes, D. A., Hubacher, D., Nanda, K., Schulz, K. F., Moher, D., & Altman, D. G. (2005). The Good Clinical Practice guideline: a bronze standard for clinical research. *The Lancet, 366*(9480), 172-174.

Guthrie, B., Payne, K., Alderson, P., McMurdo, M. E., & Mercer, S. W. (2012). Adapting clinical guidelines to take account of multimorbidity. *BMJ, 345*, e6341. doi:10.1136/bmj.e6341

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology, 8*(1), 23.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155-163.

Hansen, B. P., Lambert, M. J., & Vlass, E. N. (2015). Sudden Gains and Sudden Losses in the Clients of a" Supershrink": 10 Case Studies. *Pragmatic Case Studies in Psychotherapy, 11*(3), 154-201.

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329-343.

Hara, K. M., Westra, H. A., Aviram, A., Button, M. L., Constantino, M., & Antony, M. M. (2015). Therapist awareness of client resistance in cognitive-behavioral therapy for generalized anxiety disorder. *Cognitive behaviour therapy, 44*(2), 162-174. doi:http://dx.doi.org/10.1080/16506073.2014.998705

Hardiker, N. R., & Grant, M. J. (2011). Factors that influence public engagement with eHealth: A literature review. *International journal of medical informatics, 80*(1), 1-12.

Harré, R. S., P. . (1972). *The explanation of social behaviour.* Oxford, UK: Blackwell.

Harrell, F. (2015). *Regression Modeling Strategies* (2nd ed.). London, UK: Springer.

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy research, 16*(1), 12-25.

Hayden, J. A., Côté, P., & Bombardier, C. (2006). Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine, 144*(6), 427-437.

Hedman, E., Axelsson, E., Görling, A., Ritzman, C., Ronnheden, M., El Alaoui, S., . . . Ljótsson, B. (2014). Internet-delivered exposure-based cognitive–behavioural therapy and behavioural stress management for severe health anxiety: randomised controlled trial. *The British Journal of Psychiatry, 205*(4), 307-314.

Hedman, E., Lekander, M., Ljotsson, B., Lindefors, N., Ruck, C., Hofmann, S. G., . . . Schulz, S. M. (2014). Sudden gains in internet-based cognitive behaviour therapy for severe health anxiety. *Behaviour Research & Therapy, 54*, 22-29. doi:10.1016/j.brat.2013.12.007

Heinonen, E., Lindfors, O., Laaksonen, M. A., & Knekt, P. (2012). Therapists' professional and personal characteristics as predictors of outcome in short- and long-term psychotherapy. *Journal of Affective Disorders, 138*(3), 301-312. doi:10.1016/j.jad.2012.01.023

Henningsen, P., Zimmermann, T., & Sattel, H. (2003). Medically unexplained physical symptoms, anxiety, and depression: a meta-analytic review. *Psychosomatic Medicine, 65*(4), 528-533.

Heppner, P., Wampold, B. E., & Kivlighan, D. (2008). *Research Design in Counseling* (3rd ed.). Belmont, CA: Thomson Brooks/Cole.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., . . . Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research, 20*(10), 1727-1736.

Hibbard, J., & Gilburt, H. (2014). *Supporting people to manage their health: An introduction to patient activation.* London, UK: The King's Fund.

Hibbard, J., & Greene, J. (2013). What The Evidence Shows About Patient Activation: Better Health Outcomes And Care Experiences; Fewer Data On Costs. *Health Affairs, 32*(2), 207-214. doi:10.1377/hlthaff.2012.1061

Hibbard, J., Mahoney, E. R., Stock, R., & Tusler, M. (2007). Do increases in patient activation result in improved self-management behaviors?

Health Services Research, 42(4), 1443-1463. doi:10.1111/j.1475-6773.2006.00669.x

Hibbard, J., Mahoney, E. R., Stockard, J., & Tusler, M. (2005). Development and testing of a short form of the patient activation measure. *Health Services Research, 40*(6p1), 1918-1930.

Hibbard, J., Stockard, J., Mahoney, E. R., & Tusler, M. (2004). Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers. *Health Services Research, 39*(4 Pt 1), 1005-1026. doi:10.1111/j.1475-6773.2004.00269.x

Hippisley-Cox, J., & Vinogradova, Y. (2009). *Trends in Consultation Rates in General Practice 1995 to 2008: Analysis of the QResearch® database. Final report to the NHS information centre and department of health.* Leeds: The Health and Social Care Information Centre.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician, 55*(1), 19-24.

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognitive Therapy and Research, 36*(5), 427-440.

Hollis, C., Morriss, R., Martin, J., Amani, S., Cotton, R., Denis, M., & Lewis, S. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry, 206*(4), 263-265.

Hollon, S. D., & Beck, A. T. (2013). Cognitive and cognitive behavioral therapies. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (6th ed., pp. 393-442). Somerset, NJ: Wiley.

Horvath, A. O., Del Re, A., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of counseling psychology, 36*(2), 223.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Hunot, V., Churchill, R., Teixeira, V., & de Lima, M. S. (2007). Psychological therapies for generalised anxiety disorder. *Cochrane Database of Systematic Reviews*(1).

Ilardi, S. S., & Craighead, W. E. (1994). The role of nonspecific factors in cognitive-behavior therapy for depression. *Clinical Psychology: Science and Practice, 1*(2), 138-155.

Imel, Z. E., Baer, J. S., Martino, S., Ball, S. A., & Carroll, K. M. (2011). Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence, 115*(3), 229-236.

Jacobson, N. S., & Hollon, S. D. (1996). Prospects for future comparisons between drugs and psychotherapy: Lessons from the CBT-versus-pharmacotherapy exchange.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology, 59*(1), 12-19.

Jallais, C., & Gilet, A.-L. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior research methods, 42*(1), 318-325.

Janssen, M. F., Birnie, E., Haagsma, J. A., & Bonsel, G. J. (2008). Comparing the standard EQ–5D three level system with a five level version. *Value Health, 11*, 275-284.

Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., . . . Busschbach, J. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of life research, 22*(7), 1717-1727.

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2018). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical psychology review*.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Johnstone, L. (2018). Psychological formulation as an alternative to psychiatric diagnosis. *Journal of Humanistic Psychology, 58*(1), 30-46.

Jones, R. (1995). Why do qualitative research? In: British Medical Journal Publishing Group.

Kahneman, D. (2011). *Thinking, fast and slow*. London, UK: Penguin.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist, 64*(6), 515.

Kampman, M., Keijsers, G. P., Hoogduin, C. A., & Hendriks, G.-J. (2008). Outcome prediction of cognitive behaviour therapy for panic disorder: Initial symptom severity is predictive for treatment outcome, comorbid anxiety or depressive disorder, cluster C personality disorders and initial motivation are not. *Behavioural and cognitive psychotherapy, 36*(1), 99-112.

Kanesarajah, J., Waller, M., Whitty, J. A., & Mishra, G. D. (2018). Multimorbidity and quality of life at mid-life: A systematic review of general population studies. *Maturitas, 109*, 53-62.

Kasteridis, P., Street, A., Dolman, M., Gallier, L., Hudson, K., Martin, J., & Wyer, I. (2015). Who would most benefit from improved integrated care? Implementing an analytical strategy in South Somerset. *International journal of integrated care, 15*(1).

Kazantzis, N., Deane, F. P., & Ronan, K. R. (2005). Assessment of homework completion. In N. Kazantzis, F. P. Deane, & K. R. Ronan (Eds.), *Using homework assignments in cognitive behaviour therapy* (pp. 61-72). New York, NY: Routledge.

Kendrick, T., Moore, M., Gilbody, S., Churchill, R., Stuart, B., & El-Gohary, M. (2014). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*(7).

Kiesler, D. J., & Schmidt, J. A. (1993). *The Impact Message Inventory: Form IIA Octant Scale version.* Palo Alto, CA: Mind Garden.

Kirmayer, L. J., & Looper, K. J. (2006). Abnormal illness behaviour: physiological, psychological and social dimensions of coping with distress. *Current Opinion in Psychiatry, 19*(1), 54-60.

Koole, S. L., & Tschacher, W. (2016). Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance. *Frontiers in psychology, 7*, 862.

Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London, UK: Sage.

Kroenke, K., Spitzer, R., & Williams, J. (2001). The PHQ-9: the validity of a brief depression severity measure. *Journal of Internal Medicine, 16*, 606-613.

Kroenke, K., Spitzer, R., & Williams, J. (2002). The PHQ-15: the validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine, 64*, 258-260.

Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of consulting and clinical psychology, 64*(3), 532.

Kuyken, W. (2006). Evidence-based case formulation: is the emperor clothed? In N. Tarrier (Ed.), *Case Formulation in Cognitive Behaviour Therapy*. Hove, UK: Brunner-Routledge.

Kuyken, W., Fothergill, C. D., Musa, M., & Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behaviour research and therapy, 43*(9), 1187-1201.

Kuyken, W., Padesky, C., & Dudley, R. (2009). *Collaborative case conceptualizaion: Working effectively with clients in cognitive-behavioral therapy*. New York, NY: The Guilford Press.

Kuyken, W., Padesky, C. A., & Dudley, R. (2008). The science and practice of case conceptualization. *Behavioural and cognitive psychotherapy, 36*(6), 757-768.

Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring and feedback in clinical practice*. Washington, DC: American Psychological Association.

Lambert, M. J. (2012). Helping clinicians to use and learn from research-based systems: The OQ-analyst. *Psychotherapy, 49*(2), 109.

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (6th ed., pp. 169-217). Somerset, NJ: Wiley.

Lambert, M. J. (2015). Outcome research: Methods for improving outcome in routine care. In O. Gelo, A. Pritz, & B. Rieken (Eds.), *Psychotherapy Research: Foudations, Process and Outcome* (pp. 593-610). London, UK: Springer.

Lambert, M. J. (2017). Maximizing psychotherapy outcome beyond evidence-based medicine. *Psychotherapy and psychosomatics, 86*(2), 80-89.

Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520.

Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (in-press). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. In J. C. Norcross & L. M. J (Eds.), *Psychotherapy*

relationships that work (3rd ed.). New York, NY: Oxford University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Levitt, H. M., Pomerville, A., & Surace, F. I. (2016). A qualitative meta-analysis examining clients' experiences of psychotherapy: A new agenda. *Psychological bulletin, 142*(8), 801.

Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological bulletin, 76*(5), 365.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on psychological science, 2*(1), 53-70.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *American Psychologist, 48*(12), 1181.

Lombardi, D. R., Button, M. L., & Westra, H. A. (2014). Measuring motivation: Change talk and counter-change talk in cognitive behavioral therapy for generalized anxiety. *Cognitive behaviour therapy, 43*(1), 12-21. doi:http://dx.doi.org/10.1080/16506073.2013.846400

Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It's complicated: the relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical psychology review, 41*, 3-15.

Luborsky, L., Crits-Christoph, P., Mintz, J., & Auerbach, A. (1988). *Who will benefit from psychotherapy?: Predicting therapeutic outcomes*: Basic Books.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., . . . Schweizer, E. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*(1), 95-106.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., . . . Krause, E. D. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice, 9*(1), 2-12.

Lucock, M., Morley, S., White, C., & Peake, M. (1997). Responses of consecutive patients to reassurance after gastroscopy: results of self administered questionnaire survey. *British Medical Journal, 315*(7108), 572-575.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92.

Magill, M., Apodaca, T. R., Borsari, B., Gaume, J., Hoadley, A., Gordon, R. E., . . . Moyers, T. (2018). A meta-analysis of motivational interviewing process: Technical, relational, and conditional process models of change. *Journal of consulting and clinical psychology, 86*(2), 140.

Magill, M., Bernstein, M. H., Hoadley, A., Borsari, B., Apodaca, T. R., Gaume, J., & Tonigan, J. S. (2018). Do what you say and say what you are going to do: A preliminary meta-analysis of client change and sustain talk subtypes in motivational interviewing. *Psychotherapy research*, 1-10.

Magill, M., Gaume, J., Apodaca, T. R., Walthers, J., Mastroleo, N. R., Borsari, B., & Longabaugh, R. (2014). The technical hypothesis of motivational

interviewing: A meta-analysis of MI's key causal model. *Journal of consulting and clinical psychology, 82*(6), 973.

Malins, S., Kai, J., Atha, C., Avery, A., Guo, B., James, M., . . . Morriss, R. (2016). Cognitive behaviour therapy for long-term frequent attenders in primary care: a feasibility case series and treatment development study. *British Journal of General Practice, 66*(651), e729-736.

Malins, S., Moghaddam, N., Morriss, R., Schroder, T., Cope, N., & Brown, P. (2018). Consultation Interaction Coding Scheme 1.6. Retrieved from https://figshare.com/articles/Consultation_Interaction_Coding_Scheme_CICS_/7302386

Mallett, S., Royston, P., Dutton, S., Waters, R., & Altman, D. G. (2010). Reporting methods in studies developing prognostic models in cancer: a review. *BMC medicine, 8*(1), 20.

McAdams, D. (1993). *The stories we live by: Personal myths and the making of the self.* New York, NY: Morrow.

McAleavey, A. A., Castonguay, L. G., & Goldfried, M. R. (2014). Clinical experiences in conducting cognitive-behavioral therapy for social phobia. *Behavior therapy, 45*(1), 21-35.

McHugh, R. K., Murray, H. W., Hearon, B. A., Pratt, E. M., Pollack, M. H., Safren, S. A., & Otto, M. W. (2013). Predictors of dropout from psychosocial treatment in opioid-dependent outpatients. *The American journal on addictions, 22*(1), 18-22.

McKim, C. A. (2017). The value of mixed methods research: A mixed methods study. *Journal of Mixed Methods Research, 11*(2), 202-222.

McLaughlin, H. (2009). What's in a name: 'client', 'patient', 'customer', 'consumer', 'expert by experience', 'service user' - what's next? *British Journal of Social Work, 39*(6), 1101-1117.

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*(2), 295-314.

Mercer, S. W., Fitzpatrick, B., Guthrie, B., Fenwick, E., Grieve, E., Lawson, K., . . . O'Brien, R. (2016). The CARE Plus study–a whole-system intervention to improve quality of life of primary care patients with multimorbidity in areas of high socioeconomic deprivation: exploratory cluster randomised controlled trial and cost-utility analysis. *BMC medicine, 14*(1), 88.

Mercer, S. W., Gunn, J., Bower, P., Wyke, S., & Guthrie, B. (2012). Managing patients with mental and physical multimorbidity. *British Medical Journal, 345*, e5559.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour research and therapy, 28*(6), 487-495.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., . . . Imbens, G. (2014). Promoting transparency in social science research. *Science, 343*(6166), 30-31.

Miller, S. D., Duncan, B., Brown, J., Sparks, J., & Claud, D. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and

feasibility of a brief visual analog measure. *Journal of brief Therapy, 2*(2), 91-100.

Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology, 61*(2), 199-208.

Miller, S. D., Hubble, M. A., Chow, D., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy, 52*(4), 449.

Miller, W. R., & Johnson, W. R. (2008). A natural language screening measure for motivation to change. *Addictive Behaviors, 33*(9), 1177-1182.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. New York, NY: Guilford press.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist, 64*(6), 527.

Milne, D. L., Baker, C., Blackburn, I.-M., James, I., & Reichelt, K. (1999). Effectiveness of cognitive therapy training. *Journal of behavior therapy and experimental psychiatry, 30*(2), 81-92.

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). doi:doi:10.1371/journal.pmed1000097

Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., . . . Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS medicine, 11*(10), e1001744.

Morriss, R., Kai, J., Atha, C., Avery, A., Bayes, S., Franklin, M., . . . Yang, M. (2012). Persistent frequent attenders in primary care: costs, reasons for attendance, organisation of care and potential for cognitive behavioural therapeutic intervention. *BMC Fam Pract, 13*, 39. doi:10.1186/1471-2296-13-39

Morriss, R., Patel, S., Malins, S., Guo, B., Higton, F., James, M., . . . Kaylor-Hughes, C. (2019). Clinical and economic outcomes of remotely delivered cognitive behaviour therapy versus treatment as usual for repeat unscheduled care users with severe health anxiety: a multicentre randomised controlled trial. *BMC medicine, 17*(1), 16.

Moyers, T., Martin, T., Catley, D., Harris, K., & Ahluwalia, J. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and cognitive psychotherapy, 31*(02), 177-184.

Moyers, T., Martin, T., Houck, J., Christopher, P., & Tonigan, J. (2009). From in-session behaviors to drinking outcomes: a causal chain for motivational interviewing. *Journal of consulting and clinical psychology, 77*(6), 1113.

Moyers, T., Martin, T., Manuel, J., Hendrickson, S., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of substance abuse treatment, 28*(1), 19-26.

Moyers, T., Miller, W. R., & Hendrickson, S. (2005). How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. *Journal of consulting and clinical psychology, 73*(4), 590.

Moyers, T., Rowell, L., Manuel, J., Ernst, D., & Houck, J. (2016). The motivational interviewing treatment integrity code (MITI 4): rationale, preliminary reliability and validity. *Journal of substance abuse treatment, 65*, 36-42.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*(5), 461-464.

Muran, J. C., Safran, J. D., Gorman, B. S., Samstag, L. W., Eubanks-Carter, C., & Winston, A. (2009). The relationship of early alliance ruptures and their resolution to process and outcome in three time-limited psychotherapies for personality disorders. *Psychotherapy: Theory, Research, Practice, Training, 46*(2), 233.

Naylor, C., Das, P., Ross, S., Honeyman, M., Thompson, J., & Gilburt, H. (2016). *Bringing together physical and mental health*. London, UK: King's Fund.

Naylor, C., Parsonage, M., McDaid, D., Knapp, M., Fossey, M., & Galea, S. (2012). *Long-term conditions and mental health: The cost of co-morbidities*. London, UK: The King's Fund.

Neff, K. D., Rude, S. S., & Kirkpatrick, K. L. (2007). An examination of self-compassion in relation to positive psychological functioning and personality traits. *Journal of research in personality, 41*(4), 908-916.

Neff, K. D., & Vonk, R. (2009). Self-compassion versus global self-esteem: Two different ways of relating to oneself. *Journal of personality, 77*(1), 23-50.

Newman, C. F. (1994). Understanding client resistance: Methods for enhancing motivation to change. *Cognitive and Behavioral Practice, 1*(1), 47-69.

Newman, M. G., Szkodny, L. E., Llera, S. J., & Przeworski, A. (2011). A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? *Clinical psychology review, 31*(1), 89-103.

NHS Digital. (2018). Psychological Therapies, Annual report on the use of IAPT services - England, 2017-18. Retrieved from https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2017---18

NHS England. (2018). Patient activation: People's ability to manage their own health and wellbeing.

NHS England. (2019). *The NHS Long Term Plan*. London, UK: NHS England.

Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy, 55*(4), 303.

Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: research conclusions and clinical practices. *Psychotherapy, 48*(1), 98.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Christensen, G. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of mathematical psychology, 3*(1), 1-18.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1609406917733847.

Nugter, M. A., Hermens, M. L., Robbers, S., Van Son, G., Theunissen, J., & Engelsbel, F. (2017). Use of outcome measurements in clinical practice: How specific should one be? *Psychotherapy research*, 1-13.

Okiishi, J., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62*(9), 1157-1172.

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy, 10*(6), 361-373.

Olatunji, B. O., Kauffman, B. Y., Meltzer, S., Davis, M. L., Smits, J. A., & Powers, M. B. (2014). Cognitive-behavioral therapy for hypochondriasis/health anxiety: a meta-analysis of treatment outcome and moderators. *Behaviour research and therapy, 58*, 65-74.

olde Hartman, T. C., Borghuis, M. S., Lucassen, P. L., van de Laar, F. A., Speckens, A. E., & van Weel, C. (2009). Medically unexplained symptoms, somatisation disorder and hypochondriasis: course and prognosis. A systematic review. *Journal of Psychosomatic research, 66*(5), 363-377.

Orlinsky, D. E., & Howard, K. I. (1987). A generic model of psychotherapy. *Journal of Integrative & Eclectic Psychotherapy*.

Østergård, O. K., Randa, H., & Hougaard, E. (2018). The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy—A systematic review and meta-analysis. *Psychotherapy research*, 1-18.

Paivio, S., & Bahr, L. (1998). Interpersonal problems, working alliance, and outcome in short-term experiential therapy. *Psychotherapy research, 8*(4), 392-407.

Parker, I. (2005). *Qualitative Psychology: Introducing radical research.* Maidenhead, UK: Open University Press.

Patel, S., Kai, J., Atha, C., Avery, A., Guo, B., & Malins, S. (2015). Clinical characteristics of persistent frequent attenders in primary care: case-control study. . *BMC family practice, 32*(6), 624-630.

Patel, S., Malins, S., Guo, B., James, M., Kai, J., Kaylor-Hughes, C., . . . Morriss, R. (2016). Protocol investigating the clinical outcomes and cost-effectiveness of cognitive–behavioural therapy delivered remotely for unscheduled care users with health anxiety: randomised controlled trial. *British Journal of Psychiatry Open, 2*(1), 81-87.

Patton, M. (1980). *Qualitative evaluation methods.* . London, UK: Sage.

Pelletier, L. G., Tuson, K. M., & Haddad, N. K. (1997). Client motivation for therapy scale: A measure of intrinsic motivation, extrinsic motivation, and amotivation for therapy. *Journal of personality assessment, 68*(2), 414-435.

Pinto, A., Pinto, A. M., Neziroglu, F., & Yaryura-Tobias, J. A. (2007). Motivation to change as a predictor of treatment response in obsessive compulsive disorder. *Annals of Clinical Psychiatry, 19*(2), 83-87.

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS* (6 ed.). New York, NY: Routledge.

Poulin, L. E., Button, M. L., Westra, H. A., Constantino, M., & Antony, M. M. (2018). The predictive capacity of self-reported motivation vs. early observed motivational language in cognitive behavioural therapy for generalized anxiety disorder. *Cognitive behaviour therapy*, 1-16.

Prescott, D., Maeschalck, C., & Miller, S. D. (2017). *Feedback informed treatment in clinical practice: Reaching for excellence.* Washington, DC: American Psychological Association.

Primakoff, L., Epstein, N., & Covi, L. (1989). *Homework compliance: An uncontrolled variable in cognitive therapy outcome research.* New York, NY: Springer.

Ramseyer, F., & Tschacher, W. (2008). Synchrony in dyadic psychotherapy sessions. In *Simultaneity: Temporal structures and observer perspectives* (pp. 329-347): World Scientific.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of consulting and clinical psychology, 79*(3), 284.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: : Sage Publications.

Reich, C. M., Berman, J. S., Dale, R., & Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *Journal of Social and Clinical Psychology, 33*(5), 481-494.

Reicher, S. (2000). Against methodolatry: some comments on Elliott, Fischer, and Rennie. *British journal of clinical psychology, 39*(1), 1-6.

Rennie, D. L. (1992). *Qualitative analysis of client's experience of psychotherapy: The unfolding of reflexivity.* Paper presented at the This chapter is derived from a paper on clients' reflexivity presented at the First International Conference in Developmental Counseling, Oporto, Portugal, Jul 1988, and from a related paper presented at the Twentieth Annual Meeting of the Society for Psychotherapy Research, Toronto, Jun 1989.

Ring, A., Dowrick, C. F., Humphris, G. M., Davies, J., & Salmon, P. (2005). The somatising effect of clinical consultation: what patients and doctors say and do not say when patients present medically unexplained physical symptoms. *Social science & medicine, 61*(7), 1505-1515.

Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2016). The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. *Patient education and counseling, 99*(1), 164-168.

Rode, S., Salkovskis, P. M., Dowd, H., & Hanna, M. (2006). Health anxiety levels in chronic pain clinic attenders. *Journal of Psychosomatic research, 60*(2), 155-161.

Safran, J. D., Muran, J. C., & Eubanks-Carter, C. (2011). Repairing alliance ruptures. *Psychotherapy, 48*(1), 80.

Salkovskis, P. M. (1989). Somatic problems. In K. Hawton (Ed.), *Cognitive Behaviour Therapy for Psychiatric Problems: a Practical Guide* (pp. 235 - 276). Oxford: Oxford University Press.

Salkovskis, P. M., Rimes, K., Warwick, H., & Clark, D. (2002). The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychological Medicine, 321*, 843-853.

Salkovskis, P. M., & Warwick, H. (1986). Morbid preoccupations, health anxiety and reassurance: a cognitive-behavioural approach to hypochondriasis. *Behaviour research and therapy, 24*(5), 597-602.

Salkovskis, P. M., Warwick, H., & Deale, A. C. (2003). Cognitive-behavioral treatment for severe and persistent health anxiety (hypochondriasis). *Brief Treatment and Crisis Intervention, 3*(3), 353-367.

Salmon, P., Humphris, G., Ring, A., Davies, J., & Dowrick, C. (2006). Why do primary care physicians propose medical care to patients with medically unexplained symptoms? A new method of sequence analysis to test theories of patient pressure. *Psychosom Med, 68*, 570 - 577.

Schiefele, A.-K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., . . . Nielsen, S. L. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 598-613.

Schöttke, H., Flückiger, C., Goldberg, S. B., Eversmann, J., & Lange, J. (2017). Predicting psychotherapy outcome based on therapist interpersonal skills: A five-year longitudinal study of a therapist assessment protocol. *Psychotherapy research, 27*(6), 642-652.

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2013). *Mindfulness-based cognitive therapy for depression* (2nd ed.). London, UK: Guilford Press.

Sharabiani, M. T., Aylin, P., & Bottle, A. (2012). Systematic review of comorbidity indices for administrative data. *Medical care*, 1109-1118.

Shepherd, L., Salkovskis, P. M., & Morris, M. (2009). Recording therapy sessions: An evaluation of patient and therapist reported behaviours, attitudes and preferences. *Behavioural and cognitive psychotherapy, 37*(2), 141-150.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of consulting and clinical psychology, 78*(3), 298.

Shorten, A., & Smith, J. (2017). Mixed methods research: expanding the evidence base. In: Royal College of Nursing.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420.

Siegel, S., & Castellan, N. (1988). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.

Sijercic, I., Button, M. L., Westra, H. A., & Hara, K. M. (2016). The interpersonal context of client motivational language in cognitive-behavioral therapy. *Psychotherapy, 53*(1), 13-21. doi:http://dx.doi.org/10.1037/pst0000017

Simon, H. A. (1992). What is an “explanation” of behavior? *Psychological Science, 3*(3), 150-161.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Smith, J. (1996). Beyond the divide between cognition and discourse: Using interpretative phenomenological analysis in health psychology. *Psychology and health, 11*(2), 261-271.

Smith, J., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: theory, method and research.* . London, UK: Sage.

Smits, F., Brouwer, H., van Weert, H., Schene, A., & ter Riet, G. (2009). Predictability of persistent frequent attendance: a historic 3-year cohort study. *Br J Gen Pract, 59*, e44 - e51.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.

Soto, A., Smith, T. B., Griner, D., Domenech Rodríguez, M., & Bernal, G. (2018). Cultural adaptations and therapist multicultural competence: Two meta-analytic reviews. *Journal of Clinical Psychology, 74*(11), 1907-1923.

Spitzer, R., Kroenke, K., & Williams, J. M. G. (2006). A brief measure for assessing generalised anxiety disorder; the GAD-7. *Archives of Internal Medicine, 146*, 1092-1097.

Stafford, M., Steventon, M., Thorlby, R., Fisher, R., Turton, C., & Deeny, S. (2018). Understanding the health care needs of people with multiple health conditions. Retrieved from https://www.health.org.uk/sites/default/files/upload/publications/2018/Understanding%20the%20health%20care%20needs%20of%20people%20with%20multiple%20health%20conditions.pdf

Steventon, A., Deeny, S., Friebel, R., Gardner, T., & Thorlby, R. (2018). Emergency hospital admissions in England: Which may be avoidable and how? Retrieved from www.health.org.uk/publication/emergency-hospital-admissions-england-which-may-be-avoidable-and-how

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.), 21*(1), 128.

Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice, 5*(4), 439-458.

Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (1999). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour research and therapy, 48*(7), 599-606.

Strunk, D. R., Brotman, M. A., DeRubeis, R. J., & Hollon, S. D. (2010). Therapist competence in cognitive therapy for depression: Predicting subsequent symptom change. *Journal of consulting and clinical psychology, 78*(3), 429.

Sunderland, M., Newby, J. M., & Andrews, G. (2013). Health anxiety in Australia: prevalence, comorbidity, disability and service use. *The British Journal of Psychiatry, 202*(1), 56-61.

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: a meta-analysis. *Journal of consulting and clinical psychology, 80*(4), 547.

Swift, J. K., Tompkins, K. A., & Parkin, S. R. (2017). Understanding the client's perspective of helpful and hindering events in psychotherapy sessions: A micro-process approach. *Journal of Clinical Psychology, 73*(11), 1543-1555.

Szkodny, L. E., Newman, M. G., & Goldfried, M. R. (2014). Clinical experiences in conducting empirically supported treatments for generalized anxiety disorder. *Behavior therapy, 45*(1), 7-20.

Tanana, M., Hallgren, K. A., Imel, Z. E., Atkins, D. C., & Srikumar, V. (2016). A comparison of natural language processing methods for automated coding of motivational interviewing. *Journal of substance abuse treatment, 65*, 43-50.

Tang, T. Z., & DeRubeis, R. J. (1999a). Reconsidering rapid early response in cognitive behavioral therapy for depression. *Clinical Psychology: Science and Practice, 6*(3), 283-288.

Tang, T. Z., & DeRubeis, R. J. (1999b). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of consulting and clinical psychology, 67*(6), 894.

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of consulting and clinical psychology, 73*(1), 168-172.

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., & Shelton, R. (2007). Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *Journal of Consulting & Clinical Psychology, 75*(3), 404-408. doi:10.1037/0022-006X.75.3.404

Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: are they also found in psychotherapies other than cognitive-behavioral therapy? *Journal of consulting and clinical psychology, 70*(2), 444.

Tasca, G. A., & Gallop, R. (2009). Multilevel modeling of longitudinal data for psychotherapy researchers: I. The basics. *Psychotherapy research, 19*(4-5), 429-437.

Taskforce on Multiple Conditions. (2018). *'Just One Thing After Another': Living with Multiple Conditions*. London, UK: Richmond Group of Charities.

Tichenor, V., & Hill, C. E. (1989). A comparison of six measures of working alliance. *Psychotherapy: Theory, Research, Practice, Training, 26*(2), 195.

Timulak, L. (2009). Meta-analysis of qualitative studies: A tool for reviewing qualitative research findings in psychotherapy. *Psychotherapy research, 19*(4-5), 591-600.

Town, J. M., Abbass, A., Stride, C., & Bernier, D. (2017). A randomised controlled trial of Intensive Short-Term Dynamic Psychotherapy for treatment resistant depression: the Halifax Depression Study. *Journal of Affective Disorders, 214*, 15-25.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the working alliance inventory. *Psychological Assessment: A journal of consulting and clinical psychology, 1*(3), 207.

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research, 27*(3), 247-259.

Tyrer, H., Tyrer, P., Lisseman-Stones, Y., McAllister, S., Cooper, S., Salkovskis, P. M., . . . Murphy, D. (2015). Therapist differences in a randomised trial of the outcome of cognitive behaviour therapy for health anxiety in medical patients. *International journal of nursing studies, 52*(3), 686-694.

Tyrer, P., Cooper, S., Crawford, M., Dupont, S., Green, J., Murphy, D., . . . Bhogal, S. (2011). Prevalence of health anxiety problems in medical clinics. *Journal of Psychosomatic research, 71*(6), 392-394.

Tyrer, P., Cooper, S., Salkovskis, P., Tyrer, H., Crawford, M., Byford, S., . . . McLaren, E. (2014). Clinical and cost-effectiveness of cognitive behaviour therapy for health anxiety in medical patients: a multicentre randomised controlled trial. *The Lancet, 383*(9913), 219-225.

Tyrer, P., Cooper, S., Tyrer, H., Salkovskis, P. M., Crawford, M., Green, J., . . . Murphy, D. (2011). CHAMP: Cognitive behaviour therapy for health anxiety in medical patients, a randomised controlled trial. *BMC psychiatry, 11*(1), 99.

Tyrer, P., Eilenberg, T., Fink, P., Hedman, E., & Tyrer, H. (2016). Health anxiety: the silent, disabling epidemic.

Tyrer, P., Salkovskis, P. M., Tyrer, H., Wang, D., Crawford, M. J., Dupont, S., . . . Smith, G. (2017). Cognitive–behaviour therapy for health anxiety in medical patients (CHAMP): a randomised controlled trial with outcomes to 5 years. *Health Technology Assessment, 21*(50), 1-58.

University of Exeter. (2019). High Intensity IAPT handbooks and resources: Assessment of clinical skills / CBT competencies via formative and summative asessments using the CTS-R for recordings of clinical sessions submitted on Encrypted Storage Devices (ESDs) or via secure data transfer. Retrieved from http://cedar.exeter.ac.uk/iapt/hihandbook/assessments/ctsr/

van Ravesteijn, H., Grutters, J., olde Hartman, T., Lucassen, P., Bor, H., van Weel, C., . . . Speckens, A. (2013). Mindfulness-based cognitive therapy for patients with medically unexplained symptoms: a cost-effectiveness study. *J Psychosom Res, 74*(3), 197-205. doi:10.1016/j.jpsychores.2013.01.001

van Reenen, M., & Janssen, B. (2015). *EQ-5d-5L User Guide*. Rotterdam, NL: EuroQol Research Foundation.

Viklund, E., Holmqvist, R., & Zetterqvist Nelson, K. (2010). Client-identified important events in psychotherapy: Interactional structures and practices. *Psychotherapy research, 20*(2), 151-164.

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*(2), 639.

Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). London: Routledge.

Ware, J. E., Kosinski, M. A., & Gandek, B. (2005). *SF-36 Health Survey: Manual and interpretation guide*. Lincoln: Quality Metric Inc.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of consulting and clinical psychology, 78*(2), 200.

Webb, C. A., DeRubeis, R. J., Dimidjian, S., Hollon, S. D., Amsterdam, J. D., & Shelton, R. C. (2012). Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: The role of therapist adherence and the therapeutic alliance. *Journal of consulting and clinical psychology, 80*(3), 373.

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of social psychology, 26*(4), 557-580.

Westra, H. A. (2011). Comparing the predictive capacity of observed in-session resistance to self-reported motivation in cognitive behavioral therapy. *Behaviour Research & Therapy, 49*(2), 106-113. doi:https://dx.doi.org/10.1016/j.brat.2010.11.007

Westra, H. A. (2012). *Motivational interviewing in the treatment of anxiety.* New York, NY: Guilford Press.

Westra, H. A. (2014). Using motivational interviewing to manage resistance. In S. G. Hofmann, D. J. A. Dozois, W. Rief, & J. A. J. Smits (Eds.), *The Wiley handbook of cognitive behavioral therapy* (Vol. 1-3, pp. 331-352). Chichester, UK: Wiley-Blackwell.

Westra, H. A., & Arkowitz, H. (2010). Combining motivational interviewing and cognitive behavioural therapy to increase treatment efficacy for generalized anxiety disorder. . In D. Sookman & B. Leahy (Eds.), *Resolving treatment impasses with resistant anxiety disorders* (pp. 199-232). New York, NY: Routledge.

Westra, H. A., Arkowitz, H., & Dozois, D. J. A. (2009). Adding a motivational interviewing pretreatment to cognitive behavioral therapy for generalized anxiety disorder: A preliminary randomized controlled trial. *Journal of Anxiety Disorders, 23*(8), 1106-1117. doi:http://dx.doi.org/10.1016/j.janxdis.2009.07.014

Westra, H. A., Aviram, A., Connors, L., Kertes, A., & Ahmed, M. (2012). Therapist emotional reactions and client resistance in cognitive behavioral therapy. *Psychotherapy, 49*(2), 163.

Westra, H. A., Constantino, M., & Antony, M. M. (2016). Integrating motivational interviewing with cognitive-behavioral therapy for severe generalized anxiety disorder: An allegiance-controlled randomized clinical trial. *Journal of consulting and clinical psychology, 84*(9), 768.

Westra, H. A., & Norouzian, N. (2018). Using motivational interviewing to manage process markers of ambivalence and resistance in cognitive behavioral therapy. *Cognitive Therapy and Research, 42*(2), 193-203.

Wiborg, J. F., Knoop, H., Wensing, M., & Bleijenberg, G. (2012). Therapist effects and the dissemination of cognitive behavior therapy for chronic fatigue syndrome in community-based mental health care. *Behaviour research and therapy, 50*(6), 393-396.

Williams, A., Eccleston, C., & Morley, S. (2012). Psychological therapies for the management of chronic pain (excluding headache) in adults. In.

Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51-58.

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., . . . Sterne, J. A. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ, 336*(7644), 601-605.

Xiao, B., Can, D., Gibson, J., Imel, Z. E., Atkins, D. C., Georgiou, P. G., & Narayanan, S. S. (2016). *Behavioral Coding of Therapist Language in Addiction Counseling Using Recurrent Neural Networks.* Paper presented at the Interspeech.

Xiao, B., Imel, Z. E., Georgiou, P. G., Atkins, D. C., & Narayanan, S. S. (2015). " Rate My Therapist": Automated Detection of Empathy in Drug and Alcohol Counseling via Speech and Language Processing. *PloS one, 10*(12), e0143055.

11. APPENDICES

APPENDIX A: Systematic review search strategy

Database(s): **PsycINFO** 1806 to January Week 4 2019

#	Searches
1	exp MOTIVATION/
2	exp MOTIVATION MEASURES/
3	exp MOTIVATION TRAINING/
4	exp RESISTANCE/
5	exp PSYCHOTHERAPEUTIC RESISTANCE/
6	exp Psychological Reactance/
7	exp "Stages of Change"/
8	(motivation* or "patient activation*" or engagement or engaged or engage or engaging or activation or "change talk" or "counter?change talk").tw.
9	((readiness or motivat* or stage* or willing* or desire) adj2 (chang* or action* or act)).tw.
10	(resist* or reactan* or discord*).tw.
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12	exp ANXIETY/
13	exp ANXIETY DISORDERS/
14	(anxiet* or anxious or panic or phobi* or GAD).tw.
15	(PTSD or post?trauma*).tw.
16	(stress adj disorder*).tw.
17	(OCD or obsess* or compulsi*).tw.
18	(fear* adj2 ("air?travel" or animal* or blood* or buses or "closed space*" or "enclosed space" or "public space*" or crowd* or dark* or dental* or dentist* or dog or dogs or dying or flight* or fly or flying or health or height* or illness* or injection* or injur* or lightning or movement* or needle* or "negative evaluation" or panic* or performance or plane* or snake* or social or space* or spider* or sweat* or thunder* or train* or travel* or trembl* or water)).tw.
19	(acrophob* or agoraphob* or agrophobi* or "anxiety disorder*" or claustrophob* or emetophob* or arachnophob* or hypochondri* or kinesiophob* or shyness).tw.
20	((anxiety* or phobi*) adj2 neuros*).tw.
21	((generalised or generalized or health) adj2 anxiety).tw.
22	((interpersonal or "inter personal" or social or speak*) adj2 (anxiet* or avoid* or fear*)).tw.
23	(panic adj2 (attack* or disorder* or fear or symptom*)).tw.
24	12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23

25	exp Cognitive Behavior Therapy/
26	exp Cognitive Therapy/
27	(''cognitive behaviour therap*'' or ''cognitive behavior therap*'' or ''cognitive behaviour psychotherap*'' or ''cognitive behavior psychotherap*'' or CBT or ''cognitive behavioural therap*'' or ''cognitive behavioral therap*'' or ''cognitive behavioural psychotherap*'' or ''cognitive behavioral psychotherap*'' or ''cognitive therap*'' or ''cognition therap*'').tw.
28	25 or 26 or 27
29	11 and 24 and 28
30	limit 29 to english language

Database(s): **Ovid MEDLINE(R)** 1946 to January Week 4 2019

Search Strategy:

#	**Searches**
1	exp Motivation/
2	(motivation* or ''patient activation*'' or engagement or engaged or engage or engaging or activation or ''change talk'' or ''counter?change talk'').tw,kf,kw.
3	((readiness or motivat* or stage* or willing* or desire) adj2 (chang* or action* or act)).tw,kf,kw.
4	(resist* or reactan* or discord*).tw,kf,kw.
5	1 or 2 or 3 or 4
6	exp Anxiety/
7	exp Anxiety Disorders/
8	(anxiet* or anxious or panic or phobi* or GAD).tw,kf,kw.
9	(PTSD or post?trauma*).tw,kf,kw.
10	(stress adj disorder*).tw,kf,kw.
11	(OCD or obsess* or compulsi*).tw,kf,kw.
12	(fear* adj2 (''air?travel'' or animal* or blood* or buses or ''closed space*'' or ''enclosed space'' or ''public space*'' or crowd* or dark* or dental* or dentist* or dog or dogs or dying or flight* or fly or flying or health or height* or illness* or injection* or injur* or lightning or movement* or needle* or ''negative evaluation'' or panic* or performance or plane* or snake* or social or space* or spider* or sweat* or thunder* or train* or travel* or trembl* or water)).tw,kf,kw.
13	(acrophob* or agoraphob* or agrophobi* or ''anxiety disorder*'' or claustrophob* or emetophob* or arachnophob* or hypochondri* or kinesiophob* or shyness).tw,kf,kw.
14	((anxiety* or phobi*) adj2 neuros*).tw,kf,kw.
15	((generalised or generalized or health) adj2 anxiety).tw,kf,kw.

16	((interpersonal or "inter personal" or social or speak*) adj2 (anxiet* or avoid* or fear*)).tw,kf,kw.
17	(panic adj2 (attack* or disorder* or fear or symptom*)).tw,kf,kw.
18	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
19	exp Cognitive Behavioral Therapy/
20	("cognitive behaviour therap*" or "cognitive behavior therap*" or "cognitive behaviour psychotherap*" or "cognitive behavior psychotherap*" or CBT or "cognitive behavioural therap*" or "cognitive behavioral therap*" or "cognitive behavioural psychotherap*" or "cognitive behavioral psychotherap*" or "cognitive therap*" or "cognition therap*").tw,kf,kw.
21	19 or 20
22	5 and 18 and 21
23	limit 22 to english language

Database(s): **Embase** 1980 to 2019 Week 04

Search Strategy:

#	**Searches**
1	exp motivation/
2	exp therapy resistance/
3	(motivation* or "patient activation*" or engagement or engaged or engage or engaging or activation or "change talk" or "counter?change talk").tw.
4	((readiness or motivat* or stage* or willing* or desire) adj2 (chang* or action* or act)).tw.
5	(resist* or reactan* or discord*).tw.
6	1 or 2 or 3 or 4 or 5
7	exp anxiety/
8	exp anxiety disorder/
9	exp anxiety neurosis/
10	(anxiet* or anxious or panic or phobi* or GAD).tw.
11	(PTSD or post?trauma*).tw.
12	(stress adj disorder*).tw.
13	(OCD or obsess* or compulsi*).tw.
14	(fear* adj2 ("air?travel" or animal* or blood* or buses or "closed space*" or "enclosed space" or "public space*" or crowd* or dark* or dental* or dentist* or dog or dogs or dying or flight* or fly or flying or health or height* or illness* or injection* or injur* or lightning or movement* or needle* or "negative evaluation" or panic* or performance or plane* or snake* or social or space* or spider* or sweat* or thunder* or train* or travel* or trembl* or water)).tw.

15	(acrophob* or agoraphob* or agrophobi* or "anxiety disorder*" or claustrophob* or emetophob* or arachnophob* or hypochondri* or kinesiophob* or shyness).tw.
16	((anxiety* or phobi*) adj2 neuros*).tw.
17	((generalised or generalized or health) adj2 anxiety).tw.
18	((interpersonal or "inter personal" or social or speak*) adj2 (anxiet* or avoid* or fear*)).tw.
19	(panic adj2 (attack* or disorder* or fear or symptom*)).tw.
20	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19
21	exp cognitive behavioral therapy/
22	("cognitive behaviour therap*" or "cognitive behavior therap*" or "cognitive behaviour psychotherap*" or "cognitive behavior psychotherap*" or CBT or "cognitive behavioural therap*" or "cognitive behavioral therap*" or "cognitive behavioural psychotherap*" or "cognitive behavioral psychotherap*" or "cognitive therap*" or "cognition therap*").tw.
23	21 or 22
24	6 and 20 and 23
25	limit 24 to english language

APPENDIX B: Example completed PROBAST form

Step 1: Specify your systematic review question
State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. *The following table should be completed once per systematic review.*

Criteria	Specify your systematic review question
Intended use of model:	Predicting anxiety-related outcomes following CBT
Participants *including selection criteria and setting:*	Adults experiencing an anxiety disorder, who have been treated for it with individual CBT
Predictors (used in modelling) *including (1) types of predictors (e.g. history, clinical examination, biochemical markers, imaging tests), (2) time of measurement, (3) specific measurement issues (e.g. any requirements/ prohibitions for specialized equipment):*	Observer-rated motivation for treatment, assessed up to (and not beyond) halfway through therapy. This includes negative aspects of motivation (e.g. resistance and counter-change talk) and positive aspects of motivation (e.g. change talk).
Outcome *to be predicted:*	Self-reported anxiety using standardised assessments such as the Penn State Worry Questionnaire.

Step 2: Classify the type of prediction model evaluation
Use the following table to classify the evaluation as model development, model validation, or combination.
Different signalling questions apply for different types of prediction model evaluation.
When a publication focuses on adding one or more new predictors to established predictors then use "development only". When a publication focuses on validation of an existing model in other data though followed by updating (adjusting or extending) of the model such that in fact a new model is being developed, then use "development and validation in the same publication".
If the evaluation does not fit one of these classifications then PROBAST should not be used.

Classify the evaluation based on its aim			
Type of prediction study	**PROBAST boxes to complete**	**Tick as appropriate**	**Definition for type of prediction model study**
Development only	Development	y	Prediction model development without external validation. These studies may include internal validation methods, such as bootstrapping and cross-validation techniques.
Development and validation	Development and validation	n	Prediction model development combined with external validation in other participants in the same article.
Validation only	Validation	n	External validation of existing (previously developed) model in other participants.
This table should be completed once for each publication being assessed and for each relevant outcome in your review.			
Publication reference	Lombardi, D. R., Button, M. L., & Westra, H. A. (2014). Measuring Motivation: Change Talk and Counter-Change Talk in Cognitive Behavioral Therapy for Generalized Anxiety. Cognitive Behaviour Therapy, 43(1), 12-21.		
Models of interest	Predicting post-CBT worry from first and second session observed motivation		
Outcome of interest	Post-CBT worry		

Step 3: Assess risk of bias and applicability
PROBAST is structured as four key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signalling questions to help make judgments. Signalling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signalling questions are phrased so that "yes" indicates absence of bias. Any signalling question rated as "no" or "probably no" flags the potential for bias; you will need to use your judgment to determine whether the domain should be rated as "high", "low" or "unclear" risk of bias. The guidance document contains further instructions and examples on rating signalling questions and risk of bias for each domain.
The first three domains are also rated for concerns for applicability (low/ high/ unclear) to your review question defined above.
Complete all domains separately for each evaluation of a distinct model. Shaded boxes indicate where signalling questions do not apply and should not be answered.

DOMAIN 1: Participants			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection: *46 were allocated to receive CBT but only 37 were analysed. No information is given about the way in which those not included were excluded or if there were identified differences between samples.*			
		Dev	*Val*
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		*y*	
1.2 Were all inclusions and exclusions of participants appropriate?		*NI*	
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*	*high*	
Rationale of bias rating:			
20% of potential participants were excluded without explanation. In the original trial some had dropped out, but it is unclear whether they had received any CBT sessions by this point or if they went on to complete post-treatment outcomes. Plus, the original trial suggests 38 CBT participants' data were analysed but only 37 were included in the current analysis.			
B. Applicability			
Describe included participants, setting and dates: *Participants were recruited from public adverts and a standardised tool used to assess GAD diagnostic criteria. Participants were recruited from 2007-2009.*			
Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*	*low*	
Rationale of applicability rating:			
The participants all had an anxiety disorder validated through standard interview, were offered a quality-checked CBT treatment and observed motivation was assessed in the first and second session.			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors included in the final model, e.g. definition and timing of assessment: *Baseline work (PSWQ)* *Pre-treatment, self-reported motivation (CMOTS and CQ)* *Observed motivation at first or second session (MISC 1.1)*			
		Dev	*Val*
2.1 Were predictors defined and assessed in a similar way for all participants?		*y*	
2.2 Were predictor assessments made without knowledge of outcome data?		*y*	
2.3 Are all predictors available at the time the model is intended to be used?		*y*	
Risk of bias introduced by predictors or their assessment	**RISK:** *(low/ high/ unclear)*	*low*	
Rationale of bias rating: *All predictors were appropriately treated for all participants and outcomes were not known to any associated parties.*			
B. Applicability			
Concern that the definition, assessment or timing of predictors in the model do not match the review question	**CONCERN:** *(low/ high/ unclear)*	*low*	
Rationale of applicability rating: *The timing and relevance of all predictors help to answer the review question.*			

DOMAIN 3: Outcome

A. Risk of Bias

Describe the outcome, how it was defined and determined, and the time interval between predictor
assessment and outcome determination:

The PSWQ is a standardised self-reported measure of worry. It was the primary outcome of the original trial and is therefore an appropriate outcome for this model. 14 hours of CBT were delivered over a period of 12 weeks. Therefore, outcomes were recorded approximately 12 weeks after the predictors, which was an appropriate gap.

		Dev	*Val*
3.1 Was the outcome determined appropriately?		*y*	
3.2 Was a pre-specified or standard outcome definition used?		*y*	
3.3 Were predictors excluded from the outcome definition?		*y*	
3.4 Was the outcome defined and determined in a similar way for all participants?		*y*	
3.5 Was the outcome determined without knowledge of predictor information?		*y*	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?		*y*	
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*	*low*	

Rationale of bias rating:

There is small concern that having reported a high level of motivation at baseline this might lead to a falsely inflated degree of worry reduction due to social desirability. However, the interval between predictor and outcome should have provided a suitable washout period. Therefore, there was a low risk of bias being introduced.

B. Applicability

At what time point was the outcome determined:
Post-CBT – approximately 12 weeks after the predictors were measured.

If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:

Concern that the outcome, its definition, timing or determination do not match the review question	**CONCERN:** *(low/ high/ unclear)*	*low*	

Rationale of applicability rating:
All elements of the outcome were suitable for the review question.

DOMAIN 4: Analysis			
Risk of Bias			
Describe numbers of participants, number of candidate predictors (for DEV only), outcome events and events per candidate predictor (for DEV only): *N = 35* *Predictors = 3* *Outcomes = 35 probably?* *Outcomes per predictor = 7.4* *This is likely to be insufficient power for a reliable analysis*			
Describe how the model was developed (predictor selection, optimism, risk groups, model performance): *Predictors were selected on theoretical basis – that early motivation can predict outcomes in CBT for anxiety* *Predictors were added in four blocks:* *1 – Pre-CBT worry (PSWQ)* *2 – Self reported pre-CBT motivation CQ and CMOTS)* *3 – Observed first and second session motivation (positive: CT and negative: CCT)* *4 (in response to results of block 3) – observed first and second session negative motivation*			
Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants): *No validation was reported.*			
Describe the performance measures of the model, e.g. (re)calibration, discrimination, (re)classification, net benefit: *N/A*			
Describe any participants who were excluded from the analysis: *The reasons for exclusion are unclear, but it is likely that most dropped out of treatment and therefore outcome data were unavailable.*			
Describe missing data on predictors and outcomes as well as methods used for missing data: *Missing data is not clearly reported and no methods were reported on how missing data were managed.*			
		Dev	*Val*
4.1 Were there a reasonable number of participants with the outcome?		*n*	
4.2 Were continuous and categorical predictors handled appropriately?		*y*	
4.3 Were all enrolled participants included in the analysis?		*n*	
4.4 Were participants with missing data handled appropriately?		*pn*	
4.5 Was selection of predictors based on univariable analysis avoided?		*n*	
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?		*n*	
4.7 Were relevant model performance measures evaluated appropriately?		*n/a*	
4.8 Were model overfitting and optimism in model performance accounted for?		*n/a*	
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?		*ni*	
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*	*high*	
Rationale of bias rating: *Several routes through which bias could be introduced were not addressed. This is again most likely due to the small sample size, meaning that the all models were underpowered. Where no information was reported this is likely due to the methods not being used.*			

Step 4: Overall assessment

Use the following tables to reach overall judgments about risk of bias and concerns for applicability of the prediction model evaluation (development and/or validation) across all assessed domains.

Complete for each evaluation of a distinct model.

Reaching an overall judgment about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to **high risk of bias**. Such a model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at **high risk of bias**.
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Reaching an overall judgment about applicability of the prediction model evaluation	
Low concerns for applicability	If low concerns for applicability for all domains, the prediction model evaluation is judged to have **low concerns for applicability**.
High concerns for applicability	If high concerns for applicability for at least one domain, the prediction model evaluation is judged to have **high concerns for applicability**.
Unclear concerns for applicability	If unclear concerns (but no "high concern") for applicability for at least one domain, the prediction model evaluation is judged to have **unclear concerns for applicability** overall.

Overall judgment about risk of bias and applicability of the prediction model evaluation		
Overall judgment of risk of bias	**RISK:** *(low/ high/ unclear)*	*high*
Summary of sources of potential bias: *The unexplained sub-selection of participants and the small sample size to predictor ratio introduce bias.*		
Overall judgment of applicability	**CONCERN:** *(low/ high/ unclear)*	*low*
Summary of applicability concerns: *Low concern.*		

APPENDIX C: Developmental phase consent form

CLAHRC Nottinghamshire, Derbyshire, Lincolnshire | The University of Nottingham | NHS National Institute for Health Research

PARTICIPANT CONSENT FORM
(Version 1.2 1-8-2012)

Title of Study: Regular Attenders in Primary Care- a case study approach

REC ref: 11/ EM/ 0392

Name of Chief Investigator: Prof Richard Morriss

Name of Participant:

	Please initial box
1. I confirm I have read and understand the information sheet (Participant Information Sheet (case formulation and treatment), version 1.2, 1-8-2012 for the above study and have had the opportunity to ask questions.	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected. I understand that should I withdraw then the information collected so far cannot be erased and that this information may still be used in the project analysis.	☐
3. I understand that relevant sections of my medical notes and data collected in the study may be looked at by authorised individuals from the University of Nottingham, the research group and regulatory authorities where it is relevant to my taking part in this study. I give permission for these individuals to have access to these records and to collect, store, analyse and publish information obtained from my participation in this study.	☐
4. I agree to my GP being informed of my participation in this study.	☐
5. I agree to take part in the above study.	☐
6. I understand that should any malpractice or abuse by health service staff towards patients be identified, this will be reported to the Principal Investigator of the study and he will take appropriate steps to ensure this is dealt with by the relevant authorities.	☐
7. I give permission for my treatment sessions to be audio-recorded. I understand that any quotes used in publications, reports or training materials will be anonymous and I will not be identified.	☐

Name of Participant	Date	Signature
Name of Researcher	Date	Signature

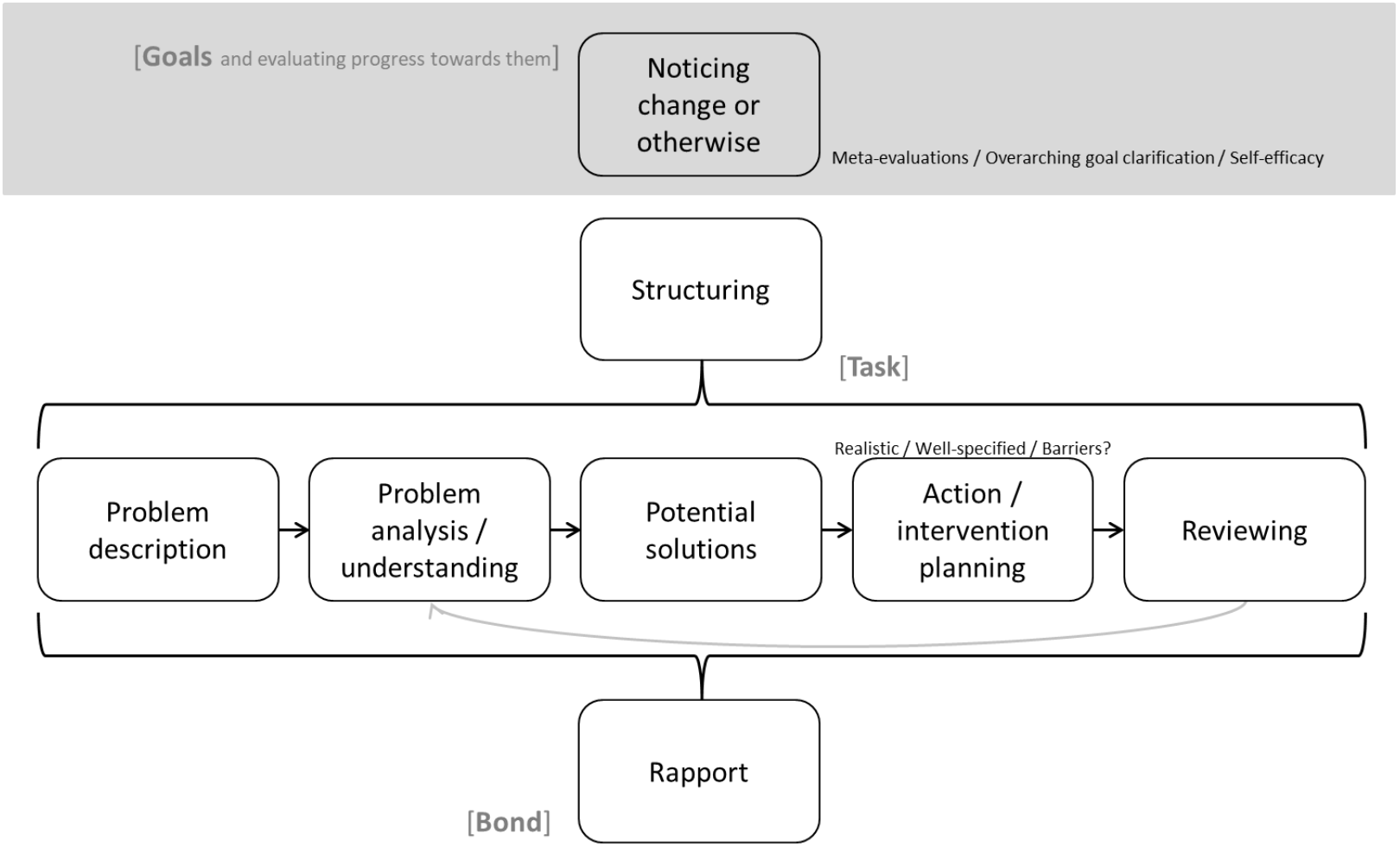
3 copies: 1 for participant, 1 for the project notes and 1 for the medical notes

APPENDIX D: Theme development examples

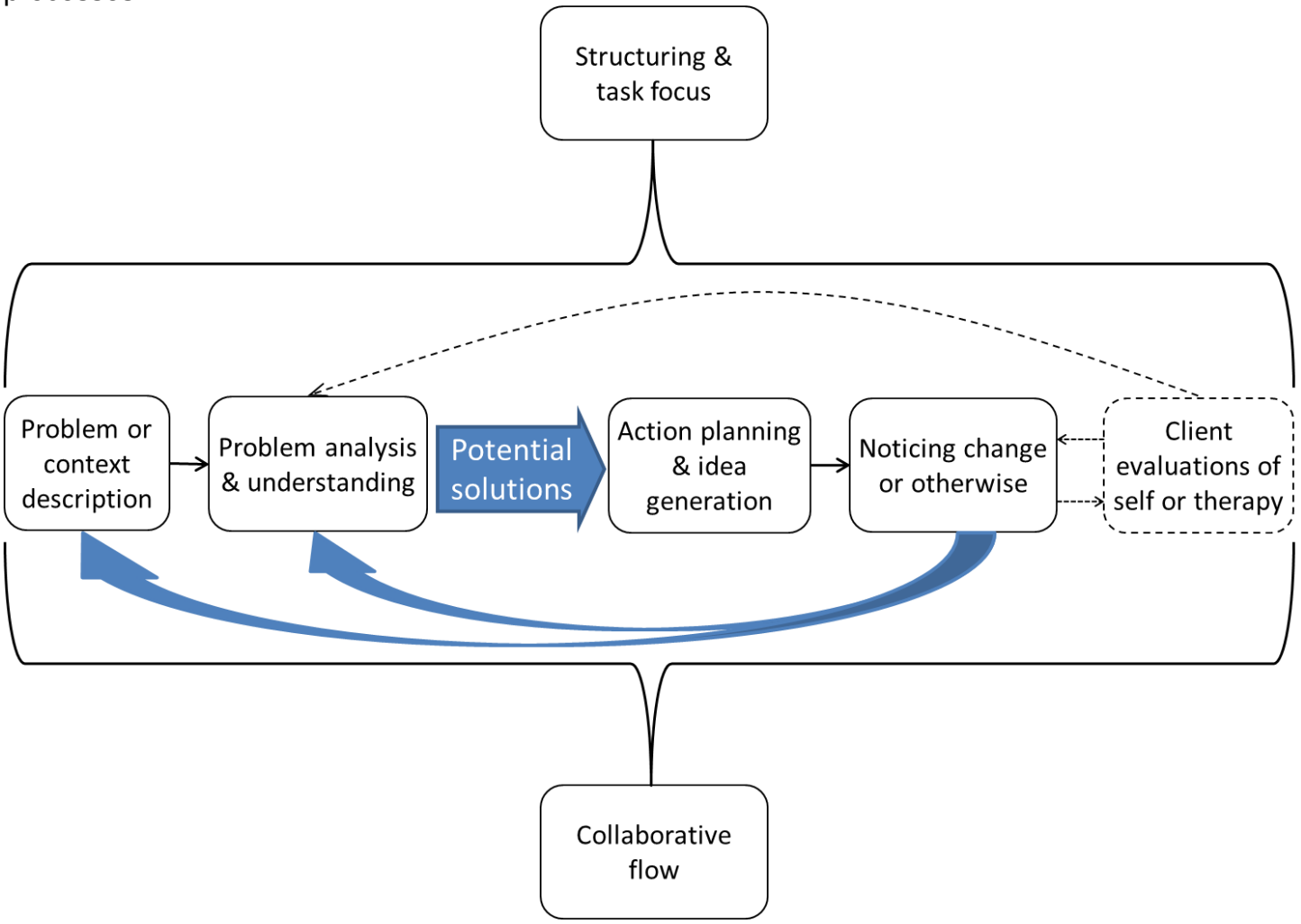
Initial Theme Template 8/10/13



Developing Theme Template 19/1/15 – Some themes were deemed to be part of a progressive process whilst others were seen throughout all interactive processes



Approaching Final Theme Template – Clearer delineation of feedback loops and integration of themes crossing interactive processes



APPENDIX E: Examples of coder reflections

Reflections on coding the later session for Participant 1 by coder 2:

Psychological mindedness?

Participant 1 seems to be unwilling or unable to identify and discuss thoughts and feelings with therapist. Therapist continually probing for psychological meaning (and sometimes resorts to directing client?) but often seems to be closed down (brief, minimising responses; sometimes with some [defensive?] laughter). Difficult to tell whether client becomes annoyed with this probing, but does seem to be some talking at cross-purposes - which may limit therapeutic connection and collaboration.

Presenting stoicism?

Minimising language very recurrent - frequently qualifies description of illness experiences with 'just' (implying that things could be worse...). 'Just' may also emphasise that it is only what it appears to be (i.e., take it at face value - don't probe for deeper meaning) and can be read as fatalistic ('just' is this way); may also position the client as stoical (getting on with it, in spite of suffering).

Comparative reflections from Participant 4's earlier session from coder 2:

Participant 4 talkative and fairly tangential, clearly preoccupied with relationship difficulties (principally with daughter) and focus of the session shifts to accommodate this. Therapist does offer points to refocus (and checks whether session had covered key concerns); seems a strength that session was responsive to client, but could also be criticised for ostensibly focussing away from presenting issue (formulation not clear, so may be seen that dealing with relationship difficulties is actually fundamental to illness presentation and attendance behaviour - however, does seem to be more of a fire-fighting session on first read).

Therapist style notably very different to first transcripts (Participant 1): I constructed codes for reflection and validation for the first time - seems subjectively stronger on therapeutic bond and client-centredness (particularly listening skills - but perhaps reflects more talkative and emotionally-expressive client?). Perhaps less structured/orientated around therapist plan than Participant 1 sessions though?

Coder 2 overall reflections part way through analysis:

Something I've noticed but am not sure is captured in my coding so far:

Problems are frequently raised (either by client or therapist - sometimes therapist helps to delineate the problem or bring in to focus when client is avoidant/minimising of problem) but variability in how these are dealt with...

Sometimes therapist helps client to elaborate on problem (and attendant responses), frame it more constructively, generate solutions, and identify personal strengths or sources of support.

Sometimes problems are raised but deflected or not processed further (may be that therapist misses cue or client avoids/shuts down further discussion). Moreover, sometimes problems are discussed but without connecting to possible solutions or resources (focus on strengths? Code for this...).

Differences here could be meaningful if we're trying to understand relative 'success' of therapy case, particularly given common factors literature; however, limits to what we can see in single transcripts (e.g., may be that problems are returned to at a later date in a more or less constructive way).

Coder 1 reflections on coding Participant 4's later session:

This session seemed more focused on fairly detailed analysis of problem situations. D40 seemed much free-er to disagree with the therapists sentiments and talked more about the barriers to change as well as talking about positive changes that had been noticed. They both talked about being "interested" in what was going on, which perhaps contributed to the extended sections analysing problems. The tasksD40 was left to go away with were less clearly defined though and at times they ended up with several explanatory strands lefts dangling only partially explored with not as much focus as the first session.

Comparative reflections from coding Participant 3's later session

Compared to D40, this participant's presentation is far less consistent, with strong sentiments expressed for not being able to change or cope alongside equally strong sentiments that change must occur. The therapist's focus seems equally variable, with a strong psychoeducation and task focus at times combined with a wandering to and fro from an assessment that is trying to be completed

Coder 1 overall reflections part way through analysis:

At times I end up thinking like a therapist and wanting to write codes which are a kind of critical evaluation of the therapy rather than a summary reflection of what has been said.

Linked to that, being therapeutically trained definitely impacts on how I view the data. I start thinking in terms of therapeutic processes probably too quickly - i.e. reducing talk down to something I am aware should be happening in therapy which risks missing something other than a therapeutic process.

Consultation Interactions Coding Scheme (CICS) 1.6



Sam Malins[1], Nima Moghaddam[2], Richard Morriss[1], Thomas Schröder[1], Paula Brown[1], Naomi Boycott[1]

[1]Division of Psychiatry and Applied Psychology, University of Nottingham, UK
[2]School of Psychology, University of Lincoln, UK

Correspondence: sam.malins@nottingham.ac.uk

Contents

Background 3

Guidance for coders 4

- Coding terms 4
- Coder training 4
- Coding levels 4
- Decision tree 7
- Decision rules for theme choice 8
- Segmenting 9
- Scoring 12
- Implications 12

Session Content Themes 13

1. Action Planning and Idea Generation (API) 14
2. Evaluations of Self or Therapy (EST) 16
3. Information Discussion (IDI) 17
4. Noticing Change or Otherwise (NCO) 19
5. Problem Analysis and Understanding (PAU) 21
6. Problem or Context Description (PCD) 22
7. Structuring and Task Focus (STF) 23
8. Other (OTH) 24

Global Themes 25

- Collaborative Flow (COF) 26
- Integration of Themes (IOT) 27

References 32

Appendix I: Theme and level summary sheet 33

Appendix II: Theme scoring sheet 35

Background (Return to top)

Previous literature has found that certain types of healthcare interaction contribute to improved health outcome more than others (e.g. Moyers, Martin, Houck, Christopher, & Tonigan, 2009). Patient activation is an identified contributor to improved health outcomes, particularly in long-term or comorbid health issues, but its impact has not been explored at an interaction level (Greene, Hibbard, Sacks, Overton, & Parrotta, 2015; Mosen et al., 2007; Schmaderer, Zimmerman, Hertzog, Pozehl, & Paulman, 2016). Patient activation is a recognised concept focused on the knowledge, skills and confidence a person has in their ability to be involved in and manage their health (Hibbard, Stockard, Mahoney, & Tusler, 2004). This coding manual categorises different interaction topics or themes and uses patient activation to characterise the types of interaction that may contribute to improved outcomes and those that contribute to non-improvement or deterioration.

The themes used in this manual are derived from a thematic analysis of 22 transcripts of Cognitive Behaviour Therapy (CBT) provided to 11 patients in a case-series study of CBT for long-term high utilisers of primary care (Malins et al., 2016). Multimorbidity was identified as a key feature of the problems experienced by this patient group (Patel et al., 2015). Patient activation and interaction-style in clinical consultations have been identified as effective intervention routes for people experiencing multiple, chronic physical and mental health problems where targeted treatments are difficult to select and apply. (Deen, Lu, Rothstein, Santana, & Gold, 2011; Derksen, Bensing, & Lagro-Janssen, 2013; Mercer, Gunn, Bower, Wyke, & Guthrie, 2012).

For each participant a session from the earlier half and latter half of therapy was randomly selected. An inductive thematic analysis was conducted by two coders working independently, using transcripts of the selected sessions. Whilst blind to participant outcomes, a theory-informed deductive analysis was then carried out. Patient activation theory was used to establish interaction types most likely to contribute to outcome. This resulted in a process model of themes grouped into three categories: Theme presentations likely to contribute to: (1) therapeutic improvement; (2) non-improvement or deterioration; and theme presentations where (3) no clear distinction could be made.

This manual is used to categorise the interactions in psychological therapy sessions on the basis of patient activation. The coding process is described below, followed by descriptions of each theme, a coding summary sheet (Appendix 1) and a code scoring sheet (Appendix 2).This manual does not aim to be used as a standalone method of training coders, but is to be used in conjunction with guided coding sessions led by experienced coders, as described in the training process below.

Guidance for Coders (Return to top)

Coding Terms

Turn - A patient or therapist turn of speech, which continues until the other person begins speaking instead.

Segment – A session extract to be coded. This could include a single turn or several turns.

Theme - The overall topic of interaction which will be coded. Each is described in "Session content themes" (p15) and "Global themes" (p26). Themes are rated at different levels of patient activation. When a specific theme and level is applied to a particular session segment it is "coded".

Coder Training

Code six transcripts in three sets pairs with two hour training sessions between each pair. These transcripts are benchmarked "gold standard" transcripts coded by at least two experienced coders where themes and levels were jointly agreed. Inter-rater reliability is deemed to be achieved when an overall Intra-class correlation coefficient of 0.7 or above is achieved across all themes and levels. Further transcripts are coded if necessary.

Coding Levels

Coding a session using this manual requires session recordings transcribed verbatim. Audio-visual recordings support and enhance coding, but are not required. First code the transcript using session content themes – note turns of speech where each theme is present. All themes are rated on one of five levels ranging from +2 to -2. There are two levels for what are deemed to be therapeutically positive interaction types (i.e. moving towards therapeutic change), graded +1 and +2. There are two levels for what are deemed to be therapeutically negative interaction types (i.e. moving away from therapeutic change), graded -1 and -2. Between them lies a neutral level (coded as 0). When coding an extract, begin at 0/neutral and move up or down if sufficient observable evidence is present to warrant doing so. Use 0/neutral as the default position or if the polarity is unclear. Each level is characterised in a similar fashion across themes, broadly focused on the level of patient activation:

+2: Hypothesised to be the optimal therapeutic interaction type within the theme. Key attributes are a high level of patient activation and focus; a therapeutic interaction or activity *led by the patient without an explicit invitation or elicitation* from the foregoing therapist turn-of-speech.

Example 1

Therapist: How is your family life? **Patient:** Well, I've realised something that would help us all. I've decided that it's really important for me to start running again. I used to run all the time and it was really helpful. **Therapist:** Great idea! When could you start?	Action Planning & Idea Generation, +2

+1: Hypothesised to be a positive interaction, but with a lower level of patient activation and leadership. Typically this would be a positive interaction led or guided by the therapist, which the patient is both endorsing and developing. A +2 level interaction would be differentiated because it would be patient-led interaction *not directly responding to the previous therapist* interaction.

Example 2

Therapist: What about starting running again? You used to enjoy that. **Patient:** Yes, I'd been thinking that too. Maybe I could do it in the morning before work.	Action Planning & Idea Generation, +1

0 (Neutral): These are interaction where *few discernible positive or negative interactive factors* are apparent and is deemed to be neutral. This includes interactions where therapists make suggestions or comments with little observable sign of how the comment has been received and limited receptivity on the patient's part.

Example 3

Therapist: What about starting running again? You used to enjoy that. **Patient:** Maybe. **Therapist:** You could start tomorrow morning. **Patient:** Yeah.	Action Planning & Idea Generation, 0

-1*:* Hypothesised to be interactions that are therapeutically unhelpful in a minor way. This includes interactions which show the beginnings *of unaddressed disagreements or reluctance to engage* with therapeutic activities. Low levels of patient activation and involvement are observable, independent of therapist activity.

Example 4

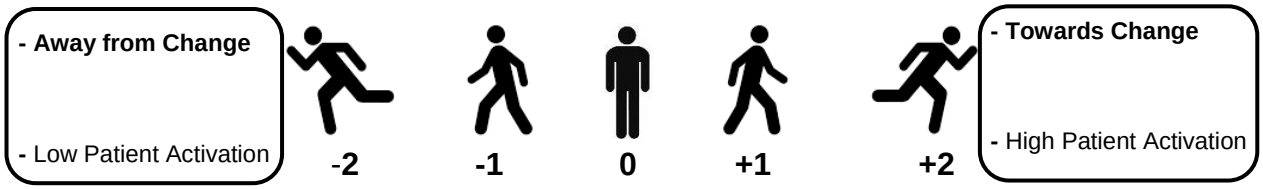
Therapist: What about starting running again? You used to enjoy that. **Patient:** I'm not sure that would work. **Therapist:** You could start tomorrow. **Patient:** I doubt it.	Action Planning & Idea Generation, -1

-2*:* Hypothesised to be interactions that would be contradictory to most therapeutic guidance. This would include argumentative or obstructive interactions where the patient (and potentially the therapist) appear(s) *disengaged and/or unfocused and/or oppositional* to therapeutic activity.

Example 5

Therapist: What about starting running again? You used to enjoy that. **Patient:** That's a terrible idea. You clearly don't understand me. **Therapist:** Maybe if you got off the couch and did something you wouldn't be so negative. **Patient:** That has nothing to do with it!	Action Planning & Idea Generation, -2

The overall approach to coding levels can be illustrated pictorially as below. The level coded is a proxy for the level of patient activation and orientation to change that can be inferred from the segment under review. A level of 0 would denote an interaction where neither movement towards or away from change can be observed nor is there sufficient observable evidence to say that the patient is either activated or disengaged. Positive levels indicate increasing evidence of movement towards change and active therapeutic engagement. Conversely, negative levels indicate increasing movement away from change and disengagement.



Apply the level that seems the best fit for the extract under review. Where applicable, the key component for the level of each theme is displayed in *italics* to emphasise its significance.

After coding session content themes:

1. Focus on the global scales. Allocate the level that is the best fit for the transcript as whole, although there are likely to be deviations from this at points in the session.
2. Review codes that have already been allocated, with knowledge of the context of the whole session. You may wish to add or remove themes that were not apparent in the first pass. This is most likely to happen in large extracts where one theme of the same level was recorded. Some of the *levels* coded may also need readjusting.

Themes can be coded concurrently at the coder's discretion. In particular, *Problem analysis and understanding (PAU)*, *Noticing change or otherwise (NCO)* and *Structuring and Task Focus (STF)* codes are likely to be coded concurrently with other themes.

Before coding any sessions, coders should familiarise themselves with each theme and the relevant level descriptors. A theme summary sheet is provided at the end of this manual which can be used as an ongoing reference whilst coding (Appendix 1).

Segmenting (Return to top)

Apply at least one theme at a specific level to each turn of speech from the therapist and patient. A single theme and level is coded for as long as it is the best fit for the segment. This can incorporate several patient and therapist turns.

Example 7

***Therapist*:** So when we stay with something that makes us anxious, over time, we often find the anxiety starts to fade as we get used to it. ***Patient*:** That makes sense, because eventually I just forgot all about that rash.	Information Discussion , +1

Longer turns may be coded with more than one theme or a change in the theme level. This is particularly likely with the *PCD* theme.

Example 8

***Patient*:** At times I end up searching the internet to see what's wrong, it's really bad and I just get myself so worried. I'm strong enough to beat this though; I know I'm going to get there.	Problem or Context Description evaluation of self or therapy, +2

Small interjections of one theme within an extract clearly linked to another theme may be omitted.

Example 9

Therapist: Have you noticed any changes in your mood? ***Patient:*** I am feeling a bit better I think since we started meeting. I have still got this niggling pain in my neck that won't go away and I'm not sure what to do about it, but on the whole things are definitely better. **Therapist:** What do you think has brought this change about?	Noticing Change or Otherwise, +1 (not coded) Problem or Context Description

In Example 9 where a patient is prompted to talk about changes they have noticed, the NCO theme would be applied at the +1 level. There is a small interjection which on its own would be coded under *PCD* (underlined) but this does little to disrupt the overall flow of the existing theme and so is not coded. Furthermore, the therapist response is focused on the NCO theme.

In Example 10 the *NCO* theme is clearly stopped and the *PCD* theme is started, which should be an additional code for that turn. In particular, the change in theme is emphasised by the fact that the next turn responds to the problem description rather than the change that has been noticed.

Example 10

Therapist: Have you noticed any changes in your mood? ***Patient:*** I am feeling a bit better I think since we started meeting. I have still got this niggling pain in my neck that won't go away and I'm not sure what to do about it. It's dragging me down. **Therapist:** That does sound bad, tell me about the pain	Noticing Change or Otherwise, +1 Problem or Context Description

If two themes are apparently present in one turn the turn is split and coded with each theme. The level of the theme is determined by the response it receives. If a theme is ignored, the level coded is 0.

Example 11

Therapist: I want to explain something about vicious cycles in panic attacks: Noticing that your heart is racing can make you worried about what it means, but this worry makes your heart race more. So I think that we will try some techniques to help you get used to the feeling of your heart racing. ***Patient:*** Some techniques to help with my heart racing would be really useful, because that is a real problem for me.	Information Discussion, 0 Action Planning & Idea Generation, +1

In this example the therapist turn begins with a segment which is clearly within the *Information Discussion (IDI)* theme. However, it transitions to an *Action Planning and Idea Generation (API)* theme. The coding decision relies on the patient's response. The patient responds to the *API* component over the *IDI* component therefore the *IDI* theme is coded at level 0 as there is no clear response from which to judge it. The *API* theme is coded at +1 level as there is a positive developing response.

Facilitative language. Where there are segments that do not seem to apply to any particular theme, but are mainly facilitative in nature they are coded at the same theme and level as the preceding segment:

Example 12

Patient: I have been suffering with this problem for several years now without any change. **Therapist:** Seven years? **Patient:** No several years. **Therapist:** Several? **Patient:** Yeah. **Therapist:** Ok, sorry.	Problem or Context Description

Returning to neutral. Where a segment is being coded at either a positive (level +1 and +2) or a negative level (-2 and -1) of a particular theme there are likely to be occurrences where coding needs to return to neutral (0).

Example 13

Therapist: It sounds like you see things differently now. **Patient:** Absolutely, I can see that I was wearing myself out and beating myself up when I would have been able to achieve more if I had paced my activity. **Therapist:** Yes, to stop when you've done enough for your body. **Patient:** Yes. **Therapist:** And accept that you can't do it all at once. **Patient:** Yeah. **Therapist:** Because that makes it worse **Patient:** That's right and what I can do is perfectly normal for a man of my age. I sometimes expect myself to work like a 20 year old! **Therapist:** You were expecting too much of yourself.	Problem Analysis & Understanding, +1 Problem Analysis & Understanding, 0 Problem Analysis & Understanding, +1

Example 13 begins with two turns coded as *PAU, +1* because the patient provides development of their understanding cued by the therapist. The underlined segment is a return to *PAU, 0* because there is no patient development of the analysis presented by the therapist for these turns. Then there is further development later on so the level returns to *PAU, +1.*

Stepping up from +1 to +2. For there to be a step up from +1 to +2 there needs to have been a deviation from the topic (typically by the therapist) to allow the patient to initiate a new segment in its own right as +2.

Example 14

Therapist: So how about if you started off by only checking the internet once a day? **Patient:** Ok, yeah so for maybe ten minutes. **Therapist:** And we'll see how you got on next week **Patient:** Yeah, ok. **Therapist:** Great. **Patient:** Also**,** I've actually had a new idea about how to tell my boss not to put so much pressure on me. I'm going to arrange a meeting and then ask her how she feels first. **Therapist:** Brilliant idea! When are you thinking of doing that?	Action Planning & Idea Generation, +1 Action Planning & Idea Generation, 0 Action Planning & Idea Generation, +2

In the example above, the action planning starts by being guided by the therapist, but there is a clear break to a level 0 and a clear restart initiated by the patient. Hence the last segment is coded at the +2 level.

Decision Tree (Return to top)

Coding involves a three stage decision-making process:

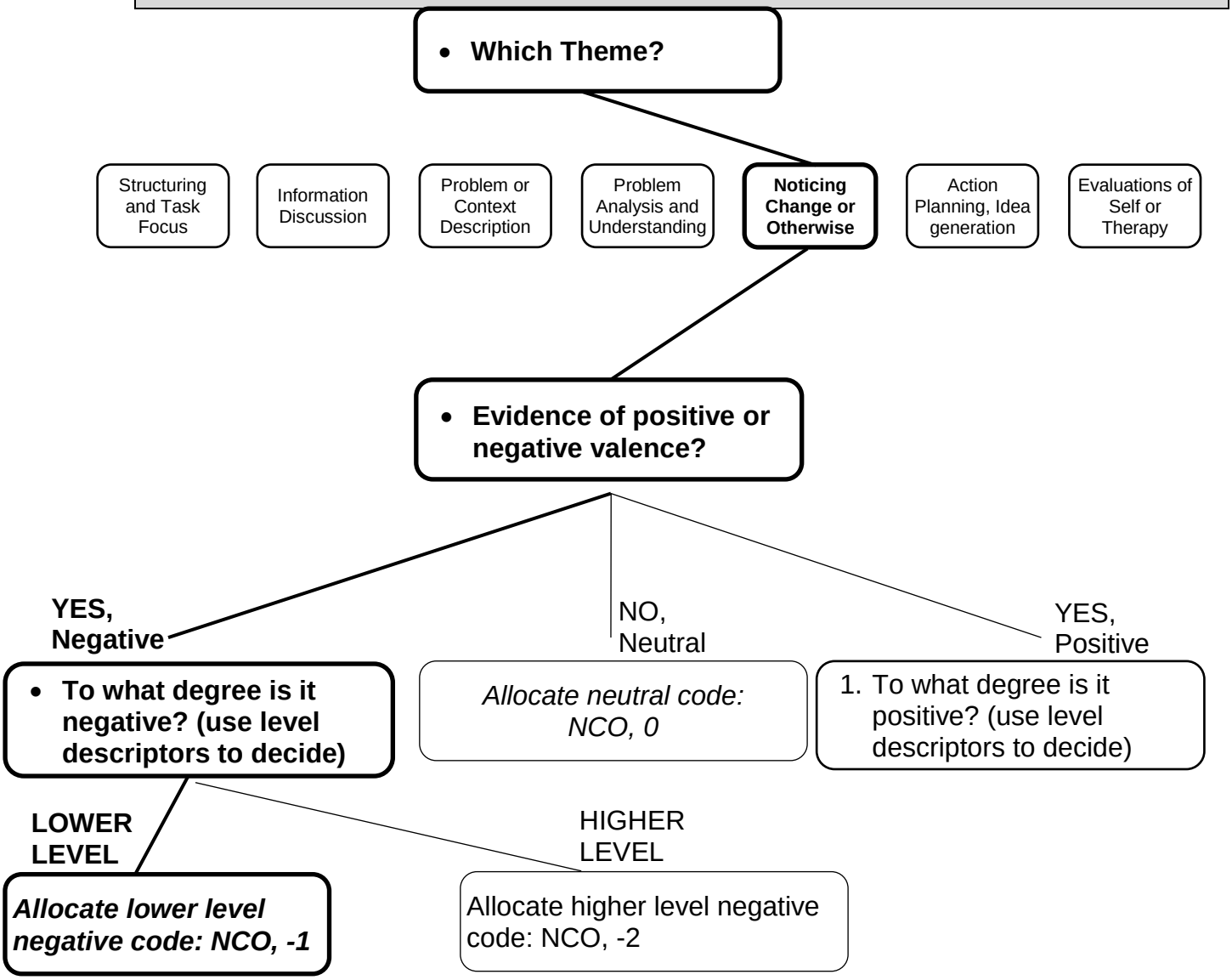
- Which theme shall I code for the current segment?
- Is there enough evidence for the level coded to be either above or below level 0? (0/neutral is the default code).
- If yes, to what degree is the current segment positive (+1 or +2) / negative (-2 or -1).

Therefore, a coder first decides which of the session content themes should be coded for the current segment, then decides whether it demonstrates sufficient difference from neutral to warrant a positive or negative valence and finally a decision is made about the degree to which it demonstrates a positive or negative valence. Overall, coders apply the level that is the best fit, but if there is a lack of clarity the more conservative score is given (i.e. +1 instead of +2 if unsure).

Decision-Tree Example:

Therapist: It seems like things are really different now, you've done well.

Patient: Hmm, I'm not so sure about that, it's still pretty rough.



Decision Rules for Theme Choice (Return to top)

Two rules are applied in deciding which theme is most appropriate for a particular segment:

Primary Rule: Could the segment stand alone, independent of the context, and still be identified correctly by the theme definition and level descriptor? If a clear theme and level cannot be achieved using this rule, the secondary rule should be applied.

Secondary Rule: Which theme is the best fit for this segment given the context, before and after it occurs? This would include themes and levels applied 2-3 turns before the segment and 2-3 turns after the segment. If still unclear a different code should be considered – PCD being the most common in areas of doubt.

Example 15

***Therapist*:** When you stop checking your body you're less bothered by pain and worries. ***Patient*:** Yes, it was keeping me worried when I thought it was helping.	Information Discussion, +1 **OR** Problem Analysis & Understanding, +1

When the segment in this example is seen independent of context it seems that two themes could be coded: *IDI* or *PAU.* To decide between these two the broader context is reviewed:

Therapist: So you tried an experiment this week to see what happened if you didn't check your body for a week. How did it go? ***Patient:*** It was so different I can't believe it. By the end of the week I was hardly thinking about the pain in my stomach at all! I hadn't realised checking was making me anxious! ***Therapist*:** When you stop checking your body you're less bothered by pain and worries. ***Patient*:** Yes, it was keeping me worried when I thought it was helping.	Noticing Change or Otherwise, +1 **AND** Problem Analysis & Understanding, +1 Problem Analysis & Understanding +1

With the context of the previous two turns it becomes clear that this is not a situation where the therapist is offering new information, but is summarising the new understanding that the patient has reached from the experiment tried between sessions. As such, the segment is coded as *PAU, +1*.

Scoring (Return to top)

Scoring aims to summarise an overall picture of the spread of themes and levels present across a segment or session. It provides a visual and analytical impression of how much of the session was spent in interactions that were positively coded or otherwise.

Multiply the proportion of the transcript coded at each level of each theme by the score for the given level (TOTAL 1). For example, if 10% of a transcript was coded as Action Planning and Idea generation (API) at the +2 level, this would contribute a score of 20 to the total for the API theme (+2 x 10). Conversely, if 10% of the same

transcript was scored as API -1, this would contribute a score of -10 to the total for the API theme (-1 x 10). If these were the only levels of API coded for the transcript a total score for API on that transcript would be 10 (20-10). To calculate overall content scores, average the scores for each theme (TOTAL 2). Global themes are scored separately with a single rating on the level scored.

NVivo 12 coding software can be used manage transcript data and to give the percentage of a transcript coded at a given theme level.

Implications

From coding scores, outcome can be predicted. This is based on the proportion of the transcript coded at high levels as opposed to neutral or low levels. If specific themes and/or rating levels predict health outcomes this would give therapeutically valuable feedback to therapists on the relationship between patient outcome and specific interactions types. As such the coding manual could be used to develop skills for therapists in training, as well as helping refine skills and enhance outcome for those in practice.

SESSION CONTENT THEMES (Return to top)

Session content theme definitions and level descriptors are given below.

Action Planning and Idea Generation (API) (Return to top)

Coded where either party suggests an action plan or potential problem solution idea linked to activities which might take place *outside the session*. The level coded is dependent on how the idea is developed or refined and the degree of observable engagement, particularly from the patient. The refinement of ideas may stray into what seems like a PCD theme, as the barriers to actions are discussed. This should be coded as PCD and return to API if further action planning is restarted. If the rationale for a particular action is given but no explicit suggestion is made that it could be done then this should be coded as IDI. For actions that are planned to be completed in the session use the STF theme.

Patient: I want to build up my fitness so I can do more things with the family, but I know I'll have to pace myself and build it up over time. ***Therapist:*** Where would you want to start with that? ***Patient:*** I was thinking that I could be doing more walking than I have been doing. I get the bus into town, and I could maybe walk to the next stop along from where I normally catch it.	API, +2

+2. *Patient initiated* action plans or change ideas that have potential to be applied through refinement and discussion.

+1. Therapist initiated action plans or change ideas, the patient agrees with *development, summary or corrections.*

0. Therapist initiated action plans or change ideas – *no development or clear sense of patient view.*

-1. Therapist or patient initiated action plan or change idea – *unresolved reluctance or ambivalence expressed* by patient or therapist towards the action plan.

-2. Therapist or patient initiated action plan or change idea – *ignored, refuted or distracted to another topic* and no agreement is reached.

Patient: I do find myself getting quite het up about it ***Therapist:*** Next time you feel that way, you should have a go at the breathing exercises in this booklet.	API, -2

Patient: I don't want to do that because the other thing is that I think I've got the wrong eye drops, but they've said I've got to wait for a month to see how it goes.	

Evaluations of Self or Therapy (EST) (Return to top)

Code where there are interactions focused on evaluating therapy and/or related activities; self-supporting statements, or instances where completion of a therapeutic task is acknowledged as an accomplishment. This also includes statements of praise or support for the therapist or patient. The level coded is dependent upon who initiates the comment and level of patient engagement and agreement. This theme is often coded alongside NCO for specific evaluative phrases.

Patient: I do think this has been helpful. I'm doing things that I thought were beyond me, and getting back my confidence.	EST, +2

+2. Patient initiated statements of *self-efficacy* OR patient *acknowledgement/pride* at therapeutic achievement OR *positive evaluations* of therapy or the therapist initiated by the patient.

+1. Therapist initiated positive evaluation, as above, patient agrees with *development, summary or corrections.*

0. The above expressed by therapist – *ignored or low level acknowledgement* by the patient.

-1. Positive comments as above are *undermined to some degree* by patient OR *somewhat negatively focused* self-statements or statements about therapy/therapist.

-2. *Self-denigrating or self-critical* statements OR a *self-critical focus* on therapeutic tasks that have not been completed to the exclusion of those that have.

Therapist: So you didn't manage to complete the diary this week? ***Patient:*** It's just another example, isn't it? I can't do anything on my own. When I go out that door, I just mess it all up.	EST, -2

Information Discussion (IDI) (Return to top)

Code where information is given or sought. The degree of relevance and reciprocity affects the level to be coded. Higher ratings demonstrate patient engagement with, and development of therapist information. Lower ratings demonstrate reluctance to engage with, or active rejection of therapist information. If information is provided by a therapist in the middle of a turn, but the patient does not directly respond to the information given then the relevant section of the therapist turn is coded as IDI, 0. This is differentiated from PAU because any information given is theoretical, abstract or general, as opposed to a personalised conceptualisation or explanation of a situation (see example 1). Key to this is the use of pronouns. Information using "we/you plural, "people" etc. would be coded as IDI. Whereas information using personal pronouns (you singular) would be coded as PAU. Segments coded as IDI contain information which could be delivered to an audience of several people and would not make sense if it was not delivered to the individual.

Patient: It would be really helpful if we could talk about how panic attacks work like you mentioned last time, so I know what I'm dealing with. ***Therapist:*** Yeah, great idea. Shall we look at a recent example?	IDI, +2

+2. *Patient initiated request or cue* for specific information, relevant to current therapeutic endeavour. Encouraged, endorsed or developed by the therapist.

+1. Therapist initiated information giving which the patient acknowledges and *develops further, summarises or seeks clarification.*

0. Therapist initiated information-giving - *no development* OR *development suggests unresolved misunderstanding* OR therapist acknowledges patient request for information, but does not develop.

-1. Therapist gives information – *patient displays ambivalence* in response OR gentle dismissal OR Information given does not seem relevant or applicable and this is clear from their responses.

-2. Information given that *elicits a negative response*, such as active rejection or mocking.

Therapist: Let me tell you what needs to happen here. People with anxiety have got to face their fears to beat them.	IDI, -2

Patient: That's easy for you to say! If I did that I'd probably wind up worse off!	

Problem Analysis and Understanding (PAU) and Noticing Change or Otherwise (NCO) *are likely to be coded in parallel. Code PAU alone where an explanation of a problem is given with no application to the patient's life. Code NCO alone where changes are described without any explicit additional understanding. Where a change is noticed and applied to a new understanding both themes are coded.*

Problem Analysis and Understanding (PAU) (Return to top)

Code where a new understanding is explained by either party. The level coded is dependent on the degree of reciprocity and development expressed. Higher ratings demonstrate new patient understandings, even if initiated by the therapist, that logically feed into interventions. Lower ratings demonstrate either clearly counterproductive understandings or oppositional responses to presentations of new understandings. Interactions would move from PCD ratings when questioning or descriptive interactions move to *explanatory* interactions. PAU is differentiated from IDI because any information given is personalised to the individual within a specific context, conceptualising an explanation of a situation (see example 1).

PAU +2 *without* NCO

Patient: I realise that I've been fighting myself all this time, I've been my own worst enemy. Trying to get away from anxiety keeps me anxious! ***Therapist:*** That sounds like an important insight. How did you come to that conclusion?	PAU, +2

PAU +2 *with* NCO

Patient: When I looked back over the last few weeks I realised as I've been staying with the anxiety I've actually stopped feeling as anxious. It's having a massive impact on what I can do. ***Therapist:*** You've noticed a real difference then.	PAU, +2 NCO, +2

+2. *Patient led* articulation of a new understanding, where the solutions may be clearer (e.g. I've realised; I now see). There is some evidence that therapeutic concepts are understood. Endorsed or encouraged by therapist.

+1. Therapist initiates the above and *it is developed, summarised or corrected* by the patient OR patient-led new understanding with *limited clarity corrected, clarified, endorsed or encouraged* by therapist.

0. Therapist presents a problem analysis of some apparent therapeutic value and this is acknowledged by the patient with *no development* OR a new understanding is presented by either party of *little or no therapeutic value*.

-1. Therapist presents a problem analysis and the patient displays *ambivalence in their response without resolution*. OR the patient presents a problem analysis *with the potential to be counterproductive.*

-2. Therapist presents a formulation and this is *refuted by the patient. No alternative is given* and *no agreement* is reached OR the patient presents an analysis that is *anti-therapeutic* and this is not opened for correction by the therapist.

PAU 0 *without* NCO

Therapist: So from what we've discussed it looks like you're bringing these problems on yourself. ***Patient:*** That's not what's going on. You really don't get it do you. ***Therapist:*** Look, this leaflet says the same as what I'm telling you.	PAU, -2

PAU 0 *with* NCO

Therapist: So you stopped seeking reassurance from your doctor and you felt less worried, maybe constantly going to him was making you anxious. ***Patient:*** It's just by chance though isn't it? Nothing's really change. I'm just on a lucky streak with my health.	PAU, -2 NCO, -2

Noticing Change or Otherwise (NCO) (Return to top)

Code where positive change (or absence thereof) is noticed and explicitly commented upon by either party. This would be linked to the patient's efforts and/or the therapist's efforts. This typically includes changes since therapy began but can include other patient-initiated changes without external support/prompting. The level coded depends upon the degree to which the patient notices, develops and personally owns positive or negative changes. If the change is extended to a meta-evaluation of the therapy or of the patient themselves it would also be coded as Evaluation of Self or Therapy (EST).

Patient: Last time I was waiting for test results, I just couldn't cope. I didn't know what to do with myself. This time though, I've been much better at calming myself and just getting on with the things that I'd normally be doing. ***Therapist:*** Tell me more about the things you've been doing differently this time.	NCO, +2

+2. Patient notices a positive change, *articulates it without prompting*. Likely to include: its impact on their life; some sense of personal ownership, or a sense of control over the change. Endorsed or encouraged by therapist.

+1. The above prompted by therapist – patient endorses *and summarises or develops*.

0. Therapist initiated acknowledgement of change - *no development* from patient OR very minor acknowledgment of change by the patient *not personally owned*, perhaps relating the change to sources outside of their control.

-1. Therapist initiated acknowledgement of change, *patient displays ambivalence* towards suggested change OR *Efforts to make a change are reported but none is seen*.

-2. Therapists initiated acknowledgement of change patient *refutes positive change*. OR Efforts to make a change are reported *but deterioration results*.

Patient: I guess I was hoping that things would improve, but it just isn't happening. ***Therapist:*** You say that, but I think you've made lots of positive changes since we started.	NCO, -2

Patient: Well, I can't think of any, and it doesn't feel that way. If anything, I feel like I'm going backwards.	

Problem or Context Description (PCD) (Return to top)

Coded for apparently neutral interactions where no other theme is appropriate to be applied. This will often constitute a problem description being given or sought. It includes general assessment interactions and descriptions of the problem context, but is not limited to these topics. Levels of coding are not applied to this theme – the theme is either present or absent. The PCD theme is only stopped where there is distinct movement to a different theme. When general discussions coded as PCD deviate off relevant topics for two turns or more STF -1/-2 is coded, as it is deemed a lack of focus. If interactions totally unrelated to therapy take place Other (OTH) is coded.

Patient: I just find that I'm always inside the house and I feel totally trapped no matter what I do. ***Therapist:*** That sounds really tough.	PCD

Structuring and Task Focus (STF) (Return to top)

Code where significant changes in session direction take place, initiated by either therapist or patient. Typically, the following six turns are tracked to see if the change in direction is maintained by both parties. This theme is also coded when working through a structured task and clear instructions are being given and followed – in this case as few as one following turn can be coded. The theme includes changes of direction initiated by *drifting or deviating* from a specified topic. This may be coded alongside another theme relevant to the content of the change in direction. It includes agenda mapping that may occur at the beginning and throughout the session. Higher ratings demonstrate maintenance of a strong focus on relevant topics, endorsed, followed and/or developed by both parties. Lower ratings demonstrate unaddressed changes in focus, moving away from or seeming to avoid relevant topics; if maintained (i.e. not corrected within two turns of speech) these passages are negatively coded for the duration of the deviation regardless of number of turns. A positive code is applied if a party draws the conversation back onto relevant topics after a deviation and this is maintained.

Patient: I would like to start today by talking about the homework, I realised there were a few things I hadn't understood that I'd like you to clarify. ***Therapist:*** Ok, that sounds good. Which bits weren't clear? ***Patient:*** Well, I filled in the diary but I wasn't sure if I was meant to put how I felt physically – like tired – or like worried? ***Therapist:*** Would it help if went through the diary to get it clear? ***Patient:*** Yeah.	STF, +2

+2.Patient *leading, cuing or initiating* a relevant session focus or task; *self-monitoring when deviating* or distracted and *redirecting* the therapist if they deviate. Structuring can still be done by therapist, but is initiated or cued by patient. This initiation is supported by the therapist.

+1.Therapist and patient maintaining a strong focus on a new direction which is *maintained for significant periods (6 turns)* or the discussion is completed. Also includes guided therapeutic tasks, to which the patient responds and follows (no specified number of turns). This can include agreeing to put off something for a later point.

0. Redirections and course corrections from the therapist or patient result in temporary shift of focus, *but focus is not maintained*.

-1. Either party drifts or deviates from relevant issues or are easily distracted onto *tangents about unrelated topics or problems*. Redirections and course corrections from either party have *little or no impact on deviating discourse*. The -1 level is coded if deviations continue for 2-4 turns of speech.

-2. Little or no direction provided by either party, leading to long deviations, general chat or topics that *do not lead to therapeutic activity*. The -2 level is coded if deviations continue for more than 4 turns of speech. OR A didactic style from the therapist receives an *oppositional response from the patient*.

Therapist: Let's move on now ***Patient:*** I want to finish what I was saying. ***Therapist:*** We need to move on, we don't have time. ***Patient:*** But it's important to me. ***Therapist:*** No we need to move on.	STF, -2

Other (OTH) – There may be session segments that cannot appropriately be coded using the other themes identified. In such instances, this is coded "OTH" and a brief description is given of what is observed. (Return to top)

GLOBAL THEMES (Return to top)

The two scales below are rated as interaction overviews of the session as a whole. Attention should be focused on these scales after using content themes to code the whole transcript and an overall feel for the types of interactions present in the session can be given.

Collaborative Flow (COF) (Return to top)

Assesses the extent of joint understanding and joint working displayed through the session as a whole, which may be difficult to characterise from individual incidents. At its best this is where both parties build on one another's ideas over several exchanges (at least two contributions from each party) in a way that moves the discussion productively forward. If taken together it could almost constitute a single paragraph which could be attributed to one or other party's speech. This is described as "collaborative flow". Assign higher ratings for transcripts showing a more ofthis type of interaction (in quality or quantity) and lower ratings for transcripts where there is a notable absence of joint understanding or the transcript is characterised by interactions where joint working is hindered.

+2. There are more than three segments of collaborative flow. Collaborative flow characterises session interactions as a whole.

+1. There are between one and three segments of collaborative flow or several brief single episodes (one turn each) with some movement away from mutual enthusiasm, engagement and understanding at points. There is an overall sense of mutual understanding, collaboration and direction for periods, but periods where this may not be as clear.

0. Little or no evidence of collaborative flow. The therapeutic partnership weighs more heavily on the therapist as director and initiator. The patient may agree with the therapist's proposals and direction, but offers little developmental contribution of their own. OR The therapist offers only agreement with the patient and little guidance or structure to develop patient input.

-1. There are significant periods where the patient and therapist are disconnected in their agendas or aims. There is little or no joint working in the session as a whole.

-2. The patient and therapist appear, to a large extent, to be talking across purposes with different agendas or goals in mind. This characterises most interactions in the session. OR the session is characterised by excessive agreement or collusion without a sense of how this agreement could move the conversation forward therapeutically.

Integration of Themes (IOT) (Return to top)

Assesses the interconnectedness of themes and the “natural” flow or fluency between them. In positive codes multiple content themes are positively coded through the transcript and there is coherent movement between themes, with logical coverage across themes. In negative codes there is imbalance in coverage of themes, most heavily weighting on PCD with few positively coded themes.

+2. Significant periods of patient led fluid movement between four or more positively scored themes. Often a logically sequenced session flow.

+1. Some fluid movement and integration of two or three positively scored themes. Some periods where movement is limited or disjointed.

0. Limited integration and flow between different themes, some seemingly important themes are notably absent or ignored. Disconnected movement between unlinked themes. Limited patient involvement in movement between themes.

-1. Some movement between themes, but often unnatural in feel with little logical sequence. Logical gaps or inconsistencies. Very little patient involvement in navigating session themes.

-2. Preoccupied and disconnected focus on single themes without links with other components. Awkward or uncoordinated movement between themes. Largely therapist led with little patient cooperation.

References (Return to top)

Deen, D., Lu, W.-H., Rothstein, D., Santana, L., & Gold, M. R. (2011). Asking questions: the effect of a brief intervention in community health centers on patient activation. Patient Education and Counseling, 84(2), 257-260.

Derksen, F., Bensing, J., & Lagro-Janssen, A. (2013). Effectiveness of empathy in general practice: a systematic review. British Journal of General Practice, 63(606), e76-e84.

Greene, J., Hibbard, J. H., Sacks, R., Overton, V., & Parrotta, C. D. (2015). When patient activation levels change, health outcomes and costs change, too. Health Affairs, 34(3), 431-437.

Hibbard, J., Stockard, J., Mahoney, E. R., & Tusler, M. (2004). Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers. Health Services Research, 39(4 Pt 1), 1005-1026. doi:10.1111/j.1475-6773.2004.00269.x

Malins, S., Kai, J., Atha, C., Avery, A., Guo, B., James, M., . . . Morriss, R. K. (2016). Cognitive behaviour therapy for long-term frequent attenders in primary care: a feasibility case series and treatment development study. British Journal of General Practice, 66(651), e729-736.

Mercer, S. W., Gunn, J., Bower, P., Wyke, S., & Guthrie, B. (2012). Managing patients with mental and physical multimorbidity. British Medical Journal, 345, e5559.

Mosen, D. M., Schmittdiel, J., Hibbard, J., Sobel, D., Remmers, C., & Bellows, J. (2007). Is patient activation associated with outcomes of care for adults with chronic conditions? The Journal of ambulatory care management, 30(1), 21-29.

Moyers, T. B., Martin, T., Houck, J. M., Christopher, P. J., & Tonigan, J. S. (2009). From in-session behaviors to drinking outcomes: a causal chain for motivational interviewing. Journal of Consulting and Clinical Psychology, 77(6), 1113.

Patel, S., Kai, J., Atha, C., Avery, A., Guo, B., & Malins, S. (2015). Clinical characteristics of persistent frequent attenders in primary care: case-control study. . BMC Fam Pract, 32(6), 624-630.

Schmaderer, M. S., Zimmerman, L., Hertzog, M., Pozehl, B., & Paulman, A. (2016). Correlates of Patient Activation and Acute Care Utilization Among Multimorbid Patients. Western Journal of Nursing Research, 38(10), 1335-1353.

Appendix I: Theme and level summary sheet (Return to top)

Theme Name	Theme Summary	+2	-2
Structuring & Task Focus (STF)	Coded where significant changes in session direction take place. This includes changes of therapeutic importance or changes of direction initiated by drifting or deviating from a specified topic	Patient *leading or initiating* a relevant session focus, self-monitoring and redirecting the therapist if they deviate.	Lengthy avoidant deviations unaddressed or unabated OR Didactic, decision-making receiving little response
Information Discussion (IDI)	Code where new information is given or sought. The degree of relevance and reciprocity affect the level to be coded. When information is given that is personalised to the patient's situation PAU is coded.	Patient initiated request for specific information, relevant to therapeutic endeavour.	Relate to en elicits negative response OR Information given not relevant
Problem or Context Description (PCD)	Coded for all apparently neutral interactions where no other theme is appropriate to be applied. This will often constitute a problem account being given or sought.	N/A	N/A
Problem Analysis & Understanding (PAU)	Coded where a new therapeutic understanding is described by either party. The level coded is dependent on the degree of reciprocity and development expressed.	*Patient led* new understanding with links between problem components, where the solutions follow naturally.	Therapist or patient presents a formulation and this is not acknowledged, no alternative is given and no agreement is reached.
Noticing Change or Otherwise (NCO)	Coded where positive change (or absence thereof) is noticed and explicitly commented upon by either party. This would be linked to one or more of: a) The patient's efforts b) The therapist's efforts.	Patient notices a positive change, articulates it without prompting. May include: its impact on their life; sense of personal ownership, or control over the change.	Therapists initiated acknowledgement of change - no patient agreement OR patient refutes positive change.
Action Planning & Idea Generation (API)	Coded where either party suggests an action plan or potential problem solution idea linked to a future action. The level coded is dependent on how the idea is developed or refined particularly by the patient.	Patient initiated action plans or change ideas that are have potential to be applied effectively through refinement and discussion.	Therapist or patient initiated action plan or change idea – *ignored, refuted or distracted to another topic* and no agreement is reached.
Evaluations of Self or Therapy (EST)	Coded where there are expressions of commitment to carry out specified tasks; self-supporting statements, or where completion of a therapeutic task is acknowledged as an accomplishment. The level coded is dependent upon who initiates the comment and level of patient development and agreement.	Patient initiated expression of commitment to therapeutic tasks OR patient statements of self-efficacy OR Patient acknowledgement of therapeutic achievement.	Self-denigrating or self-critical statements OR a self-critical focus on therapeutic task that have not been completed to the exclusion of those that have.
Other (OTH)	Coded where none of the above codes can be applied, alongside a brief description of the segment.	N/A	N/A

Appendix II: Theme scoring sheet (Return to top)

Instructions:
Multiply the proportion of the transcript coded at each level of each theme by the score for the given level (TOTAL 1). For example, if 10% of a transcript was coded as Action Planning and Idea generation (API) at the +2 level, this would contribute a score of 20 to the total for the API theme (+2 x 10). Conversely, if 10% of the same transcript was scored as API -1, this would contribute a score of -10 to the total for the API theme (-1 x 10). If these were the only levels of API coded for the transcript a total score for API on that transcript would be 10 (20-10). To calculate overall content scores, average the scores for each theme (TOTAL 2). Global themes are scored separately with a single rating on the level scored.

Theme Name	-2	-1	0	+1	+2	TOTAL 1
Action Planning & Idea Generation (API)						
Evaluations of Self or Therapy (EST)						
Information Discussion (IDI)						
Noticing Change or Otherwise (NCO)						
Problem Analysis & Understanding (PAU)						
Problem or Context Description (PCD)						
Structuring & Task Focus (STF)						
TOTAL 2						
Integration of Themes (IOT)						
Collaborative Flow (COF)						

Helping Urgent Care Users Cope with Distress about Physical Complaints

Screening ID No:

Date of Interview:

Participant Initials:

Participant DOB:

Trial Identity Code (TIC) - Insert when randomised:

Interviewer:

**** Detach page 1-2 from rest of questionnaire and file separately

Gender: **Male** ☐ **Female** ☐		**Date of Birth:** **_____/_____/_____** **(DD / MM / YYYY)**	
Ethnicity: **White** **English/Welsh/Scottish/** ☐ **Northern Irish/British** ☐ **Irish** ☐ **Gypsy or Irish traveller** ☐ **Any other White background, please describe** ☐ **Mixed/multiple ethnic groups** ☐ **White and Black Caribbean** ☐ **White and Black African** ☐ **White and Asian** ☐ **Any other Mixed / Multiple ethnic background, please describe** **Other ethnic group** ☐ **Arab** ☐ **Any other ethnic group, please describe**		**Asian/Asian British** Indian ☐ Pakistani ☐ Bangladeshi ☐ Chinese ☐ Any other Asian background, please describe ☐ **Black / African / Caribbean / Black British** Black Caribbean ☐ Black African ☐ Black other ☐ Any other Black / African / Caribbean background, please describe ☐	
Highest Qualification: **Higher Degree** ☐ **First Degree** ☐ **Other Higher Qualification** ☐ **A-Level** ☐ **O-Level / GCS** ☐ **Other Qualification** ☐ **No Qualifications** ☐		**Marital Status:** Married ☐ Single ☐ Partner ☐ Divorced ☐ Widowed ☐	
Occupational Status: **Employed** ☐ **Self-employed** ☐ **Voluntary employment** ☐ **Student / Training** ☐ **Housewife/husband** ☐ **Unemployed** ☐ **Retired** ☐ **<u>Number of Hours Worked Each Week:</u>** **__Hours**		**IF Employed: Please state your current occupation** Manager/administrator ☐ Professional (e.g. health, teaching, legal) ☐ Clerical worker/secretary ☐ Services/sales ☐ Skilled labourer/craftsman (e.g. building, electrical) ☐ Factory worker ☐ Other (please specify) ☐	
Home address: **Contact numbers:** ***Home:*** ***Mobile:*** ***Work:*** ***Email address:***		**GP name and address:**	

Baseline Assessment

Screening ID No:

Date of Interview:

Participant Initials:

Participant DOB:

Trial Identity Code (TIC) - Insert when randomised:

Interviewer:

PHQ15

During the *past 4 weeks*, how much have you been bothered by any of the following problems?

		Not bothered at all	Bothered a little	Bothered a lot
a.	Stomach pain	☐	☐	☐
b.	Back pain	☐	☐	☐
c.	Pain in your arms, legs, or joints (knees, hips, etc.)	☐	☐	☐
d.	Menstrual cramps or other problems with your periods [Women only]	☐	☐	☐
e.	Headaches	☐	☐	☐
f.	Chest pain	☐	☐	☐
g.	Dizziness	☐	☐	☐
h.	Fainting spells	☐	☐	☐
i.	Feeling your heart pound or race	☐	☐	☐
j.	Shortness of breath	☐	☐	☐
k.	Pain or problems during sexual intercourse	☐	☐	☐
l.	Constipation, loose bowels, or diarrhoea	☐	☐	☐
m.	Nausea, gas, or indigestion	☐	☐	☐
n.	Feeling tired or having low energy	☐	☐	☐
o.	Trouble sleeping	☐	☐	☐

EQ-5D-5L

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

MOBILITY

I have no problems in walking about ❑
I have slight problems in walking about ❑
I have moderate problems in walking about ❑
I have severe problems in walking about ❑
I am unable to walk about ❑

SELF-CARE

I have no problems washing or dressing myself ❑
I have slight problems washing or dressing myself ❑
I have moderate problems washing or dressing myself ❑
I have severe problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities ❑
I have slight problems doing my usual activities ❑
I have moderate problems doing my usual activities ❑
I have severe problems doing my usual activities ❑
I am unable to do my usual activities ❑

PAIN / DISCOMFORT

I have no pain or discomfort ❑
I have slight pain or discomfort ❑
I have moderate pain or discomfort ❑
I have severe pain or discomfort ❑
I have extreme pain or discomfort ❑

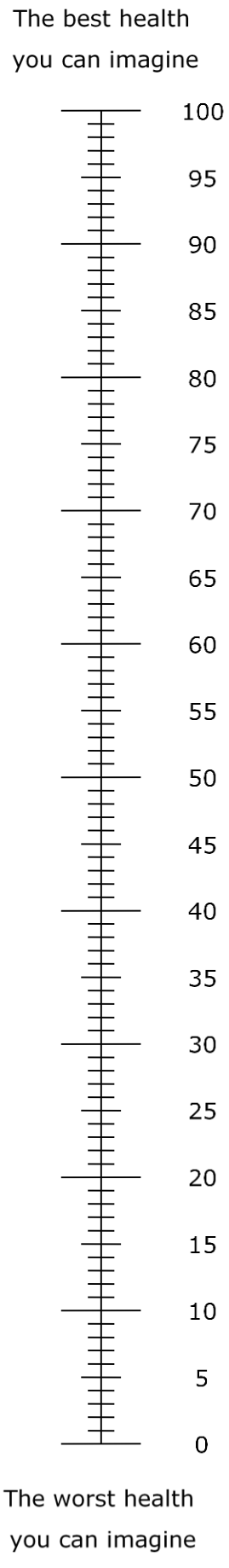
ANXIETY / DEPRESSION

I am not anxious or depressed ❑
I am slightly anxious or depressed ❑
I am moderately anxious or depressed ❑
I am severely anxious or depressed ❑
I am extremely anxious or depressed ❑

- We would like to know how good or bad your health is

  TODAY.
- This scale is numbered from 0 to 100.
- 100 means the <u>best</u> health you can imagine.
  0 means the <u>worst</u> health you can imagine.
- Mark an X on the scale to indicate how your health is TODAY.
- Now, please write the number you marked on the scale in the box below.

YOUR OWN HEALTH TODAY = ☐



PHQ9

Over the last 2 weeks, how often have you been bothered by any of the following problems:

		Not at all	Several days	More than half the days	Nearly every day
1.	Little interest or pleasure in doing things	0	1	2	3
2.	Feeling down, depressed, or hopeless	0	1	2	3
3.	Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4.	Feeling tired or having little energy	0	1	2	3
5.	Poor appetite or overeating	0	1	2	3
6.	Feeling bad about yourself – or that you are a failure or have let yourself or your family down.	0	1	2	3
7.	Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8.	Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.	0	1	2	3
9.	Thoughts that you would be better off dead or of hurting yourself in some way.	0	1	2	3

If you checked off <u>any</u> problems, how <u>difficult</u> have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not Difficult	Somewhat Difficult	Very Difficult	Extremely Difficult
☐	☐	☐	☐

GAD7

Over the last 2 weeks, how often have you been bothered by the following problems?

Please circle one response for each question

	Not at all	**Several days**	**More than half the days**	**Nearly every day**
1. Feeling nervous, anxious or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3
If you circled any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?				
Circle one	Not difficult at all	Some what difficult	Very difficult	Extremely difficult

WSAS

Rate each of the following questions on a 0 to 8 scale:

0	1	2	3	4	5	6	7	8
Not at all	Slightly		Definitely		Markedly		Very severely	

• **Because of my [symptoms/pain], my ability to work is impaired. 0 means not at all impaired and 8 means very severely impaired to the point I can't work**	
• **Because of my [symptoms/pain], my home management (cleaning, tidying, shopping, cooking, looking after home or children, paying bills) is impaired. 0 means not at all impaired and 8 means very severely impaired.**	
• **Because of my [symptoms/pain], my social leisure activities (with other people, such as parties, bars, clubs, outings, visits, dating, home entertainment) are impaired. 0 means not at all impaired and 8 means very severely impaired.**	
• **Because of my [symptoms/pain], my private leisure activities (done alone, such as reading, gardening, collecting, sewing, walking alone) are impaired. 0 means not at all impaired and 8 means very severely impaired.**	
• **Because of my [symptoms/pain], my ability to form and maintain close relationships with others, including those I live with, is impaired. 0 means not at all impaired and 8 means very severely impaired.**	

APPENDIX H: Participant information sheet for chapters 6, 7 and 8

Collaboration for Leadership in Applied Health Research and Care East Midlands

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

NHS National Institute for Health Research

Participant Information Sheet (Version 3.0 Date: 31.1.17)

Helping Urgent Care Users Cope with Distress about Physical Complaints

Name of Researchers: Shireen Patel and Michelle Stubley

We would like to invite you to take part in our research study

- Before you decide we would like you to understand why the research is being done and what it would involve for you.
- One of our team will go through the information sheet with you and answer any questions you have.
- Talk to others about the study if you wish.
- Ask us if there is anything that is not clear.

Contents

1. What is the purpose of the study?
2. Why have I been invited?
3. Do I have to take part?
4. What will happen to me if I take part?
5. Expenses and payment
6. What are the possible disadvantages and risks of taking part?
7. What are the possible benefits of taking part?
8. What happens when the research study stops?
9. What if there is a problem?
10. Will my taking part in the study be kept confidential?
11. What will happen to me if I don't want to carry on with the study?
12. Involvement of the General Practitioner (GP)
13. What will happen to the results of the research study?
14. Who is organising and funding the research?
15. Who has reviewed the study?
16. Further information and contact details

If you have any questions about this study, please talk to study researchers Shireen Patel/Michelle Stubley on **0115 8231434**

1 What is the purpose of the study?

(1) We want to arrive at a better understanding of the needs of patients who use unscheduled/urgent care.

- Each year many people make use of emergency services, such as going to the A & E department, a walk in centre or making an urgent same day appointment with their GP. This is termed "unscheduled/urgent care".
- Many patients have had to attend one of these services more than once in the past year but the doctors have been unable to tell them what exactly is causing their symptoms.
- These symptoms can be painful and cause distress if they are not able to be treated, yet repeated investigation may not establish the cause.

(2) We want to find out if providing CBT via video calling/over the telephone is acceptable and effective in improving physical and mental health.

- It may be possible to reduce the pain and discomfort using a talking therapy called Cognitive Behavioural Therapy (CBT) which can be delivered via a video calling system (similar to Skype) or the telephone.
- CBT aims to teach you how best to cope when you get distressed. It will also help you understand your body's responses better and how you can manage them.
- To find out whether CBT is effective we will carry out a "randomised controlled trial" in which we make comparisons between two groups. Half of the people we interview will be offered CBT as well as their usual care and the other half will continue to receive their usual care only. For statistical reasons, each patient is put into a group by chance (randomly).

2 Why have I been invited?

You are being invited to take part because you have accessed unscheduled/urgent care more than once in the last year. We are inviting 144 participants like you to take part.

3 Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. This would not affect your legal rights.

4 What will happen to me if I take part?

- A member of staff from the hospital, GP Practice or out of hours service you recently visited will introduce the study to you. They will tell you more about the study and answer any questions you may have. If you agree to be contacted by us, we will call you to see whether the study will be suitable for you, this will take no longer than fifteen minutes. Based on the answers you provide we may then ask to meet with you.
- If we contact you again, we will ask you to meet with a member of the research team at a time and place convenient to you or ask to carry out an interview over the phone, whichever approach is most suitable to you. If you are happy for us to meet or contact you we will go through the study in more detail. We will answer any questions you may have and ask for your written consent to take part in the study. We will then ask you some questions about your physical and mental health and your use of health services.
- The interview is expected to last around one and a half hours. We will only need to meet with you once for this, if you are randomly selected to receive CBT we will contact you again and ask if you would like to participate in the therapy part of the study.
- If you are randomised to the treatment arm of the study you will be contacted by a CBT therapist (trained in providing psychological therapy but is not a medical doctor) who will arrange for you to receive between 6-12 sessions of CBT either via video calling or over the telephone. During CBT sessions, you and the therapist will identify how your thoughts, feelings and actions affect the symptoms you are experiencing.
- You will then work together to find ways of helping you cope with the physical symptoms and any linked emotional difficulties. As part of this process, you will be asked to test and practise strategies between sessions.
- We will ask some people who take part in the initial interview whether they would like to be interviewed again. This interview will allow you to talk in more detail about your thoughts about the CBT. The interview will take place within 18 months of your initial interview.
- We will also ask all participants to complete a set of questionnaires three, six, nine and twelve months after they have been interviewed. These can be completed over the phone or sent to you by post or email, whichever is most suitable to you. The questions will be similar to the ones we ask you at your interview and will ask about your physical and mental health and use of health services so that we can compare any changes. They will take around 45 minutes to complete.
- Participation in this research, does not affect your rights to consult your GP or seek treatment from other health care providers. If you have any concerns about your symptoms you should continue to see your GP and other health care providers you normally consult.
- With your consent, meetings with the CBT therapist may be audio/video recorded to help maintain the quality of the treatment and better understand how we can help you. We may also offer a recording of the session to you so that you can listen to and reflect on previous sessions. You are free to decide not to have the sessions recorded and this would not affect the care you receive in any way.
- We may use anonymised quotes in reports or training materials arising from this study.
- After completing CBT sessions you will be invited to use the Florence texting system to help maintain any benefits from CBT. In your final CBT sessions your therapist will help you identify key ideas that you would like to remember in future. These can then be part of a plan to help you stay well. The Florence system will text you reminders of the ideas that you have produced, that may help you feel better for longer. For example, when if you were feeling worried you may write a reminder for yourself:

- "Remember that you don't need to avoid anxiety, it passes in time." Florence will text this, or other ideas, to you when you tell the system you are worried. If you choose to use the system you will receive weekly text reminders for up to six months. You can stop using the system at any time by texting STOP in reply to a message. If you use Florence you will be asked to complete a survey about the experience.
- Your responses to the research questions will be anonymised and stored safely according to clinical trial regulations (see below).
- If you decide to take part in this study you will be given a copy of this information sheet to keep and a copy of the signed consent form.
- By taking part you agree for a member of the research team to examine your medical notes. Your notes will not be taken from where they are usually kept and any information taken from them will remain confidential and will be anonymised so that you cannot be identified

5 Expenses and payments

To thank you for your time, you will be given a £20 High Street Voucher, £10 High Street Voucher at the end of the interview and a £5 voucher each time you complete a questionnaire.. You can also claim travel expenses for any visits incurred as a result of participation. We will always contact you for the CBT sessions and research assessments so that you incur no extra cost for participating in this study.

6 What are the possible disadvantages and risks of taking part?

Some of the questions we will be asking will enquire about symptoms including emotions such as feeling anxious or low. Whilst most people do not mind answering these questions, some people may feel upset. It is important that we ask these questions and find out if treatment can improve these symptoms. Many people find talking about or sharing concerns in a safe and confidential way can be helpful.

7 What are the possible benefits of taking part?

By taking part you may receive treatment which helps with your symptoms. The questionnaires and interview you complete as part of the research will allow you to reflect on your symptoms and emotions and how these have changed over the research period.

We cannot promise the study will help you but the information we get from this study may help patients in the future to get treatment that helps them manage their difficulties, and cope better with their pain or associated symptoms.

8 What happens when the research study stops?

Your care will not change as a result of taking part in the study. If you are interested in finding out the results of the study we can keep your contact details and inform you of results when the study is completed.

9 What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the researchers or the Clinical Trials Manager who will do their best to answer your questions. Their contact details are given at the end of this information sheet. If you remain unhappy you may contact the Patient Advice Liaison Service (PALS) on 0115 924 9924 ext: 63187 and if you wish to complain formally, you can do this by contacting NHS Complaints. Details can be obtained from your hospital.

10 Will my taking part be kept confidential?

- We will follow ethical and legal practice and all information about you will be handled in confidence at all times.

- We will follow ethical and legal practice and all information about you will be handled in confidence at all times.
- If you join the study, some parts of your medical records and the data collected for the study will be looked at by authorised persons from the University of Nottingham who are organising the research. Data may also be looked at by authorised people to check that the study is being carried out correctly. All will have a duty of confidentiality to you as a research participant and we will do our best to meet this duty.
- All information which is collected about you during the course of the research will be kept **strictly confidential** stored in a secure and locked office, and on a password protected database**.** Any information about you which leaves the hospital or GP practice will have your name and address removed (anonymised) and a unique code will be used so that you cannot be recognised from it.
- The only time that we would break confidentiality is if we felt that we need to share information to protect your safety or the safety of others.
- Your personal data (address, telephone number) will be kept for twelve months after the end of the study so that we are able to contact you about the findings of the study and possible follow-up studies (unless you advise us that you do not wish to be contacted). All other data (research data) will be kept securely for seven years in accordance with the University of Nottingham regulations. After this time your data will be disposed of securely. During this time all precautions will be taken by all those involved to maintain your confidentiality, only members of the research team will have access to your personal data.

11 What will happen to me if I don't want to carry on with the study?

Your participation is voluntary and you are free to withdraw from the study at any time, without giving any reason, and without your legal rights being affected. If you decide to withdraw from the study, any information collected so far cannot be erased and this information may still be used in the project analysis.

12 Involvement of the General Practitioner/Family

With your written consent, we will send a letter to your usual doctor informing them that you are taking part in this study.

13 What will happen to the results of the research study?

The results will be publicised through the extensive arrangements for dissemination locally within the University and local NHS services (through road shows, websites, and conferences) as well as publication in peer reviewed journals, local, national and international scientific conferences. We will also send all participants a summary of the findings. The results will be published at the end of the study which is expected to be the end of 2018.

14 Who is organising and funding the research?

This study is funded by National Institute for Health Research (NIHR) and matched funding from local partner organisations as part of a research and implementation centre called CLAHRC East Midlands, one of 13 such research centres in the country. The Chief Investigator and researchers you will meet are based at the University of Nottingham.

15 Who has reviewed the study?

All research in the NHS is looked at by an independent group of people called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given favourable opinion by National Research Ethics Service (NRES) London-Riverside Committee.

16. Further information and contact details

If you have any queries or would like to talk more about the study you can contact:

Study researchers

Shireen Patel/Michelle Stubley
Tel: 0115 82314 34/25
CLAHRC EM
IMH Building
University of Nottingham
Nottingham
NG7 2TU

Clinical Trials Manager

Dr Catherine Kaylor-Hughes
Tel: 0115 8232478
CLAHRC EM
IMH Building
University of Nottingham
Nottingham
NG7 2TU

Chief Investigator

Professor Richard Morriss
Tel: 0115 8230427
CLAHRC EM
IMH Building
University of Nottingham
Nottingham
NG7 2TU

APPENDIX I: CBT information sheet

Collaboration for Leadership in Applied Health Research and Care East Midlands

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

NHS National Institute for Health Research

CBT Information Sheet

Helping Urgent Care Users Cope with Distress about Physical Complaints

We would like you to receive remotely delivered CBT treatment

- We would like you to understand why the research is being done and what it will involve for you.
- Before you decide whether to take part, one of our CBT therapists can go through the information sheet with you and answer any questions you may have.
- Do talk to others about the study if you wish.
- Please ask us if there is anything that is not clear.

Contents

1. What is CBT?
2. Why have I been selected to receive the CBT treatment?
3. What will my taking part involve?
4. How will CBT be delivered via video calling or telephone?
5. Is receiving CBT in this way secure?
6. Can I choose how the CBT sessions are delivered?
7. How many CBT sessions will I receive?
8. Will my sessions be audio or video recorded?
9. Who would listen to the session recordings?
10. How will my session recordings be kept?
11. What if I don't want my sessions to be recorded?
12. Will there be a change in my usual care?
13. What happens if I want to stop the sessions?
14. What happens next?
15. Do I have to take part?
16. Who do I contact for further information?

1 What is CBT?

Cognitive Behavioural Therapy (CBT) is a type of talking therapy based on the idea that how people feel physically is affected by the way they think, emotions they experience and what they do.

2 Why have I been selected to receive the CBT treatment?

To find out whether CBT is effective we are carrying out a Randomised Controlled Trial (RCT) to compare the two treatment groups. Half of the people we interview will continue to receive their usual care and the other group will be offered CBT as well as their usual care. Each participant is put into a group by chance (randomly). You have been placed in the group that will receive CBT sessions via a video calling system (similar to Skype) or the telephone.

3 What will my taking part involve?

During CBT sessions, you and the therapist will identify how your thoughts, feelings and actions affect the symptoms you are experiencing. You will then work together to find ways of helping you cope with the physical symptoms and any linked emotional difficulties.

If you have any questions about this study, please talk to Julie Moss on **0115 8231439** or Dr Catherine Kaylor-Hughes on **0115 8232478**

As part of this process, you will be asked to test and practise strategies between sessions.
Previous research suggests that CBT may be able to support you to develop ways of managing your symptoms and cope with the distress and pain that your symptoms are causing

4 How will CBT be delivered?

In both video calling and telephone CBT the therapist will ask that you are able to find a quiet place where you will not be disturbed or overheard by others. It is also helpful to find a place where you have good telephone reception or internet Wi-Fi.

Video calling

The therapist will call you and explain how video calling will be set up. Your therapist will then set up a video call with you. Once this is set up the therapist will contact you via a video calling system similar to Skype at a time chosen by you. If you choose to, the therapist can text/email you 24 hours before your session to remind you of your appointment. The therapist will always contact you. You will need to have an internet connection. You may set up a password or some security questions with the therapist if you prefer. The session will normally last one hour. Both you and the therapist will be able to see each other throughout the session. No previous experience of using video calling facilities is required. This facility gives you the opportunity to have as close to "face-to-face" contact as possible.

Telephone

The therapist will contact you on the telephone number that you provide at a time chosen by you. If you choose to, the therapist can text/email you 24 hours before your session to remind you of your appointment. The therapist will always call you and you will not be charged for any of the sessions.

You may set up a password or some security questions with the therapist if you prefer. The session will normally last one hour.

5 Is receiving CBT in this way secure?

Yes, the system we use for video calling will adhere to the strictest security standards and the data protection policy. In addition the system will be user friendly with minimal downloading of software for you.

6 Can I choose how the CBT sessions are delivered?

Yes, it is entirely up to you whether you would like to receive CBT via video calling or over the telephone. The number of sessions and the content will remain the same regardless of how you receive the therapy. The main difference will be that over the telephone you will not be able to see the therapist.

7 How many CBT sessions will I receive?

You will receive between 6-10 sessions of CBT.

8 Will my sessions be audio or video recorded?

With your consent the therapist may record your session. Prior to recording they will always inform you that the recording will start. If you are receiving CBT over the telephone the session will be recorded using an audio-recording device. If the CBT is via video calling the session will be recorded using a recording device on the computer. The sessions will be recorded to maintain and enhance the quality of the treatment. Session recordings will also be available for you to listen to between sessions, which can be a helpful reminder of what was discussed and planned. Session recordings will also be analysed by the research team to establish the best therapeutic approaches and recordings can inform training of clinical staff.

9 Who would listen to the session recordings?

We will ask for your consent to use session recordings for the following purposes:

- To be listened to by your therapist in supervision with other study therapists.
- To be given to you for additional therapeutic benefits.
- To be listened to by study researchers in an analysis of therapeutic sessions.
- For extracts to be reviewed in training for clinical staff.

10 How will my session recordings be kept?

After the session the therapist will transfer your recording onto the computer. All recordings will be anonymised and stored securely in password protected storage areas. Only authorised members of the study team will have access to the recordings. Original recordings will be destroyed at the same time as all other study documentation.

11 What if I don't want my sessions to be recorded?

You are free to decide not to have the sessions recorded and this would not affect the care you receive in any way. You can also withdraw consent at any time until the end of the study. You can still withdraw consent for any session even if you have consented to audio-visual recordings without having to give a reason until the end of the study.

12 Will there be a change in my usual care?

No, your participation in receiving CBT, does not affect your rights to consult your GP or seek treatment from other health care providers.

13 What happens if I want to stop the sessions?

Some people may find that CBT is not for them. Sometimes it may take a number of sessions before you notice any benefits. Whilst we hope that CBT will support you to develop ways of managing your symptoms and cope with distress this cannot be guaranteed. If you do not feel that CBT is helping you please talk to your therapist. If you decide to stop your CBT sessions, this will not affect your usual care in any way.

It also does not affect your participation in the study. The researchers will continue to contact you for follow up assessments at 3, 6, 9 and 12 months.

If you no longer wish to take part in the study at any point and do not want to be contacted by the researchers please let us know by contacting the research team on the contact numbers given below.

14 What happens next?

We would like you to have some time to think about whether or not you would like to receive CBT. Please talk to others if you wish or contact the study team if you have any further questions. A CBT therapist will contact you within the next 10 days to introduce themselves and answer any additional questions you may have.

15 Do I have to take part?

No, it is entirely up to you to decide whether or not to take part. If you decide not to take part it will not affect the standard of care you normally receive. You are free to withdraw at any time and without giving a reason.

17. Further information and contact details

If you would like further information or have any questions or if you decide that you would not like to be contacted by a CBT therapist please contact:

Administrative Support Officer

Julie Moss
Tel: 0115 8231439
CLAHRC EM
IMH Building
University of Nottingham
Nottingham
NG7 2TU

Clinical Trials Manager

Dr Catherine Kaylor-Hughes
Tel: 0115 8232478
CLAHRC EM
IMH Building
University of Nottingham
Nottingham
NG7 2TU



APPENDIX J: Patient-participant consent form

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Collaboration for Leadership in Applied Health Research and Care East Midlands

NHS National Institute for Health Research

Local letter head to be added

TIC:

PARTICIPANT CONSENT FORM
(Version 1.1)

Helping Urgent Care Users Cope with Distress about Physical Complaints

REC ref:

Name of Chief Investigator: Professor Richard Morriss

Name of Participant:

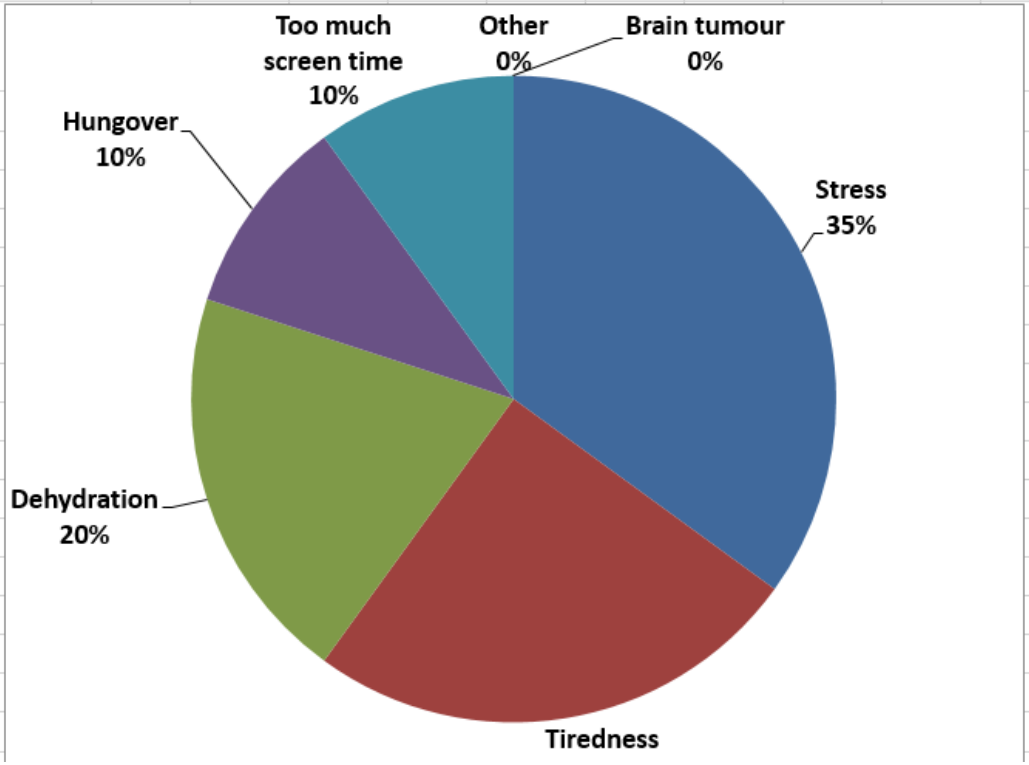
	Please Initial box
1. I confirm I have read and understand the information sheet (Participant Information Sheet Version 1.1 dated 23.7.14) for the above study and have had the opportunity to ask questions.	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected. I understand that should I withdraw then the information collected so far cannot be erased and that this information may still be used in the project analysis.	☐
3. I understand that relevant sections of my medical notes and data collected in the study may be looked at by authorised individuals from the University of Nottingham, the research group and regulatory authorities where it is relevant to my taking part in this study. I give permission for these individuals to have access to these records and to collect, store, analyse and publish information obtained from my participation in this study.	☐
4. I agree to my GP being informed of my participation in this study.	☐
5. I give permission for my treatment sessions to be audio/video recorded. I understand that any quotes used in publications, reports or training materials will be anonymous and I will not be identified.	☐
6. I agree to being contacted in the event that the research team wish to carry out a further interview.	☐
7. I understand that if I agree to a further interview it will be recorded and transcribed and that anonymous direct quotes from the interview may be used in the study reports.	☐
8. I agree to take part in the above study.	☐

_____	_____	_____
Name of Participant	**Date**	**Signature**
_____	_____	_____
Name of Researcher	**Date**	**Signature**

3 copies: 1 for participant, 1 for the project notes and 1 for the medical notes

APPENDIX K: Cause pie chart example

Headache	
Cause	**Percentage of people**
Stress	35
Tiredness	25
Dehydration	20
Hungover	10
Too much screen time	10
Brain tumour	?
Other	0
Subtotal	100



APPENDIX L: Transcriber confidentiality agreement

TRANSCRIBER CONFIDENTIALITY AGREEMENT

Data Protection Act 1998 Confidentiality Agreement for Transcribers

This Agreement is made as of 20/10/17 (Date), by and between the University of Nottingham, with principal offices at Jubilee Campus, Wollaton Rd, Nottingham NG8 1BB (the University) and Shirley Woolley with principal offices at 10 Bentley Avenue Bakersfield, Notts NG3 7Ax (the Transcriber).

The Transcriber has been appointed by the University of Nottingham to transcribe audiovisual/audio files and documentation resulting from the Helping Urgent Care Users Cope With Distress About Physical Complaints study (Principle Investigator Richard Morriss) which will involve the disclosure to the Transcriber of personal data held by the University. Accordingly the Transcriber is required to deal with any such information in accordance with the terms of this Agreement and the Data Protection Act 1998.

The Transcriber undertakes to respect and preserve the confidentiality of personal data. Accordingly, for an indefinite period after the date of this Agreement the Contractor shall:

- maintain the personal data in strict confidence and shall not disclose any of the personal data to any third party;
- restrict access to employees, agents or sub-contractors who need such access for the purposes of the contract (and then only if the employee, agent or subcontractor is bound by conditions of confidentiality no less strict than those set out in this agreement, which the Transcriber shall enforce at the University's request);
- ensure that its employees, agents or sub-contractors are aware of and comply with the Data Protection Act 1998; and
- not authorise any sub-contractor to have access to the personal data without obtaining the University's prior written consent to the appointment of such sub-contractor and entering into a written agreement with the subcontractor including conditions of confidentiality no less strict than those set out in this agreement, which the Transcriber shall enforce at the University's request.

The Transcriber agrees to indemnify and keep indemnified and defend at its own expense the University against all costs, claims, damages or expenses incurred by the University or for which the University may become liable due to any failure by the Transcriber, its employees, agents or sub-contractors to comply with any of its obligations under this Agreement.

For the avoidance of doubt, the confidentiality imposed on the Transcriber by this Agreement shall continue in full force and effect after the expiry or termination of any contract to supply services.

The restrictions contained in this Agreement shall cease to apply to any information which may come into the public domain otherwise than through unauthorised disclosure by the Transcriber.

This Agreement shall be governed by and construed in accordance with the laws of England and the parties hereby submit to the exclusive jurisdiction of the English courts.

Signed for and on behalf of SHIRLEY WOOLLEY

Signed: S. Woolley Name: SHIRLEY WOOLLEY

Title: MRS Date: 20/10/17

Signed for and on behalf of the University of Nottingham

Signed: [signature] Name: SAM MALINS

Title: CLINICAL PSYCHOLOGIST Date: 30/10/17

Recd 23/10/17.

APPENDIX M: Notes from coder discussion

All Themes

- Strong endorsement can move IDI and all other themes from 0 to +1 (e.g. Therapist: “You’ve got to face your fears to overcome anxiety” Patient: “I completely agree, definitely.”)
- If we find ourselves doing a lot of interpreting or unsure about whether the sentiment is strong enough to code a particular theme, then err towards coding PCD. PCD is the default unless clear evidence to code otherwise.
- If there is a “hanging” turn after a theme has been coded this is just added to the previously coded theme.
- Code each turn in its own right if it is clear what to coder, if unclear use the context around the turns (including audio or video) to help decide what to code.
- All “grace turns” are removed, if PAU +1 deviates to PAU 0 the level is changed to PAU 0. If API +1 deviates to PCD the theme is changed. No allowance is given for these types of change
- For there to be a step up from +1 to +2 there needs to have been a deviation from the topic (typically by the therapist) to allow the patient to *initiate* a new segment in its own right as +2
- Large turns of speech may require breaking up into different themes within the turn. The level is defined by the response received.

API

- This is strictly referring to actions or ideas to take place *outside* the session. Reference to actions taking place later in the session coded as STF.

STF

- Minor agreement to discuss something later in the session is STF +1 and does not require six turns (e.g. Therapist: Lets talk about how you can deal with anxiety at work later in the session; Patient: Yes, ok.)
- The standard most common coding is 6 turns led by the therapist
- However it also includes when a therapist led activity is being guided (e.g. Therapist: “now close your eyes and picture a stream” Patient: “I love streams” = STF +1)
- STF +2 ratings do not need to be 6 turns either and can be structural promptings that receive a positive response from the therapist

IGI

- The decision between IDI and PAU remains based on pronouns used (i.e. IDI= we, people, everyone, you plural; PAU=you singular, examples from patient’s experience etc)
- If there is mixed use of pronouns or it is unclear whether individualised or general use IDI
- Think of IDI as something that could be delivered in a presentation to an audience, i.e. no individualised content.

NCO

- There must be a reference to before/after change even if implied

PAU

- This refers only to analysis that directly relates to the patient. So if an analysis of someone else's behaviour is discussed, this must be linked to how it affects the patients responses. Otherwise coded as PCD.

OTH

- Technical workings out are coded OTH (e.g. Therapist: "So if you just click the top corner you can view this in full-screen mode." Patient: "Errm, no it's still not working…")
- Code any actions transcribed without words as OTH.

APPENDIX N: Chapter 6 bootstrapped CICS MLM results

Table 1. Multilevel model of noticing change interactions at session two predicting all health outcomes across time. Confidence intervals estimated with 1,000 bootstrapped samples

Health anxiety (SHAI)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	11.48	-0.36	8.47	.008	1.30	19.82
Assessment time point	-1.56	0.00	0.27	.001	-2.14	-0.99
Baseline severity (SHAI)	0.56	0.00	0.15	.001	0.40	0.74
Therapeutic alliance at session 2 (WAI-O)	-0.87	0.06	1.14	.101	-1.81	0.42
Noticing change at session 2 (CICS)	*-0.40*	*-0.03*	*0.20*	*.001*	*-0.70*	*-0.24*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	13.61	-4.19	2.41	.150	5.27	14.56
Intercept variance	23.77	4.00	3.97	.001	20.08	35.96

Somatic symptoms (PHQ-15)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	8.13	-0.29	2.51	.002	3.00	12.96
Assessment time point	-0.91	0.03	0.24	.003	-1.37	-0.40
Baseline severity (PHQ-15)	0.61	0.01	0.05	.001	0.52	0.71
Therapeutic alliance at session 2 (WAI-O)	-0.52	0.04	0.40	.151	-1.32	0.26
Noticing change at session 2 (CICS)	*-0.35*	*-0.05*	*0.20*	*.020*	*-0.75*	*-0.13*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	8.22	-2.77	0.89	.089	3.78	7.29
Intercept variance	8.43	2.42	2.37	.001	6.22	15.43

Generalised anxiety (GAD-7)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.75	-0.02	3.05	.038	-0.28	11.55
Assessment time point	-1.02	0.01	0.26	.001	-1.52	-0.50
Baseline severity (GAD-7)	0.41	0.00	0.06	.001	0.30	0.52
Therapeutic alliance at session 2 (WAI-O)	-0.18	0.00	0.47	.683	-1.08	0.75
Noticing change at session 2 (CICS)	*-0.50*	*-0.01*	*0.07*	*.001*	*-0.68*	*-0.38*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	9.73	-3.01	1.30	.080	4.30	9.44
Intercept variance	10.46	2.80	1.88	.001	9.71	17.20

Depression (PHQ-9)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.68	-0.67	2.99	.042	-0.92	11.03
Assessment time point	-0.56	0.00	0.26	.044	-1.11	-0.05
Baseline severity (PHQ-9)	0.49	0.01	0.05	.001	0.40	0.59
Therapeutic alliance at session 2 (WAI-O)	-0.19	0.11	0.47	.696	-1.01	0.89
Noticing change at session 2 (CICS)	*-0.51*	*-0.07*	*0.16*	*.001*	*-0.91*	*-0.35*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	9.36	-2.83	1.36	.102	3.99	9.41
Intercept variance	14.97	2.24	2.03	.001	13.28	21.23

Functional impairment (WSAS)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	6.48	0.21	4.25	.082	-2.13	14.51
Assessment time point	-1.43	0.03	0.42	.008	-2.24	-0.64
Baseline severity (WSAS)	0.73	0.00	0.04	.001	0.65	0.80
Therapeutic alliance at session 2 (WAI-O)	-0.62	-0.02	0.74	.341	-2.00	0.89
Noticing change at session 2 (CICS)	*-0.85*	*-0.06*	*0.17*	*.001*	*-1.30*	*-0.64*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	24.53	-7.56	5.09	.207	8.10	27.34
Intercept variance	33.81	6.38	5.90	.001	29.38	53.00

General health (VAS)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	53.05	-1.09	8.80	.001	34.80	69.34
Assessment time point	0.62	0.02	0.89	.520	-1.11	2.32
Baseline health (VAS)	0.37	-0.01	0.05	.001	0.25	0.45
Therapeutic alliance at session 2 (WAI-O)	-2.30	0.24	1.49	.094	-4.98	1.03
Noticing change at session 2 (CICS)	*1.77*	*0.14*	*0.42*	*.001*	*1.32*	*2.71*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	108.65	-33.77	17.38	.122	43.36	111.90
Intercept variance	167.21	32.14	24.60	.001	153.67	251.15

Quality of life (EQ5D-5L)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	0.313	0.001	0.110	.001	0.102	0.548
Assessment time point	0.024	0.000	0.012	.052	0.002	0.048
Baseline QoL (EQ5D-5L)	0.636	0.004	0.071	.001	0.497	0.794
Therapeutic alliance at session 2 (WAI-O)	-0.024	-0.002	0.020	.188	-0.069	0.010
Noticing change at session 2 (CICS)	*0.016*	*0.002*	*0.005*	*.001*	*0.012*	*0.030*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	0.017	-0.006	0.003	.148	0.006	0.017
Intercept variance	0.026	0.005	0.005	.001	0.022	0.041

Note. Abbreviations: SHAI: Short health anxiety inventory; GAD-7: Generalised anxiety disorder – 7 items; PHQ-9: patient health questionnaire – 9 items; PHQ-15: Patient health questionnaire – 15 items; WSAS: Work and social adjustment scale; VAS: Visual analogue scale; EQ5D-5L: EuroQol – 5 dimensions – 5 levels; QoL: Quality of Life; B: unstandardized β coefficient; SE: standard error; p: significance p value with an alpha of .05; CI: confidence interval; WAI-O: working alliance inventory - observer rated; CICS: consultation interactions coding scheme.

APPENDIX O: Chapter 8 bootstrapped ORS MLM results

Table 1. Predictive validity of patient wellbeing at the last treatment session on follow-up outcomes over 3, 6, 9 and 12-month follow-ups, controlling for baseline severity, comorbidities and changes over time. Confidence intervals estimated using 1,000 bootstrapped samples

Health anxiety (SHAI)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	22.53	-0.19	2.55	.001	17.16	27.28
Assessment time point	-1.53	-0.01	0.28	.001	-2.11	-1.01
Baseline severity (SHAI)	0.22	0.01	0.08	.002	0.07	0.38
Session 1 wellbeing (ORS)	-0.08	0.00	0.05	.046	-0.18	0.00
Total morbidities	0.44	-0.01	0.13	.001	0.17	0.72
Last session wellbeing (ORS)	*-0.34*	*0.01*	*0.04*	*.001*	*-0.40*	*-0.26*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	13.59	-4.11	2.22	.128	5.44	13.92
Intercept variance	15.48	3.91	3.09	.001	13.65	26.23

Somatic symptoms (PHQ-15)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	13.21	-0.16	1.85	.001	9.44	16.85
Assessment time point	-0.76	0.00	0.22	.003	-1.22	-0.32
Baseline severity (PHQ-15)	0.52	0.01	0.05	.001	0.42	0.63
Session 1 wellbeing (ORS)	-0.11	-0.01	0.06	.031	-0.22	0.00
Total morbidities	0.00	0.00	0.07	.962	-0.12	0.14
Last session wellbeing (ORS)	*-0.19*	*0.01*	*0.04*	*.001*	*-0.27*	*-0.09*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	7.84	-2.56	0.84	.058	3.75	6.99
Intercept variance	9.72	2.31	2.02	.001	8.24	16.12

Depression (PHQ-9)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	17.83	0.18	2.02	.001	14.22	21.92
Assessment time point	-0.44	-0.01	0.26	.112	-0.94	0.07
Baseline severity (PHQ-9)	0.32	0.00	0.04	.001	0.24	0.41
Session 1 wellbeing (ORS)	-0.13	0.00	0.06	.021	-0.25	-0.01
Total morbidities	0.00	-0.01	0.07	.981	-0.15	0.12
Last session wellbeing (ORS)	*-0.34*	*0.00*	*0.04*	*.001*	*-0.42*	*-0.25*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	10.43	-3.12	1.31	.054	4.86	9.89
Intercept variance	10.81	2.75	2.02	.001	9.40	17.54

Generalised anxiety (GAD-7)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	16.27	0.14	1.87	.001	12.72	20.13
Assessment time point	-0.90	-0.02	0.25	.002	-1.41	-0.42
Baseline severity (GAD-7)	0.15	0.00	0.06	.003	0.05	0.27
Session 1 wellbeing (ORS)	-0.11	0.00	0.04	.003	-0.20	-0.03
Total morbidities	0.30	0.00	0.07	.001	0.15	0.44
Last session wellbeing (ORS)	*-0.33*	*0.00*	*0.04*	*.001*	*-0.41*	*-0.25*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	10.63	-3.40	1.31	.063	4.78	9.93
Intercept variance	5.12	3.13	1.74	.001	5.05	11.83

Functional impairment (WSAS)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	17.30	0.64	3.54	.001	11.63	24.98
Assessment time point	-1.01	-0.07	0.42	.032	-1.99	-0.25
Baseline severity (WSAS)	0.59	0.00	0.04	.001	0.50	0.67
Session 1 wellbeing (ORS)	-0.06	0.00	0.09	.469	-0.23	0.12
Total morbidities	0.02	-0.01	0.14	.830	-0.26	0.30
Last session wellbeing (ORS)	*-0.46*	*-0.01*	*0.09*	*.001*	*-0.66*	*-0.31*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	30.65	-9.27	5.49	.154	11.99	32.97
Intercept variance	32.82	8.52	6.33	.001	29.54	53.84

General health (VAS)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	22.46	-1.71	6.32	.001	7.88	32.25
Assessment time point	1.41	0.03	0.71	.068	0.02	2.86
Baseline health (VAS)	0.21	0.00	0.05	.001	0.12	0.31
Session 1 wellbeing (ORS)	0.47	0.02	0.18	.002	0.16	0.85
Total morbidities	-0.33	0.04	0.32	.240	-0.89	0.38
Last session wellbeing (ORS)	*0.73*	*0.02*	*0.20*	*.001*	*0.37*	*1.16*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	104.10	-32.54	18.40	.154	40.40	112.91
Intercept variance	180.27	30.44	29.63	.001	154.84	271.54

Quality of life (EQ5D-5L)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	-0.018	-0.013	0.065	.761	-0.160	0.095
Assessment time point	0.019	0.001	0.010	.095	0.000	0.042
Baseline QoL (EQ5D-5L)	0.643	-0.014	0.084	.001	0.464	0.792
Session 1 wellbeing (ORS)	-0.002	0.000	0.002	.332	-0.007	0.003
Total morbidities	0.002	0.000	0.004	.601	-0.005	0.010
Last session wellbeing (ORS)	*0.008*	*0.000*	*0.002*	*.001*	*0.005*	*0.013*

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	0.015	-0.005	0.003	.183	0.005	0.015
Intercept variance	0.025	0.004	0.004	.001	0.021	0.038

Note. Abbreviations: B: unstandardised β coefficient; SE: standard error; p: significance at an alpha of .05; CI: confidence interval; ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; VAS: visual analogue scale; EQ5D-5L: Euroqol –5 dimensions –5 levels; QoL: Quality of life.

Table 2. MLMs showing the difference in outcome trajectory between responders experiencing sudden gains and responders who did not.

Health anxiety (SHAI)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	6.46	0.08	2.25	.006	1.68	10.62
Assessment time point	-1.65	0.00	0.26	.001	-2.21	-1.16
Baseline severity (SHAI)	0.45	0.00	0.10	.001	0.28	0.70
Total morbidities	0.23	0.01	0.15	.090	-0.05	0.52
Sudden gainers	-1.82	-0.09	0.78	.028	-3.68	-0.46

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	13.47	-4.24	2.14	.132	5.28	13.71
Intercept variance	22.72	3.93	3.91	.001	18.54	34.49

Somatic symptoms (PHQ-15)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	5.38	0.07	1.05	.001	3.42	7.56
Assessment time point	-0.82	-0.01	0.21	.001	-1.24	-0.39
Baseline severity (PHQ-15)	0.57	0.00	0.06	.001	0.44	0.68
Total morbidities	0.03	0.00	0.07	.683	-0.11	0.17
Sudden gainers	-1.67	0.01	0.57	.003	-2.78	-0.52

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	7.94	-2.53	0.85	.053	3.85	7.13
Intercept variance	12.76	2.25	2.00	.001	11.07	18.93

Depression (PHQ-9)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	4.27	0.01	0.94	.001	2.45	6.28
Assessment time point	-0.67	-0.01	0.26	.025	-1.17	-0.14
Baseline severity (PHQ-9)	0.41	0.00	0.05	.001	0.31	0.51
Total morbidities	0.13	-0.01	0.07	.043	-0.02	0.26
Sudden gainers	-2.54	0.06	0.61	.001	-3.72	-1.28

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	11.28	-3.43	1.35	.051	5.21	10.57
Intercept variance	15.17	3.07	2.20	.001	14.22	23.08

Generalised anxiety (GAD-7)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Intercept	3.85	0.01	0.82	.001	2.36	5.57
Assessment time point	-0.88	-0.03	0.26	.002	-1.42	-0.44
Baseline severity (GAD-7)	0.28	0.01	0.06	.001	0.17	0.41
Total morbidities	0.25	0.00	0.08	.001	0.10	0.41
Sudden gainers	-2.01	-0.06	0.61	.001	-3.32	-0.88

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI Lower Bound	95% CI Upper Bound
Residual variance	10.71	-3.31	1.28	.055	4.91	9.98
Intercept variance	10.09	3.04	1.86	.001	9.76	17.08

Functional impairment (WSAS)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI	
					Lower Bound	Upper Bound
Intercept	3.56	0.11	1.44	.005	0.75	6.41
Assessment time point	-1.09	-0.01	0.40	.016	-1.89	-0.35
Baseline severity (WSAS)	0.60	0.00	0.04	.001	0.51	0.68
Total morbidities	0.13	-0.01	0.12	.201	-0.10	0.36
Sudden gainers	-4.38	0.04	1.07	.001	-6.45	-2.38

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI	
					Lower Bound	Upper Bound
Residual variance	29.55	-8.81	5.24	.137	11.13	31.58
Intercept variance	31.31	8.31	5.64	.001	29.33	51.85

General health (VAS)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI	
					Lower Bound	Upper Bound
Intercept	51.65	-0.39	3.92	.001	43.47	59.06
Assessment time point	1.62	0.11	0.84	.089	-0.03	3.29
Baseline health (VAS)	0.28	0.00	0.05	.001	0.18	0.38
Total morbidities	-0.70	0.00	0.31	.014	-1.31	-0.05
Sudden gainers	5.86	0.28	2.23	.005	1.39	10.10

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI	
					Lower Bound	Upper Bound
Residual variance	130.22	-41.17	20.79	.124	51.46	134.06
Intercept variance	194.32	40.50	30.44	.001	178.14	301.79

Quality of life (EQ5D-5L)

Fixed effects

Parameter	B	Bias	SE	*p*	95% CI	
					Lower Bound	Upper Bound
Intercept	0.216	0.000	0.068	.001	0.088	0.356
Assessment time point	0.018	0.001	0.010	.100	-0.001	0.039
Baseline QoL (EQ5D-5L)	0.616	-0.011	0.078	.001	0.453	0.763
Total morbidities	-0.002	0.000	0.004	.500	-0.009	0.005
Sudden gainers	0.070	0.005	0.025	.005	0.026	0.126

Random effects

Parameter	Variance	Bias	SE	*p*	95% CI	
					Lower Bound	Upper Bound
Residual variance	0.015	-0.005	0.003	.151	0.006	0.016
Intercept variance	0.026	0.004	0.004	.001	0.022	0.039

Note. Abbreviations: B: unstandardised β coefficient; SE: standard error; p: significance at an alpha of .05; CI: confidence interval; ORS: outcome rating scale; SHAI: Short form health anxiety inventory; PHQ-9: Patient health questionnaire 9 items; GAD-7: generalised anxiety disorder 7 items; PHQ-15: patient health questionnaire 15 items; WSAS: work and social adjustment scale; VAS: visual analogue scale; EQ5D-5L: Euroqol –5 dimensions –5 levels; QoL: Quality of life.