

The Clinical Effectiveness of the Person-Centred Psychotherapies: The Impact of the Therapeutic Relationship

By Tony Weston

Thesis submitted in part-fulfilment of the requirements for the award of Doctor (PhD) in Counselling at the University of East Anglia, School of Lifelong Learning. This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Corrections submitted February 2011 – Original Date of Submission 2007

Document word count 109,585 words of which 9,699 words in appendices, thesis word count 99,886 words, which is less than the UEA PhD word limit of 100,00 words.

Abstract

Background: Person-centred psychotherapies have fared poorly in reviews of 'empirically supported therapies', compared with cognitive-behaviour therapy (CBT). Yet there is evidence of comparable efficacy and effectiveness of person-centred therapies (PCT), of elements of the therapeutic relationship as defined in PCT as a therapeutic process of change and an NHS research funding gap (£m CBT >> PCT). The author wondered if PCT was an effective intervention for a range of different symptoms and about the role of therapeutic relationships.

Aim: The aim was to measure the clinical effectiveness of PCT as practised in the author's private practice and by colleagues at the University of East Anglia Counselling Service and to assess the therapeutic relationship as a putative predictor of outcome.

Method: This was primarily an uncontrolled naturalistic experiment; outcome measures were completed at first therapy session and subsequently, along with a measure of the relationship.

Results: There was evidence that PCT was an effective intervention for clients who completed subsequent questionnaires who started with symptoms of depression (ES(d) =1.48, n =111), anxiety (ES(d) =1.15, n=91) and distress (ES(d) =1.80, n=79). These outcomes were broadly comparable with the literature. Some of the difficulties identified with uncontrolled naturalistic experiments described in the literature are addressed in the text, further supporting the validity of these findings. There was no evidence of the role of the therapeutic relationship (Rogers 1957) as a predictor of outcome for depression (n=92), anxiety (n=75) or distress (n=54). Further analysis of outlier and influential cases suggested the therapeutic relationship had an effect on depression outcome, $r = .22$. Illustrative analysis suggested the therapeutic relationship could have an effect on outcomes for anxiety in the order of $r = .25$ and distress $r = .29$. Non-positive findings may have been due to problems with the protocol and sample, these are discussed in the text and recommendations for future research made.

Conclusions: PCT warrants further outcome and change process research and inclusion as a comparator treatment condition in NHS-sponsored trials of CBT.

Declaration of interest: The author was trained and practices as a person-centred psychotherapist. During the period 19/5/5 – 13/7/6 the author received £8,167.07 from the University of East Anglia University Counselling Service; this was for sessional counselling work and included a contribution towards the costs of this research.

Table of Contents

Abstract	2
List of Figures	7
List of Tables	11
1. General Introduction	15
1.1 The source of this research.....	16
1.2 Overview of this research.....	17
2. Person-centred psychotherapies.....	23
2.1 Early research leading up to the statement of Rogers' theories.	23
2.2 Main statements of Roger's theories and Rogers' attempts to validate his theories.....	37
2.3 Contemporary person-centred psychotherapies.	42
3. Placing this research in the context of the literature	44
3.1 Outcomes literature	45
3.1.1 Overview of outcomes literature	46
3.1.2 Anxiety outcomes	51
3.1.3 Depression outcomes	54
3.1.4 Distress outcomes and related commentary on methodologies.....	55
3.1.5 The therapy workforce in the UK and investment in therapy research	58
3.1.6 Outcomes management.....	59
3.1.7 Concluding comments on outcome research.....	61
3.2 Process-outcome research literature.....	62
3.2.1 The case against Rogers' theory	64
3.2.2 Abuse of the drug metaphor	70
3.2.3 Causation approaches	75
3.2.4 Common factors	79
3.2.5 Meta-analytic approaches	82
3.2.6 Some Recent studies	85

3.2.7 Anti-depressant medication	91
3.2.8 Dose effect relationships	93
3.2.9 Concluding comments on process-outcome research.....	94
3.3 Research Design.....	95
3.3.1 Overview of design issues.....	95
3.3.2 Measuring outcomes.	96
3.3.3 Measuring process-outcome correlations.....	98
4. Introduction to this study	100
4.1 The need for this research	100
4.2 Overview of hypotheses to be tested.....	107
4.3 Specific hypotheses to be tested.....	109
4.4 Concluding comments on the introduction to the research.....	111
5. Method	113
5.1 Participants.....	114
5.1.1 Clients.....	114
5.1.2 Therapists.....	120
5.2 Measures.....	121
5.2.1 Outcome measures.....	121
5.2.2 Process measure.....	127
5.3 Research protocol.....	130
5.4 Treatment duration	133
5.5 Resulting sample	136
5.6 Participant flow and demographic characteristics for each sample analysed	138
5.7 Analytical approach.....	140
6. Results.....	141
6.1 Outcomes.....	141
6.1.1. Depression (BDI-II) outcomes	142

6.1.2 Anxiety (BAI) outcomes	149
6.1.3 Distress (CORE-OM) outcomes	155
6.1.4 Personality Disorder Prevalence (PBQ)	162
6.2 Predictor analysis	165
6.2.1 Depression (BDI-II) outcome prediction by therapeutic relationship (BLRI)	165
6.2.2 Anxiety (BAI) outcome prediction by therapeutic relationship (BLRI)	167
6.2.3 Distress (CORE-OM) prediction by therapeutic relationship (BLRI).....	169
7. Further results	171
7.1 Outcomes research: Addressing some of the criticisms of uncontrolled naturalistic research.....	171
7.1.1 Missing cases: Depression.....	172
7.1.2 Controlling for other causes of recovery	180
7.1.3 Lack of randomisation to different treatments	219
7.1.4 No evidence that treatments appropriately delivered	220
7.2 Process-outcome research.	225
7.2.1 The range of client perceptions of the therapeutic relationship tested in this research.....	225
7.2.2 The depression process-outcome data.	228
7.2.3 The anxiety process-outcome data.	230
7.2.4 The distress process-outcome data.	234
7.2.5 The predictive ability of within and between therapist variability.....	238
7.2.6 Possible sources of client ‘rating tendencies’ for the therapeutic relationship.....	238
8. Discussion	244
8.1 Summary of findings.....	244
8.1.1 Depression	245
8.1.2 Anxiety	251
8.1.3 Distress	255
8.2 Process-outcome findings	259

8.2.1 Introduction	259
8.2.2 Depression	260
8.2.3 Anxiety	261
8.2.3 Distress	262
8.3 The case for causal inference based on the outcomes and process-outcomes findings.	263
8.4 Strengths and weaknesses of this research	268
8.4.1 Questionnaires and related issues	268
8.4.2 Frequency of measurement and related issues	271
8.4.3 Measurement of the therapeutic relationship and related issues	274
8.4.4 Sampling – Clients, and related issues	276
8.4.5 Sampling – Therapists, and related issues	277
8.4.6 Methodology and related issues	278
8.4.7 Analysis and related issues	280
8.4.8 Summary of strengths and weaknesses	280
8.5 Discussion of results.....	282
8.5.1 Outcomes and related issues.....	282
8.5.2 Process-outcome correlations with therapeutic relationship and related issues	285
8.5.3 Person-centred theory and related issues.....	287
8.6 Implications of this research	290
9. References.....	291
10. Appendices.....	320
Appendix 1: Research Ethics Committee submission	320
Appendix 2: Characteristics of sample upon which coefficient alphas were estimated.	346
Appendix 3: Participant flow and demographic characteristics for each sample analysed	347

List of Figures

Figure 1: For University Counselling Service clients, flow diagram of participation at each stage of research.....	116
Figure 2: For Private Practice clients, flow diagram of participation at each stage of research.	119
Figure 3: Extract of Barrett-Lennard Relationship Inventory as used in this research: Tick box format as agreed with Barrett-Lennard.	128
Figure 4: Research protocol: The six phases of development of the research protocol.	132
Figure 5: Jacobson Plot for Depression (BDI-II) Outcomes.	145
Figure 6: Jacobson Plot for Depression (BDI-II) Outcomes using transformed data.....	147
Figure 7: Jacobson Plot for Anxiety (BAI) Outcomes.....	152
Figure 8: Jacobson Plot for Anxiety (BAI) Outcomes using transformed data.	154
Figure 9: Jacobson Plot for Distress Outcomes.	158
Figure 10: Example of client PBQ score: Black bars indicate mean z-scores for patients with diagnosis of each personality disorder whilst this questionnaire was being tested (Beck, et al., 2001) and grey bars indicate this client's score for each personality disorder subscale.....	162
Figure 11: Regression of depression (BDI-II) at first and subsequent session for full sample (n=92): Two outliers with standardised residuals ≥ 3	166
Figure 12: Regression of anxiety (BAI) at first and subsequent session for full sample (n=75): Two outliers with standardised residuals ≥ 3	168
Figure 13: Regression of anxiety at first and subsequent session for full sample (n= 54): No outliers with standardised residuals ≥ 3	170
Figure 14: Hypothetical impact of 'missing cases' on effect size: removal of cases from the research for reasons given in text (x-axis is descending number of cases left 'in the research' once these cases are removed and y-axis is mean effect size on an LOCF basis).....	178
Figure 15: Scatter plot of transformed BDI-II at first session versus subsequent counselling session minus score at first session.	182
Figure 16: Scatter plot of regression to mean corrected transformed BDI-II at subsequent session minus first session score versus first session score.	183
Figure 17: Scatter plot of treatment period (number of days from first session to subsequent session) versus improvement in depression (BDI-II) during treatment period: One outlier with standardised residual ≥ 3	186

Figure 18: Estimated marginal means at exploratory session (stage 1) and first session (stage 2) for 36 UCS clients in the ‘wait and treatment’ condition and 32 UCS clients in the ‘wait only’ condition: stage (1 = exploratory session, 2 = first session) is x-axis and depression (BDI-II) score is y-axis.	189
Figure 19: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 36 UCS clients in the ‘wait and treatment’ condition and 75 PP clients in the ‘treatment only’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and depression (BDI-II) score is y-axis.....	189
Figure 20: Box plot of depression (BDI-II) scores at first session (y-axis) for clients with no subsequent BDI-II measurement (81 clients) and a subsequent BDI-II measurement (124 clients): Clients with a subsequent measurement started therapy with a significantly higher level of depression $p < .001$	191
Figure 21: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 25 clients in the ‘relevant medication taken’ condition and 86 clients in the ‘no relevant medication taken’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and depression (BDI-II) score is y-axis.....	192
Figure 22: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 21 clients in the ‘anti-depressant medication taken’ condition and 86 clients in the ‘no relevant medication taken’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and depression (BDI-II) score is y-axis.....	193
Figure 23: Scatter plot of transformed BAI at first session versus subsequent counselling session minus score at first session.....	196
Figure 24: Scatter plot of regression to mean corrected transformed BAI at subsequent session minus first session score versus first session score.	197
Figure 25: Scatter plot of wait period (number of days from exploratory session to first session) versus improvement in anxiety (BAI) during wait period: A negative improvement was a worsening in anxiety score.....	200
Figure 26: Estimated marginal means at exploratory session (stage 1) and first session (stage 2) for 36 UCS clients in the ‘wait versus treatment’ condition and 30 UCS clients in the ‘wait only’ condition: stage (1 = exploratory session, 2 = first session) is x-axis and anxiety (BAI) score is y-axis.....	203
Figure 27: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 36 UCS clients in the ‘wait versus treatment’ condition and 54 PP clients in the ‘treatment only’	

condition: stage (1 = first session, 2 = subsequent session) is x-axis and anxiety (BAI) score is y-axis.	204
Figure 28: Box plot of anxiety (BAI) scores at first session (y-axis) for clients with no subsequent BAI measurement (102 clients) and a subsequent BAI measurement (102 clients): Clients with a subsequent measurement started therapy with a significantly higher level of anxiety $p < .001$	205
Figure 29: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 20 clients in the 'relevant medication taken' condition and 71 clients in the 'no relevant medication taken' condition: stage (1 = first session, 2 = subsequent session) is x-axis and anxiety (BAI) score is y-axis.	206
Figure 30: Scatter plot of CORE-OM at first session versus subsequent counselling session minus score at first session.	208
Figure 31: Scatter plot of regression to mean corrected CORE-OM at subsequent session minus first session score versus first session score.	209
Figure 32: Scatter plot of treatment period (number of days from first session to a subsequent session) versus improvement in distress (CORE-OM) during treatment period: A positive improvement in CORE-OM score was a lessening in distress.	211
Figure 33: Estimated marginal means at exploratory session (stage 1) and first session (stage 2) for 37 UCS clients in the 'wait versus treatment' condition and 34 UCS clients in the 'wait only' condition: stage (1 = exploratory session, 2 = first session) is x-axis and distress (CORE-OM) score is y-axis.	215
Figure 34: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 37 UCS clients in the 'wait versus treatment' condition and 41 PP clients in the 'treatment only' condition: stage (1 = first session, 2 = subsequent session) is x-axis and distress (CORE-OM) score is y-axis.	216
Figure 35: Box plot of distress (CORE-OM) scores at first session (y-axis) for clients with no subsequent CORE-OM measurement (66 clients) and a subsequent CORE-OM measurement (89 clients): Clients with a subsequent measurement started therapy with a non-significant higher level of distress ($p=.054$).	217
Figure 36: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 12 clients in the 'relevant medication taken' condition and 67 clients in the 'no relevant medication taken' condition: stage (1 = first session, 2 = subsequent session) is x-axis and distress (CORE-OM) score is y-axis.	218

Figure 37: Box plot of BLRI subscale scores for 118 clients who completed the BLRI in this research.	221
Figure 38: ‘Regard’ BLRI subscale scores: Distribution was significantly different from normal.	222
Figure 39: ‘Empathy’ BLRI subscale scores: Distribution was significantly different from normal.	223
Figure 40: Distribution of BLRI scores for the process-outcome research.	227
Figure 41: Histrionic personality disorder (PBQ-HI subscale score) as a predictor of client perception of the therapeutic relationship (BLRI score)....	240
Figure 42: Box plot of therapeutic relationship (BLRI) subscale scores.....	241
Figure 43: Origin of clients (Phase of Protocol) and questionnaire completion for clients completing BDI-II at First Session.	349
Figure 44: Origin of clients (Phase of Protocol) and questionnaire completion for clients completing BAI at First Session.	351
Figure 45: Origin of clients (Phase of Protocol) and questionnaire completion for clients completing CORE-OM at First Session.....	353
Figure 46: Origin of clients (Phase of Protocol) in depression predictor sample: Clients who completed BDI-II at First Session, BLRI at a subsequent session and BDI-II at a subsequent session.	359
Figure 47: Origin of clients (Phase of Protocol) in anxiety predictor sample: Clients who completed BAI at First Session, BLRI at a subsequent session and BAI at a subsequent session.	359
Figure 48: Origin of clients (Phase of Protocol) in distress predictor sample: Clients who completed CORE-OM at First Session, BLRI at a subsequent session and CORE-OM at a subsequent session.	360

List of Tables

Table 1 : Summary of meta-analysis of key elements of the therapy relationship based on Norcross, 2010	84
Table 2: Calculation of reliable change criteria (Reliable Change Index, RCI) for 95% confidence in each outcome measure: Data from this research.	122
Table 3: For all clients in the research, numbers of sessions this episode.....	133
Table 4: For University Counselling Service clients, numbers of sessions this episode.....	134
Table 5: For Private Practice clients, numbers of sessions this episode.....	135
Table 6: For Private Practice clients paid for by their employer, number of sessions this episode.	136
Table 7: Summary of outcomes hypotheses.	141
Table 8: Depression (BDI-II) outcomes.	143
Table 9: Depression (BDI-II) outcomes: Comparison with other studies.	144
Table 10: Depression (BDI-II) severity at first and subsequent measurement.	146
Table 11: Percentages of clients with reliable change and recovered from depression (BDI-II).....	146
Table 12: Anxiety (BAI) outcomes.....	150
Table 13: Anxiety (BAI) outcomes: Comparison with other studies.	151
Table 14: Anxiety (BAI) severity at first and subsequent measurement.	153
Table 15: Percentages of clients with reliable change and recovered from anxiety (BAI). ..	153
Table 16: Distress (CORE-OM) outcomes.	156
Table 17: Distress (CORE-OM) outcomes: Comparison with other studies.	157
Table 18: Distress (CORE-OM) severity at first and subsequent measurement.	159
Table 19: Percentages of clients with reliable change or recovered from symptoms of distress (CORE-OM). ..	160
Table 20: Benchmarking percentages of clients with reliable change or recovered from symptoms of distress (CORE-OM): Compared with published benchmarks by Mullin et al. , 2006.....	161
Table 21: Primary Personality Disorder within Private Practice sample: Highest PBQ subscale z-scores for each client (n = 53).	163
Table 22: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Depression (BDI-II) at First Session in Predicting Depression at Last Session for 92 clients with measurement of depression at first and subsequent session plus BLRI measurement. .	165

Table 23: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Depression (BDI-II) at First Session in Predicting Depression at Subsequent Session for 90 clients with measurement of depression at first and subsequent session plus BLRI measurement.	166
Table 24: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Anxiety (BAI) at First Session in Predicting Anxiety at Subsequent Session for 75 clients with measurement of anxiety at first and subsequent session plus BLRI measurement.	167
Table 25: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Anxiety (BAI) at First Session in Predicting Anxiety (BAI) at Subsequent Session for 73 clients with measurement of anxiety at first and subsequent session plus BLRI measurement.	168
Table 26: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Distress (CORE-OM) at First Session in Predicting Distress (CORE-OM) at Subsequent Session for 54 clients with measurement of distress at first and subsequent session plus BLRI measurement.	169
Table 27: Numbers of sessions attended by 19 PP ‘missing cases’	173
Table 28: Numbers of sessions attended by 32 UCS ‘missing cases’	174
Table 29: Number of sessions attended by UCS clients who might be considered ‘treatment failures’	175
Table 30: Groups of clients who could be included or excluded from outcome measurement and estimated impact on effect size and reliable change percentage.	177
Table 31: For different denominators: percentages of clients with reliable change and recovered from depression.	179
Table 32: Depression (BDI-II) outcomes: Comparing ‘wait’ with ‘treatment’	185
Table 33: Summary of Regression Analysis for Wait period (days) in predicting change in Depression (BDI-II) during wait: Model 1 has no outlying cases (standardised residuals ≥ 3).	185
Table 34: Summary of Regression Analysis for Treatment period (days) in predicting change in Depression during treatment: Model 1 with outlying case included (standardised residual ≥ 3) and Model 2 with outlying case deleted from the model.	187
Table 35: Anxiety (BAI) outcomes comparing wait and treatment.	199
Table 36: Summary of Regression Analysis for Wait period (days) in predicting change in Anxiety (BAI) during wait: Model 1 had no outlying cases (standardised residuals ≥ 3). .	201

Table 37: Summary of Regression Analysis for Treatment period (days) in predicting change in Anxiety (BAI) during treatment: Model 1 had no outlying cases (standardised residuals ≥ 3)	201
Table 38: Distress (CORE-OM) outcomes comparing wait and treatment	210
Table 39: Summary of Regression Analysis for Wait period (days) in predicting change in Distress (CORE-OM) during wait: Model 1 has no outlying cases (standardised residuals ≥ 3)	212
Table 40: Summary of Regression Analysis for Treatment period (days) in predicting change in Distress (CORE-OM) during treatment: No outlying cases (standardised residual ≥ 3)	212
Table 41: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Depression Score (BDI-II) at First Session in Predicting Depression Score at Subsequent Session.	229
Table 42: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Transformed Anxiety Score (BAI) at First Session in Predicting Transformed Anxiety Score at Subsequent Session – Controlling for Site (PP versus UCS).....	231
Table 43: Summary of Regression Analysis for Therapeutic Relationship (BLRI), Site and Transformed Anxiety Score (BAI) at First Session in Predicting Transformed Anxiety Score at Subsequent Session: Deleting highest SDBETA BLRI sequentially.....	232
Table 44: Characteristics of cases that did not fit the anxiety predictor structure, together with possible reason(s) for non-fit.	233
Table 45: Summary of Regression Analysis for Therapeutic Relationship (BLRI), Number of Sessions and Distress Score (CORE-OM) at First Session in Predicting Distress Score at Subsequent Session: Deleting outliers and influential cases.	235
Table 46: Summary of Regression Analysis for Therapeutic Relationship (BLRI), Number of Sessions and Distress Score (CORE-OM) at First Session in Predicting Distress Score at Subsequent Session: Deleting highest SDBETA BLRI sequentially.	237
Table 47: Regression analysis of outcome measures at the start as a predictor of the therapeutic relationship (BLRI score): Possible sources of within therapist variability.	239
Table 48: Summary of further results for impact of therapeutic relationship on outcome: Significant effect on depression and illustrative results for anxiety and distress.	242
Table 49: Characteristics of overall sample plus subsets of University Counselling Service and Private Practice samples.....	348
Table 50: Characteristics of sample for clients with depression (BDI-II) outcomes.....	350
Table 51: Characteristics of sample for clients with anxiety (BAI) outcomes.	352

Table 52: Characteristics of sample for clients with distress (CORE-OM) outcomes.	354
Table 53: Characteristics of sample for clients in wait time analysis: Clients with clinical scores at first session, any score at exploratory and last session plus numbers of days in wait and treatment conditions.	356
Table 54: Characteristics of clients in hypotheses testing.	358

1. General Introduction

This section introduces the research that follows. Firstly a general introduction is provided that describes the source of this research for the present author and gives a brief overview of the need for this research, in particular to provide a context for the subsequent literature review. The literature review seeks to place this research in an appropriate context of what is known about person-centred psychotherapy outcomes and the impact of the therapeutic relationship. The literature review is in two parts, firstly to introduce person-centred psychotherapies and secondly to review outcomes literature and process-outcomes literature. The introduction to person-centred psychotherapies describes the early research leading up to the statement of Rogers' theory, describes Rogers' theory and his attempts to validate his own theory. Contemporary person-centred psychotherapies are briefly described, since these are not simply about 'reflections of feelings'. The subsequent literature review is broadly divided into a review of outcomes literature and a review of process-outcomes literature, since these are the two main foci for the experimental part of this thesis. It is important to note that this is an arbitrary division between outcome and process-outcome literature, with the purpose of providing some structure to this part of the thesis, since researchers have frequently addressed both issues in the same article, as does this thesis. This part of the literature review is brought together by an introduction to the research that follows.

In subsequent sections there is a description of the methodology, a statement of the results, especially as these relate to the hypotheses, and presentation of some further results that further explore the results as related to the hypotheses. Finally a discussion is provided that summarises the results, the strengths and weaknesses of the research and draws some conclusions.

For consistency UK spellings are used throughout, so that 'counselling' is used instead of interchanging with 'counseling' for US authorship. The terms counselling and psychotherapy have been used interchangeably. Rogers first used the term 'counselling' because in 1940s America it was necessary to be a qualified medical practitioner to practice 'psychotherapy' (Thorne, 2003, p. 14 and p. 60).

1.1 The source of this research

In the UK the National Institute for Health and Clinical Excellence (NICE) has made recommendations on what treatments should be offered for particular diagnoses including depression and anxiety (NICE, 2002a), schizophrenia (NICE, 2002b), panic disorder, with and without agoraphobia and generalised anxiety disorder (NICE, 2004a), depression (NICE, 2004b), anorexia, bulimia and related eating disorders (NICE, 2004c), self harm (NICE, 2004d), post traumatic stress disorder (PTSD, NICE, 2005a), depression in children and young people (NICE, 2005b), obsessive-compulsive disorder (NICE, 2005c), borderline personality disorder (draft guidelines, NICE, 2008) and revised (draft) depression guidelines (NICE, 2009a). Whilst cognitive-behavioural therapies (CBT) have fared very well in these reviews and are recommended for all diagnoses, the person-centred psychotherapies have fared poorly in these recommendations and in fact are recommended only for mild to moderate depression (NICE, 2004b). This situation is likely to change such that patients are given 'warnings' about the absence of an evidence-base for counselling for depression (NICE, 2009a).

The trend towards evidence-based recommendations for mental health problems was begun in the US with the American Psychological Association (APA), Division of Clinical Psychology, Task Force on Promotion and Dissemination of Psychological Procedures (1995). Dianne Chambless has been a key supporter of this move (e.g. Chambless 1996) in a process described by Dobson and Craig (1998). These recommendations have been subject to some updates (e.g. Chambless & Hollon, 1998) and some reframing, for example Gone and Alcantra, 2007, reviewed the literature to identify effective mental health interventions for American Andians and natives of Alaska. Again, person-centred psychotherapies have fared poorly in reviews in the US too.

As a recently qualified person-centred psychotherapist the author was concerned about this situation and decided to investigate further by doing a Masters dissertation in this subject area (Weston, 2005). This revealed to the author that not all researchers agreed with these recommendations, neither in terms of their methodologies nor their findings. Bohart, O'Hara and Leitner (1998) wrote about what they termed 'empirically violated treatments' and described what they called the 'disenfranchisement of humanistic and other psychotherapies'. In a large meta-analysis Elliott, Greenberg and Lietaer (2004) presented evidence that in their view suggested experiential therapies (humanistic therapies including person-centred, Gestalt,

existential, process-experiential, etc.) were as effective as other therapies, including CBT, for a wide range of client issues including depression, anxiety disorders, trauma and marital problems. These authors argued that using the criteria developed by the APA (1995) and subsequently made stricter (Chambless & Hollon, 1998) experiential therapies were 'efficacious' for depression (the process-experiential suborientation was 'specific and efficacious'); 'possibly efficacious' for anxiety disorders; the process-experiential suborientation was 'specific and efficacious' for traumatic and abusive events, and; 'efficacious and possibly specific' for emotion focused therapy with couples. These authors acknowledged that more outcome research was needed for all types of humanistic therapies across different client issues, particularly in the political context that existed in the US, UK, Germany, Netherlands, Austria, etc.

From reading material similar to that described above the present author was left wondering whether person-centred psychotherapies (also known as 'counselling') did have any beneficial effect for clients and decided to conduct some primary research into the clinical effectiveness of the person-centred psychotherapies. Furthermore, the author wanted to research the validity of Rogers' theory (e.g. 1957, 1959) and to assess the impact of the therapeutic relationship as described by Rogers upon client outcomes. The next section gives an overview for why this research was needed to set the context for the subsequent review of literature.

1.2 Overview of this research.

In some respects person-centred psychotherapy is in crisis. In the UK the only diagnosis that achieves any kind of 'recommendation' for counselling from the main body that informs the NHS about treatment policy is that patients with depression should be warned about the absence of an evidence-base for counselling for depression (NICE, 2009a). Otherwise clinicians are warned against offering counselling for all other psychologically-based diagnoses that NICE has reported on.

Yet, non-directive therapy was founded upon empirical principles. During the course of the 1940s – 1960s Carl Rogers' team made early attempts at establishing pretty much every research technique currently in use. It is widely believed that Rogers was the first researcher

to record therapy interviews, transcribe these and analyse them. Rogers established the team-based approach to research wherein many researchers used the same experimental material for many different forms of analysis. These people worked in a manner that was a fore-runner of the National Institute for Mental Health (NIMH) Treatment of Depression Collaborative Research Programme (TDCRP, Elkin, et al, 1989), probably the most widely researched data in psychology history. Rogers established a massive, in its day, research programme at Wisconsin that employed hundreds of people and millions of dollars in today's money to test his theory.

Rogers' theory, that the perception of congruent empathy and unconditional positive regard by the client from the therapist, was evidence-based. Initially, CBT, now the most widely recommended therapy in the UK and beyond, rejected the idea of the therapeutic relationship as causative. Beck's (1976) theory was that a good therapeutic relationship was simply a convenient atmosphere to teach the techniques of CBT that would cause outcome.

It would be too simplistic to describe the therapeutic alliance (Horvath & Greenberg, 1989) as a re-branding of the therapeutic relationship. Yet, this change of name, heralded a huge research interest in the goals of therapy, the tasks of therapy and the bond between client and therapist. The therapeutic alliance is not only shown to predict outcome, it is also increasingly seen as causing outcome. Even CBT researchers (Strauss, et al., 2006, Spinhoven, Giez, van Dyck, Kooiman, & Arntz, 2007) are finding that good relationships appear to cause outcomes and wondering how they can get along better with their patients because of it.

In the UK a very large number of the 'therapy work-force' are not trained in CBT, perhaps 26,000 individual therapists (A. Couchman, personal communication, 3rd January 2008); do all non-CBT trained therapists need to be re-trained in CBT?

When this research was begun in 2004 the present author joked with the Head of the Department that if person-centred therapy didn't sort itself out soon it would find itself supplanted by a new form of CBT-relationship therapy wherein the therapist does not use techniques, because these might get in the way of the relationship, and instead focused on trying hard to understand what it felt like to be the client, unconditionally accepting what and how the client was. Even NICE write about the importance of good therapy relationships (NICE, 2004).

NICE (2004) also appear to accept that over three billion pounds of public money spent on anti-depressants in the past couple of decades (Weston & Weston, 2008) was not money well spent because anti-depressants don't make much of a difference to most people (Kirsch, et al., 2008) and that much of the apparent improvement could be due to the therapeutic relationship (Brown, 2007).

Many/most (but not all) researchers agree that the main schools of therapy are approximately equivalent in effect (Lambert & Barley, 2002) and that the single most important thing a therapist can do for a client is to have a good working alliance/therapeutic relationship with them (Norcross, 2010).

Yet NICE don't think patients should have person-centred psychotherapy, a therapy founded on the idea that it is the relationship that cures.

There was a need for this research to look at the clinical effectiveness of person-centred psychotherapy. Reviewers of therapy want evidence of efficacy/effectiveness when they seek to make evidence-based recommendations. Seemingly reviewers of research have found insufficient evidence for person-centred therapy and this research was needed to see if person-centred psychotherapy was an effective intervention.

During the early part of this century a lot of effort has been put into developing a broadly-based measure of distress (CORE-OM, Evans, et al., 2002) to encourage psychotherapists to routinely measure outcomes. Yet, when NICE (2009a) reviewed treatments for depression they rejected any evidence based on CORE-OM because, they argued, CORE-OM was not a specific diagnosis of depression, even though there is evidence that depression can be diagnosed with CORE-OM and that there is large covariance with measures of depression (Gilbody, Richards, & Barkham, 2007).

This research was needed to do research on the effectiveness of person-centred psychotherapy with diagnostic specific measures, in addition to the more usual broadly based psychotherapy outcome measure (CORE-OM).

This research was needed to look at person-centred outcomes for depression and anxiety as these are two of the most common psychologically-based diagnoses (NICE, 2004a, 2004b) with huge economic cost.

Ideally, reviewers of therapy want evidence from 'randomised controlled trials' (RCTs). These are difficult to organise and expensive, certainly beyond the spending of an un-funded PhD student. It would be ideal to conduct an RCT for person-centred psychotherapy and this research was needed to make the case for an RCT. It seemed unlikely that funders of research would come up with a large amount of money to fund an RCT without any prior case.

Whilst the needed research was necessarily naturalistic, this is not necessarily a 'poor' method. Some researchers favour naturalistic research because it is 'real world' evidence of effectiveness e.g. Des Jarlais, Lyles, Crepaz, and TREND Group (2004), Victora, Habicht, and Bryce (2004), Schwartz, Trask, Shanmugham, and Oswald Townsend (2004), Westen, Novotny, and Thompson-Brenner, (2004).

One of the challenges with naturalistic outcomes research is to make a 'well controlled' study, wherein alternate causes of outcome are ruled out as far as possible. For example, it could be that concurrently prescribed medications were really responsible for any observed outcome, that any observed changes were simply regression to the mean, that the clients would have got better anyway, etc. Whilst naturalistic studies can be 'cheap' they require pain-staking analysis. In addition to finding out about antidepressants this research needed to find out about personality disorders. The presence of a co-morbid personality disorder is known to reduce the size of outcomes from therapy (Clarkin & Levy, 2004). In addition to considering alternate causes this research needed to find out about the presence of co-morbid personality disorders as a putative moderator of effectiveness.

This research needed to do the analysis of alternate causes or moderators, so that any observed changes could be interpreted appropriately.

Ideally reviewers of therapy want evidence from not just one study but many. In the US the definition of an empirically supported therapy is one that includes research by more than one research team (Chambless & Hollon, 1998). Some might say that person-centred therapy has already been researched and shown to have good outcome, say for depression, on a basis comparable with CBT (Ward, et al., 2000). This research was needed to add to the evidence-base for person-centred psychotherapy (Elliott, et al., 2004).

Contrary to Rogers' early interest in quantitative research, towards the end of his life he wrote a paper (1985) that many in the person-centred community appear to have taken as a rebuff to quantitative methods. Despite person-centred therapy leading the way in the early

years with research, the present author formed the impression from the literature review that person-centred therapy was in danger of being left behind as researchers from other schools adopted increasingly complex and sophisticated methods capable of reaching powerful conclusions. Without being overly presumptuous the present author formed the view that this research was needed in the hope that others from the person-centred approach would consider and do quantitative research. As the literature review shows research methodologies are continually evolving and there is no room for the sense that 'research has been done, no more required'. Consequently the literature review shows there has been a huge leap forward in what is now possible compared with the early efforts of Rogers and his colleagues. This research was needed to promote the idea of quantitative research amongst person-centred therapists and to begin a process of adopting increasingly complex methods.

As mentioned above, person-centred psychotherapy was founded on the idea that the relationship was the therapy. In addition to outcomes evidence researchers and reviewers want to see evidence for the proposed treatment rationale (Elliott, 2010). The present author reviewed the literature in the hope of finding research evidence to support the idea that the relationship as defined by Rogers *causes* outcome. The review of literature showed this was not straight-forward to do, methodologically, and early attempts at this were subsequently ruled out as invalid by reviewers.

Some subsequent research has established process-outcome correlations for some of the elements of Rogers' theory (Norcross, 2010). Drawing on the criticisms of prior attempts at establishing a correlation between the therapeutic relationship as defined by Rogers and outcome the present author sought to define an experiment that could test this element of Rogers' theory.

The present author could find no research that satisfied both the requirements of Rogers' theory and subsequent reviewers of the early research. There was a need for this research to do a simultaneous test of the Rogerian relationship elements with both 'congruent' and 'incongruent' clients to see if any observed effect of person-centred psychotherapy had anything to do with the therapeutic relationship.

The present author was unaware of any comparable methodologically balanced study of the impact of the therapeutic relationship as defined by Rogers on depression, anxiety or distress outcomes as a consequence of person-centred psychotherapy and this was one of the things that made this research unique.

Mention was made above about the impact of moderators on therapy outcome. In addition to moderating outcome these variables can also moderate process-outcome correlations. This research was needed to consider whether the effect of therapy on outcome had anything to do with the relationship as defined by Rogers whilst considering the impact of potentially moderating variables and extreme cases and this was one of the things that made this research unique.

Given the interest in the field about 'good relationships with clients' this research was needed as part of the research effort to establish that person-centred therapists 'know about relationships' and that the approach has something to offer both in research and practice 'about relationships' e.g. in therapy, medicine, management and beyond.

Given the needs of this research, to conduct a well-controlled outcome study on depression, anxiety and distress and to look for evidence that the therapeutic relationship had an impact on outcome, the following literature review considers what is known about person-centred psychotherapy, outcomes research and process-outcomes research.

2. Person-centred psychotherapies.

Person-centred psychotherapy is largely credited to the work of Carl Rogers and his collaborators. This section describes the development and testing of Rogers' theory in the context of the evolution of psychotherapy research wherein key themes have been 'does it work?' (Outcome research) and 'if it does work, how does it work?' (Causation research). The contemporary view appears to be that psychotherapy does work and works because of a number of factors common to different therapies, such as a bond between client and therapist and agreement on the tasks of therapy in pursuit of the client's goals; the perception of empathy from the therapist for the client's situation; the perception of positive regard and affirmation from the therapist for the client; the perception of genuineness from the therapist by the client; the repair of ruptures in the therapeutic alliance; the management of counter transference, and the adaptation of the therapeutic relationship to suit the needs of the particular client (Norcross, 2010). There are some who argue that particular therapies are more effective than other therapies for particular diagnoses, or client goals, and that particular therapies have unique 'ingredients' (e.g. Siev and Chambless, 2007). These differing views are considered below as part of a description of the history of outcome and process-outcome research.

2.1 Early research leading up to the statement of Rogers' theories.

This section sets Rogers' theories in the context of what was known about therapy outcomes and the impact of the therapeutic relationship from research findings at the time that Rogers wrote and published his theories (1957, 1959), specifically from the research that Rogers referred to, to the extent that this material was available.

Carl Rogers (1902 – 1987) developed an approach to psychotherapy that has variously been called 'non-directive', 'client-centred' and 'person-centred'. His early years are traced by a number of authors (e.g. Kirschenbaum 1979, 2007, Barrett-Lennard, 1998 and Thorne, 2003). He began his career in psychology studying clinical and educational psychology at Teachers College, Columbia University. Here he was exposed to what Barrett-Lennard called 'an objective measurement-oriented ethos' (p. 5). Rogers' first job as a psychologist was at the Child Study Department established by the Rochester Society for the Prevention of Cruelty to

Children. Whilst at the Child Study Department Rogers continued to study part-time for his PhD. Rogers also held a fellowship at the Institute of Child Guidance and this had the context of what Barrett-Lennard called 'eclectic Freudianism' (p. 5) and Rogers was exposed to psychoanalytic theory (Thorne, 2003, p. 8). Rogers completed his doctorate by developing a measure of personality adjustment in children aged nine to thirteen years (1931), although his first journal article was about 'intelligence as a factor in camping activities' (Rogers & Carson, 1930).

Around this time a paper was published that would become important, in the sense of being continually referred to, on a number of occasions right up until the present day. Writing in 1936 Saul Rosenzweig used the phrase the 'Dodo bird verdict' (derived from Lewis Carroll's 1865 Alice's adventures in Wonderland wherein the Dodo bird when asked to judge a race declared '*Everybody* has won, and all must have prizes', p. 34, emphasis original) to suggest that common factors (e.g. the therapy relationship) shared between different types of psychological therapies caused different therapies to be similarly effective. Rogers himself would make an important contribution to the 'common factors' idea in his 1957 theory paper, as well as setting out the key theoretical statement for what would become 'person-centred therapy' in his 1959 theory paper.

Meanwhile, also in 1930s America, Rogers went on to write about 'The Clinical Treatment of the Problem Child' (1939) based upon his own practical experiences of working with children and their parents.

Barrett-Lennard (1998, pp. 8-9) pointed out that Rogers' 1939 description of what Rogers called 'relationship therapy' foreshadowed many of the therapeutic principles that Rogers subsequently elaborated upon:

1. It applies only to those parents who have a desire to be helped...
2. The relationship between the worker and the parent is the essential feature... The worker endeavours to provide an atmosphere in which the parent can come to freely experience and realise his own attitudes...
3. The effects of this relationship upon the parent may be characterised by the terms 'clarification of feelings' and 'acceptance of self'...
4. ...another characteristic of this viewpoint is its reliance on the parent himself to determine independently the manner of dealing with the child. (Rogers, 1939, pp. 197-9).

Later Rogers was to drop his idea of 'client motivation' mentioned in 1939 from his 1957 and 1959 statements of his theory, although client motivation has subsequently been shown to make an important contribution to outcome e.g. Lambert and Barley (2002).

Following publication of his book in 1939 Rogers was offered and accepted an appointment as full professor at Ohio State University. He took up his appointment in January 1940 and had an article published later that year in the September/October edition of the Journal of Consulting Psychology (now the Journal of Consulting and Clinical Psychology) entitled 'The processes of therapy'. Writing in 1940 Rogers set out what could be described as a manifesto to scientifically research 'the processes of therapy', because:

Recent years have brought significant progress in the field of psychotherapy. The help obtained by the individual in a series of interviews is no longer a vague mystery impossible of serious investigation... The time is perhaps ripe for various workers to endeavour to formulate and describe the fundamental aspects of this process, in order that such descriptions may serve as hypotheses to be tested by research (p. 161).

In his 1940 paper Rogers set out his hypotheses and described how:

It is essential that certain basic conditions be met...It is probably necessary that the client, whether child or adult, should feel some dissatisfaction with present adjustment, some fundamental need of help... Therapy has no chance of being successful if there is too heavy a weight of adverse social factors making adjustment impossible except through radical alteration of circumstances (p. 161).

In addition to dropping 'client motivation' from his later theory statements Rogers also dropped 'social context'; again subsequently this was shown to play an important contribution to successful therapy outcome e.g. Lambert & Barley (2002).

One of the features of what Rogers later called 'non-directivity' that he appeared to recognise was, given the freedom to focus on either past or present, or a mixture of both, the client would get to whatever was most important to them:

It is worth noting that some schools of thought encourage expression of material related to past experience, others material related to present feelings. There seems to be no evidence that one is more therapeutic than the other, since in an important sense "all roads lead to Rome" (p. 162).

Rogers described how 'granted these conditions and a skilled therapist whose purpose is to release and strengthen the individual, rather than to intervene in his life, certain processes seem to take place, or if they do not take place, therapy is likely to be unsuccessful' (p. 161). Rogers described six characteristics of 'most successful therapeutic experiences':

1. Rapport is established.
2. There is free expression of feeling on the part of the client.
3. Recognition and acceptance, by the client, of his spontaneous self.
4. The making of the responsible choices.
5. The gaining of insight through assimilated interpretation.
6. Growing into independence – with support. (p. 162-3).

Rogers appeared to anticipate what he would subsequently call 'empathy' and 'congruence' in his main theory papers in his 1940 paper: 'There must be on the part of the counsellor a genuine interest in the individual, a degree of identification which is none the less real because it is understood and to some extent controlled' (p. 162).

In terms of the research 'manifesto' Rogers described the necessity of 'stripping therapy... [to the] bare bones of the therapeutic process... if we are to make progress...' (p. 163-4) and:

In closing, attention might be called to the research opportunities with which the therapeutic process bristles. If clinical and applied psychology is to win the status it desires, if it is to find sound answers to the problems of human relationships which are so urgently needed in a distraught world, then we will need to promote much more study and effort than heretofore, in this dynamic field of therapy (p.164).

Rogers closed the year of 1940 by giving a talk at the University of Minnesota in December at which he talked about 'newer concepts in psychotherapy' and this seemed to mark some kind of turning point for him personally and perhaps for psychotherapy:

I was totally unprepared for the furore the talk aroused. I was praised, I was attacked, I was looked on with puzzlement. By the end of my stay in Minneapolis it struck me that perhaps I was saying something new that came from me... I began to believe that I might personally, out of my own experience, have some original contribution to make to the field of psychotherapy (Rogers, 1974, p. 8 quoted in Barrett-Lennard, 1998, p. 10).

In 1942 Rogers published his second book 'Counselling and Psychotherapy' and used material from his talk at Minnesota as chapter two. The new therapy was called 'non-directive' therapy because the therapist did not give advice or tell the client what to do to resolve their difficulties; clients were encouraged to be self directing. This was based on the modest proposition that, rather than all-knowing therapists, clients knew how best to live their lives:

The counselling relationship is one in which warmth of acceptance and absence of any coercion or personal pressure on the part of the counsellor permits the maximum expression of feelings, attitudes and problems by the counselee. The relationship is a well structured one with limits of time, of dependence and of aggressive action, which apply particularly to the client, and limits of responsibility and of affection which the counsellor imposes on himself. In this unique experience of complete emotional freedom within a well defined framework the client is free to recognise and understand his impulses and patterns, positive and negative, as in no other relationship (p. 113-4).

The 1942 book contained the full transcript of Rogers' eight counselling sessions with 'Herbert Bryan' and this was probably the first time a full therapy transcript had ever been published. As a full professor Rogers could now supervise students to complete masters and doctoral degrees, and his students used sound recordings, transcriptions and analyses of these in their research. Rogers would subsequently draw upon these published and unpublished works in his main theory papers.

In 1944 Rogers was invited to spend the summer teaching at the University of Chicago and was subsequently offered a permanent position with the opportunity to establish a counselling centre. He was by now recognised as the creator of a new and distinctive approach to therapy (Barrett-Lennard, 1998). Rogers took up the position at Chicago in 1945 and stayed for the next twelve years.

Barrett-Lennard (1998) pointed out that America's contemporaneous war with what he called 'totalitarian anti-democracies' gave a context to the 'democratic' nature of non-directive therapy. Rogers co-authored a book about providing support to service personnel returning from the Second World War in which it was stated 'It is perhaps no accident that this emphasis in counselling [non-directivity] has reached its fruition in America' (Rogers & Wallen, 1946, p. 23).

Whilst at Chicago Rogers was elected President of the American Psychological Association (APA) for 1946-7 and in his retiring address made 'some observations on the organisation of personality' (1947) which showed how his thinking was developing:

The counsellor attitude of warmth and understanding... helps to maximise the freedom of expression by the individual. The client experiences sufficient interest in him as a person, and sufficient acceptance, to enable him to talk openly, not only about surface attitudes, but increasingly about intimate attitudes and feelings hidden from himself (p. 358-9).

Rogers was clear that what he was presenting was a series of qualitative observations upon which certain hypotheses had been formed and these observations and hypotheses were being presented prior to adequate quantitative confirmation:

I wish in this paper to try to bring you some of the clinical observations which we have made as we have repeatedly peered through these psychological windows into personality, and to raise with you some of the observations about the organisation of personality which these observations have forced upon us... What I shall offer is not a series of research findings but only the first step in that process of gradual approximation which we call science, a description of some observed phenomena which appear to be significant, and some highly tentative explanations of these phenomena (p. 359).

Rogers key hypothesis was '... that given certain psychological conditions, the individual has the capacity to reorganise his field of perception, including the way he perceives himself, and that a concomitant or a resultant of this perceptual reorganisation is an appropriate alteration of behaviour' (p361).

Rogers' idea was that it was the therapeutic relationship that helped the client to accomplish this because:

Client-centred therapy is different from other life situations inasmuch as the therapist tends to remove from the individual's immediate world all those aspects of the field which the individual can reorganise except the self. The therapist, by reacting to the client's feelings and attitudes rather than to the objects of his feelings and attitudes, assists the client in bringing from background into focus his own self, making it easier than ever before for the client to perceive and react to the self (p. 366).

In Rogers' hypothesis the therapeutic relationship facilitated this change:

By offering only understanding and no trace of evaluation, the therapist removes himself as an object of attitudes, becoming only an alternate expression of the client's self. The therapist by providing a consistent atmosphere of permissiveness and understanding removes whatever threat existed to prevent all perceptions of the self from emerging into figure. Hence, in this situation all the ways in which the self has been experienced can be viewed openly, and organised into a complex unity. It is then this complete absence of any factor which would attack the concept of self, and second, the assistance in focusing upon the perception of self, which seems to permit a more differentiated view of self and finally the reorganisation of self (p. 366).

Rogers noted that once all of these denied perceptions were integrated into awareness this was accompanied by '...feelings of comfort and freedom from tension which are experienced as psychological adjustment' (p. 364).

In contrast to the perhaps 'marginal' role of person-centred therapy that could be inferred from the NICE and APA reviews, in 1940s America Rogers' non-directive therapy was what might be termed 'cutting edge'. This 'new approach' embraced empiricism and with its creator as President of the APA was certainly 'respectable', if not 'mainstream'.

In his 1959 theory paper Rogers referred to the work of Assum and Levy (1948) who reported findings from 15 non-directive counselling interviews with one client over a period of four months, 'made from the verbatim notes of the counsellor' (p. 78) and a follow up interview one year later 'electrically recorded' (p. 78). In addition to their qualitative analysis these authors conducted quantitative analysis using the Discomfort Relief Quotient (DRQ, Dollard & Mowrer, 1947) and showed in the early stages of therapy the client experienced more 'discomfort' and this reduced as therapy progressed. Here was some evidence to support Rogers' 1947 hypothesis that therapy led to a reduction in 'tension' and he referred to this finding in his 1959 paper as one of five research papers showing '... the reduction in psychological tension' (p. 219) as both outcome and process of non-directive therapy. The idea that 'outcome' and 'process', which at first sight appear distinct, although with further consideration are perhaps less distinguishable, is a recurrent one, e.g. Stiles (1996).

Rogers also referred to Virginia Axline's (1948) paper wherein Axline described the impact of the 'permissive' therapeutic relationship in helping children to come to terms with their own attitudes and emotions.

Victor Raimy completed his PhD with Rogers at Ohio in 1943 and published an article in 1948 based on records of 111 therapy sessions, 24 sessions from counsellor notes and 87 from electrical recordings from 11 therapists counselling 14 college students. Rogers used Raimy's detailed analysis of Raimy and four judges evaluating 874 and 356 client responses, respectively, in his 1959 paper as evidence that, amongst other things, client's self regard increased as both an outcome and process of therapy.

Schwebel and Asch (1948) evaluated two different relationship conditions that psychology students were exposed to with their teachers. Two classes received 'non-directive teaching' and one control group received 'usual teaching'. These authors found students in the non-directive condition did more reading of both set texts and outside texts, and those in the 'directive group' were found to be non-participatory in class and poorly adjusted to classroom situations. They concluded that '1. Non-directive teaching might encourage a greater drive towards maturity and self realisation. 2. The amount of work accomplished is related to the degree of freedom afforded to the student and more specifically to his readiness to accept his independence' (p. 363). Related to this research was that by Volney Faw who also published research on teaching relationships with psychology students and, along with the work of Schwebel and Asch, Rogers referred to Faw's work in his 1959 paper. Faw found students taught in a non-directive manner increased their amount of participation ('statistically significant at the one per cent level', p. 104-5) and had better grade point averages at the end ('statistically reliable at the five per cent level' p. 108). Rogers referred to Faw's research as suggestive that:

To the extent that education is concerned with learnings which significantly influence behaviour and facilitate change in personality, then the conditions of therapy and the conditions of an improving relationship apply. This leads, among other things, to more responsible basing of behaviour upon these perceptions (1959, p. 241).

The 'non-directive' school was becoming increasingly 'evidence-based'. Whilst the standards of research were not those which would necessarily pass peer review nowadays they were of an adequate standard then for acceptance by the major psychology journals. Clearly research methodologies have evolved since the 1940s and it is important to note that Rogers and his

colleagues were developing the techniques and methods which would be built upon for contemporary research.

In 1949 a 'special edition' of the Journal of Consulting Psychology (now the Journal of Consulting and Clinical Psychology) was published containing seven research reports by 'non directive authors'. These were based upon the same group of ten completely recorded and transcribed cases and became known as 'the parallel studies project' (Rogers, 1949). Barrett-Lennard referred to the fact of the 'special edition' in a major APA journal as the 'high point of the opening phase' in client-centred/non-directive therapy (1998, p. 236).

The non-directive authors conducted six different investigations and these were brought together in one paper (Raskin, 1949a), with an introduction (Raskin, 1949b). Seeman's paper was first in the journal and, leaving aside his other findings, Rogers later referred to this paper as one of seven references supporting 'the increase in the client's positive self regard' (1959, p. 217) as a consequence of non-directive therapy. Sheerer's work was also referred to by Rogers as supporting an increase in positive self regard as a consequence of non-directive therapy, although her analytical approach was quite different to Seeman. In addition Rogers referred to Sheerer's article as one of two papers supporting his assertion that as a consequence of non-directive therapy 'others are perceived in a more acceptant fashion' (p. 219). Rogers used Haigh's research as evidence that as a consequence of non-directive therapy there was a decrease in client defensiveness as both a process and an outcome. Hoffman and four judges analysed transcripts and Hoffman concluded that as a consequence of non-directive therapy the client's behaviour became more mature and this was later referred to by Rogers in his 1959 theory.

Rogers referred to a paper published the following year by Cowen and Combs (1950) as evidence that as a consequence of non-directive therapy, client '...adjustment is improved is supported by evidence based on TAT [Thematic Apperception Test], Rorschach, counsellor rating and other indexes' (1959, p. 219). In fact Cowen and Combs' 'Follow-up study of 32 cases treated by non-directive psychotherapy' was based upon pre and post tests using the Bernreuter Personality Inventory and their t-test analyses of the four subscales showed statistically significant improvements in neurotic tendency, introversion, confidence and sociability as a consequence of non-directive counselling. They also concluded their mixture of qualitative and quantitative analysis with their assessment that two important features of

the therapeutic relationship for successful outcome were ‘personal “warmth” of the counsellor and his ability to create a non-threatening atmosphere’ (p. 257).

Fred Fiedler completed a PhD in 1949 on ‘A comparative investigation of early therapeutic relationships created by experts and non-experts of the psychoanalytic, non-directive and Adlerian schools’. There was a subsequent paper in 1950 which Rogers referred to in both his 1957 and 1959 theory statements. Fiedler described himself as having ‘had some non-directive and psychoanalytic training but at present considers himself to be psychoanalytically oriented’ (1950, p. 437). He asked three ‘divergent’ judges to rate a sample of ten electrically recorded therapy sessions from a mixture of ‘experts’ and ‘non-experts’ in each of the three schools. Fiedler concluded:

1. Expert psychotherapists of any of the three schools create a relationship more closely approximating the Ideal Therapeutic Relationship than relationships created by non-experts within the same school.
2. The therapeutic relationship created by experts of one school resembles more closely that created by experts of other schools than it resembles relationships created by non-experts within the same school.
3. The most important dimension (of those measured) which differentiates experts from non-experts is related to the therapist’s ability to understand, to communicate with, and to maintain rapport with the patient (p.444).

Fiedler noted he used ‘expertness’ as a proxy for ‘effectiveness’ and these were opinions of judges, not quantitative measures of what was effective. To some extent it would appear Fiedler held an idea of integrative therapy and he noted if one therapy could succeed without a component considered essential to another therapy then that component was not essential for effective outcome. He was interested in finding what component(s) were essential for effective therapy and held an idea of the ‘drug metaphor’ (cf. Stiles & Shapiro, 1994); more of an essential component must produce proportionally better outcomes. His research convinced him ‘relationship is therapy, that the goodness of therapy is a function of the goodness of the therapeutic relationship... this does not necessarily mean, however, that the relationship alone can lead to eventual cure’ (p. 443, emphasis original).

Rogers referred to Fiedler’s work, along with that of Quinn (1950) who undertook a similar exercise, as confirmation of Rogers’ six conditions of the therapeutic process (in both the 1957 and 1959 works) and particularly for the role of empathy in the 1959 paper (p213-5). In the 1957 paper Rogers also referred to Fiedler’s work:

That such penetrating empathy is important for therapy is indicated by Fiedler's (1950) research in which items such as the following placed high in the description of relationships created by expert therapists: The therapist is well able to understand the patient's feelings; the therapist is never in any doubt about what the patient means; the therapist's remarks fit in just right with the patient's mood and content; the therapist's tone of voice conveys the complete ability to share the patient's feelings (p. 99).

In 1951 Rogers published 'Client-centred therapy: Its current practice, implications and theory'. Whilst this did not present any new research findings, this did set out Rogers' evolving thinking about his theory. Rogers saw the non-directive therapeutic relationship as having an impact upon the client so the client came to realise 'that he is responsible for himself in this relationship' (p. 71). Importantly Rogers referred to the 1951 book as being a description of the theory and practice of client-centred therapy in his 1957 theory statement. The context for this was to position the 1957 theory as a statement of therapy integration and Rogers pointed to client-centred therapy, as he had described it in 1951, as being just one of a number of the various therapies, and techniques of various therapies simply served as 'a channel by which the therapist communicates a sensitive empathy and an unconditional positive regard... by which the essential conditions of therapy are fulfilled' (1957, p. 102). The 1957 theory was to some extent an idea of the integration of the various different schools and their therapeutic relationship 'common factors' as the cause of effective therapy outcome.

In terms of the chronology of this process, it was in 1952 that Eysenck stirred up a huge amount of controversy when he wrote that following his evaluation, published in the APA Journal of Consulting Psychology, that the effects of psychotherapy were 'unproven'. This created a stir within the therapeutic community and much subsequent research, some of which is described below, to consider the effects of psychotherapy.

By 1954 Rogers was ready to report upon a large scale research programme that had been underway at Chicago for over four years (Rogers & Dymond, 1954). The format chosen for this was a book, rather than peer-reviewed journal articles and this absence of formal peer-review may not have helped the subsequent development of client-centred therapy; see also Rogers' 1967 research findings were also published in book form. The chapters of the 1954 book were authored by a number of researchers, in different combinations, as they described what they had found from analysing a number of fully transcribed courses of therapy. The

authors wrestled with what, for the field, were new issues that contemporary researchers would recognise and probably now take for granted: client inclusion/exclusion criteria, counsellor inclusion/exclusion criteria, appropriate outcome measures, control groups, criterion to structure control groups by, pre and post outcome measures, when to test, process measures, validity, reliability, analytical techniques to be used, confidence intervals, rival hypotheses, etc.

Some of the approaches taken by Rogers and his team were what Cook and Campbell (1979) would later call ‘quasi-experimentation’; experiments that had treatments, outcome measures and experimental units but did not or could not have random assignment. Lack of random assignment meant researchers had to work harder at interpreting the results, in terms of separating the effects of treatment from all possible threats to internal validity.

One of the more extraordinary issues described by Rogers and Dymond was that of ‘success criteria’ and the radically ‘honest’ approach taken by these researchers probably did not serve them well in the long-term. The team discussed how they would determine ‘effectiveness’, what measures would demonstrate ‘success’ or ‘failure’. The prevailing contemporary model (e.g. NICE, APA approach described above) was to determine a ‘medical condition’ (e.g. DSM-IV-TR diagnosis, American Psychiatric Association, 2000) and *evaluate* whether therapy has an impact upon improving this condition, perhaps compared with a control group, to which clients have been randomly assigned. What Rogers and his team did was so extraordinary as to be worth quoting at length:

Our thinking often took the form of discussing what measures would demonstrate the ‘success’ or ‘failure’ of therapy. Certainly this is the criterion which occurs to most people when they think of studying psychotherapy. From this point of view, psychotherapy is conceived as something which makes people ‘better’ or ‘adjusted’, and hence the therapy is successful or unsuccessful in achieving this aim. Or it is conceived as a ‘cure’ for ‘mental illness’, and research then becomes involved in ambiguity piled upon ambiguity, in which the question is whether a mythical entity has or has not been removed. The consequence of this use of criteria based upon value judgements has been that each investigator endeavours to prove that therapy does produce certain changes which have value to him, a rather unsatisfactory basis for science. The fact that there are various more or less competitive therapeutic orientations still further complicates this matter of using selected definitions of success (1954, p. 28).

To some extent Westen, Novotny and Thompson-Brenner (2004) provided something of a contemporary discussion of this and related issues. On the basis that 'it is quite impossible at the present time to define 'success' or 'adjustment' in such a way that the definition is both operationally clear and acceptable to all' (1954, p. 29) Rogers' team simply did away with 'success criteria' and decided instead to document what changes did or did not occur as a consequence of client-centred therapy. In a radically market-oriented approach their idea was that clients could then decide, on the basis of the evidence as presented and with their own criteria, whether they thought client-centred therapy was what they wanted. Norris (1990) differentiated 'research' and 'evaluation' and on the basis of his descriptions, strictly what Rogers and team were engaged in was 'research' (seeking to find out what happens) as opposed to 'evaluation' (finding out if a particular outcome occurs).

It is because of this that the 1954 findings were expressed in terms of changes in the self concept, changes in psychological tension, changes in psychological adjustment, changes in personality and behaviour, the reorganisation of self, changes in attitudes towards others and so on, rather than improvements in anxiety, depression, obsessive-compulsive disorder, personality disorder and so on. Whilst there was certainly an 'honesty' to science and the strict focus upon research of this endeavour, perhaps this took client-centred therapy down a 'blind alley', as opposed to, for example Beck's (e.g. 1976) background in medicine and his approach to evaluating outcome based upon diagnostic criteria.

To the extent that Rogers and team considered 'outcome' there was certainly evidence of effective outcome as a consequence of client-centred therapy. There was also a comparative outcome study (ch. 12) where it was reported that there was no difference in effectiveness between Sullivanian and Rogerian therapies. This chapter described the difficulties of working with what were referred to as 'ethnocentric' clients and it seemed plausible that this research was tapping similar attitudes and beliefs to some personality disorders e.g. anti-social personality disorder (Beck, Freeman, Davis, & Associates, 2004). Contemporary literature recognises the added complications of work with clients with personality disorders and the impact this might have on, along with other client variables, on outcome e.g. Clarkin and Levy (2004). Whilst Rogers was in possession of this kind of information, the impact of client variables upon outcome, in 1954, he appeared to ignore this and the impact of client variables played no part in his 1957 or 1959 theory statements. Rogers was perhaps reluctant to 'give up' on those who might not be helped or were more difficult to work with and

seemingly does not follow through on the questions posed about determining which clients might best profit from therapy (1954, p. 214).

The 1954 volume was concluded by:

In our judgement the research sets forth for the first time objective evidence that one defined approach to psychotherapy produces certain measurable and significant changes in the individual coming for help and that certain other changes which have also been hypothesised failed to occur in significant degree (p. 433).

In terms of the research questions for this thesis the 1954 volume provided evidence of the effectiveness of client-centred therapy, although not in terms contemporary psychologists would likely find particularly helpful or interesting and there is precious little about the impact of the therapeutic relationship. In their critique of the evidence for whether 'therapist interpersonal skills' had an impact on outcome, Lambert, DeJulio, and Stein (1978) wrote 'these early crude studies... generally did not involve the specification of therapist behaviour and its differential relation to outcome... Rogers and Dymond (1954) did not look at therapist interpersonal skills but concentrated on the overall effects of treatment' (p. 468). Beyond specifying 'client-centred therapy' was offered there was no greater specification of what the treatment was, what process was going on in sessions, or what might now be termed 'adherence to treatment protocol'. There was little attempt to consider causation in the sense of defined process leading to, or at least correlating with, defined outcome.

Rogers was able to provide evidence for changes as a consequence of client-centred therapy in both his 1957 and 1959 theory statements based on the 1954 work. In the 1957 paper Rogers had evidence to substantiate such claims as changes in the personality structure of the individual at both surface and deeper levels, greater integration, less internal conflict, more energy utilisable for effective living, change in behaviour from immature to mature and so on (p. 95). In the 1959 paper Rogers pointed to changes such as the concept of self becomes reorganised to assimilate and include experiences previously denied to awareness (p. 216), new and emergent self perceptions (p. 217), the increased congruence of self and ideal, the self-ideal becomes more achievable, concept of self improves and psychological adjustment is improved (p. 219), the proportion of behaviours which can be owned as belonging to self increases, the proportion of behaviours disowned as self decreases, behaviour becomes increasingly perceived as being more within control (p. 220), greater maturity in behaviour (p. 220), both the clients and friends notice differences in terms of the way a person drives a

car makes choices, behaves in group discussion, treats other people, and so on (pp. 236-7). Unfortunately it is not so clear what these changes amounted to in terms of patients with diagnosed conditions that a physician might want to refer to a client-centred therapist, nor was it clear, what if anything the therapist had done, or had to do again, to accomplish or at least facilitate these kinds of changes.

In 1954 Chodorkoff published, based on his PhD with Rogers, a paper that Rogers used in his 1959 paper to substantiate his argument that more congruent clients are less defensive and better psychologically adjusted (p. 202); and because of this changes in congruence and self-concept are important outcomes, which client-centred therapy was effective at accomplishing. In his 'integrative theory' (1957) Rogers refers to Chodorkoff's work as evidence clients were incongruent, vulnerable or anxious (p. 96) and went on to further define these terms (p. 97). Another paper published in 1954 by Hanlon, Hofstaetter, and O'Connor was referred to by Rogers in his 1959 theory as evidence that self-ideal congruence was related to level of psychological adjustment (p. 219) and by implication because self-ideal congruence changed with client-centred therapy, then it seemed likely that psychological adjustment also changed with client-centred therapy.

2.2 Main statements of Roger's theories and Rogers' attempts to validate his theories.

In 1957 Rogers published what in contemporary terms might be called his integrative theory based on common factors. Rogers explained to Hart (in Hart and Tomlinson, 1970) the 1959 'theory of therapy' paper published in 1959 was written in 1953-4 and preceded the 'necessary and sufficient conditions' paper published in 1957 (Wyatt, 2001, p. ii). This temporal precedence is important and the dates are unfortunately confusing. Importantly the 1959 paper was 'a' theory of therapy and specifically related to 'the client-centred framework', whereas the 1957 paper was an integrative theory about all types of psychotherapy.

At the centre of Rogers' 1957 and 1959 theory papers was the proposition that six conditions were 'necessary and sufficient' for 'constructive personality change to occur'. Wyatt (2001, p. ii) pointed out the minor differences in the 1957 and 1959 descriptions of the six

conditions. Rogers expressed these in the 1959 paper (written in 1953-4) as follows, with his later 1957 amendments shown in italics:

1. That two persons are in (*psychological*) contact.
2. That the first person, whom we shall term the client, is in a state of incongruence, being vulnerable or anxious.
3. That the second person, whom we shall term the therapist, is congruent (*or integrated*) in the relationship.
4. That the therapist is experiencing unconditional positive regard toward the client.
5. That the therapist is experiencing an empathic understanding of the client's internal frame of reference (*and endeavours to communicate this to the client*).
6. That the client perceives, at least to a minimal degree, conditions 4 and 5, the unconditional positive regard of the therapist for him and the empathic understanding of the therapist. (*The communication to the client of the therapist's empathic understanding and unconditional positive regard is to a minimal degree achieved*).

Rogers' theory was that the provision of congruent empathy and unconditional positive regard, as he defined these, by the therapist enabled clients to undergo 'constructive personality change' (again, as defined).

At the centre of Rogers' theory was the client who 'is in a state of incongruence, being vulnerable or anxious' (Rogers, 1957, p. 96). This incongruence, vulnerability or anxiety arises because of 'a discrepancy between the actual experience of the organism and the self picture of the individual, insofar as it represents that experience' (Rogers, 1957, p. 96). The theory was that by genuinely empathising with the client and unconditionally accepting the client's viewpoint the client gradually came to understand and accept 'the actual experience of the organism' cf. 'the self picture of the individual'. It is important to note that Rogers uses the term 'anxiety' in a particular way and this was precisely defined as 'phenomenologically a state of uneasiness or tension whose cause is unknown. From an external frame of reference, anxiety is a state in which the incongruence between the concept of self and the total experience of the individual is approaching symbolisation in awareness... Anxiety is the response of the organism to the "subception" that such discrepancy may enter awareness' (Rogers, 1959, p. 204). This definition of 'anxiety' is different to the way in which this term is now commonly used. For example the Diagnostic and Statistical Manual (DSM-IV-TR)

lists at least thirteen types of 'anxiety disorders' (American Psychiatric Association, 2000, p. 429). Rogers' precise use of 'anxiety' as a kind of 'cognitive crisis' is different to the current use(s) of the term.

Following the statements of his theory published in 1957 and 1959, Rogers and colleagues went on to try to test the theory with people hospitalised for a diagnosis of schizophrenia at Mendota State Hospital with the sponsorship, encouragement and financial assistance of the Wisconsin Psychiatric Institute (Rogers 1967). Known as 'the Wisconsin project' some 200 staff members worked on the project at a total cost estimated to be of the order of £4.6m in 2003 terms (Weston, 2005, p. 26).

This was a huge study working with a difficult client group in a difficult setting and, as with the work published in 1949 and 1954, the team were creating outcome measures and research methodologies that contemporary researchers would take for granted. The study created conflict and splits within the research team, that nearly ended up in a court case (Kirschenbaum, 2007). Writing about 'the bucket theory of containment and displacement', Hawkins and Shohet (2006) made the point that:

All helping organisations are, by their very nature, importing distress, disturbance, fragmentation and need. This is usually met by individual workers, who, if they are empathically relating to the client's distress, will experience parallel distress and sometimes disturbance and fragmentation within themselves. How much of this they will be able to contain and work through will depend on the size of their emotional container (or bucket), and will relate to their personality, their emotional maturity and professional development, the amount of pressure and stress they are currently under at work and at home and, most importantly, the quality and regularlity of the supervision they receive. What is not contained at this level will lead to decreased functioning in the worker and can also lead to fragmentation in the team, This comes about as those who are stressed quite often act out this stress on their colleagues. They can get irritable with the secretary, angry with their boss and non-cooperative with their colleagues. Fights can develop about who is responsible for what, and arguments over duty rotas. Team meetings begin to start later and later and become more fractious (p. 183).

Looking back it seems plausible that at least some of the 'disturbance' experienced by Rogers et al. was a consequence of 'under supervision'.

As with the 1954 book, publishing the Wisconsin findings in book form took away the external peer review process, which may have been to the detriment. The 1967 book is very difficult to read and has a number of omissions that reduce its credibility as a supposed scientific report of a research study.

Rogers sought to test six main hypotheses and amongst many non-positive findings Rogers and his team found some evidence that people with a diagnosis of schizophrenia receiving client-centred therapy had a small and statistically significant better outcome in terms of constructive personality changes and hospital release than matched patients receiving treatment as usual; and that ‘... the greater the degree to which the conditions of therapy existed in the relationship, the greater... the evidence of constructive outcome’ (p. 91). To that extent this appeared to be some evidence of the clinical effectiveness of client-centred psychotherapy and the impact of the therapeutic relationship. The absence of peer review and the sheer difficulty of making sense of this undertaking from the book suggest some caution is perhaps appropriate with these findings (and see later e.g. Gurman, 1977).

Subsequent research suggested ‘counselling’ can be effective for people with a diagnosis of schizophrenia e.g. Tarrier, et al. (2000).

After his experience with the Wisconsin project Rogers effectively ceased to be a researcher and found considerable success with the publication of several ‘populist’ psychology books, the first of which was published prior to the Wisconsin project on schizophrenia (1961) and others subsequently; Rogers and Stevens, 1967, Rogers, 1978 and Rogers, 1980. Thorne (2003) described how Rogers moved on from Wisconsin:

It could justifiably be claimed that the powerful desire to be more influential which took Rogers to Wisconsin was in no way fulfilled by his daily work there. Yet it was his fifth book, ‘On becoming a person’, published in 1961 that, almost overnight, catapulted him into the limelight and brought him more fame and influence than he could ever have hoped for. The book broke him free from the professional world of psychology... he [Rogers] set out for La Jolla in California to join WSBI, a non-profit making organisation concerned chiefly with humanistically oriented research in interpersonal relationships (pp. 17-8).

Once Rogers was in California he enjoyed celebrity status and devoted his time to the encounter movement and world peace, never again venturing to prove his theories.

Later Rogers described the Wisconsin project as ‘without doubt the most painful and anguished episode of my whole professional life’ (in Burton, 1972, p. 62 quoted in Thorne, 2003, p. 17).

By the 1980s Rogers appeared to have changed his perspective on research. In his 1985 article Rogers made the case for a ‘new science’, no longer constrained by the ‘straitjacket of logical empiricism’. Thorne (2003) attempted to put this into the context of Rogers’ life and described this as:

No longer is he content to pay even lip service to the supremacy of the conventional view of science, the Newtonian, mechanistic, linear cause-effect understanding of reality. He does not throw it out but considers it singularly inappropriate for exploring the questions that now need to be addressed in the psychotherapeutic relationship where living human persons deserve to have researchers who are prepared to commit themselves to their studies in a way that enhances the dignity of everyone involved (p. 63).

In identifying ‘some common elements’ of this ‘new science’ based around Polanyi’s (1958) philosophical work Rogers praised the work of Mearns and McLeod (1984):

....there are no longer ‘subjects’ of research, but ‘co-researchers’, ‘research partners’, ‘participants’. Mearns and McLeod (1984) carry this to an extreme. In their paper they advocate having these research partners involved in every step of the study – the planning, the data gathering, the analysis, the interpretation, the conclusions. They make psychological science a cooperative enterprise in which everything is above board, a participatory endeavour (Rogers, 1985, in Kirschenbaum & Land Henderson, 1990, p. 285)

Rogers goes onto to praise the research of other key figures in the person-centred community. Looking back it seemed the whole direction of person-centred research for the next twenty years and more is encapsulated in Rogers’ 1985 paper with its encouragement of qualitative methods and the endorsement of particular figures. Yet, Rogers also states ‘...the conventional methods are not to be thrown out, but they are often inappropriate for questions we wish to study’ (1985, in 1990, p. 281). Compared with Thorne’s emphasis, above, Rogers says ‘the Newtonian, mechanistic, reductionistic, linear cause-effect, behaviourist view of science is not thrown out but it is seen as simply one aspect of science, a perfectly good way of investigating certain questions, but decidedly inappropriate for others’ (1985, in 1990, p. 284).

In addition to the recent biography mentioned above (Kirschenbaum, 2007), other books have detailed the work of Rogers (Kirschenbaum & Land Henderson, 1990), his life (Thorne, 2003) and the interplay between the two (Barrett-Lennard, 1998).

2.3 Contemporary person-centred psychotherapies.

The approach developed by Rogers has been adapted in a number of different ways (Sanders, 2004), commonly termed ‘the tribes of the person-centred approach’ and this includes tribes known as ‘classical’, ‘focusing’, ‘experiential’, ‘existential’ and ‘integrative’. It is important to note that person-centred therapists are unified in their idea of the centrality of the therapeutic relationship (Zuroff & Blatt, 2006) and unlike say, CBT (Beck, 1976), provision of the relationship conditions is ‘principled’ rather than ‘instrumental’ (Grant, 1990), i.e. the relationship conditions are not offered as an ‘instrument’ of ‘seeking to be effective with clients’ but from the ‘principle’ of ‘how to be with clients’. In addition to the largely historical quantitative research on ‘classical’ person-centred therapy the process-experiential ‘tribe’ has recently received the most quantitative research attention e.g. Watson (1996), Elliott, et al. (2004), Watson, et al. (2003), Watson and Bedard, (2006). This thesis is titled to encompass these research findings, hence ‘person-centred psychotherapies’.

In later years Rogers became interested in the concept of ‘emotional processing’ during counselling, particularly because of the influence of one member of his research team, Eugene Gendlin (1962, 1978 and 1996). To study ‘emotional processing’ in therapy the ‘experiencing scale’ was developed (Klein, Mathieu, Gendlin, & Kiesler, 1969). In addition to the importance of the therapeutic relationship, what ‘goes on’ in therapy was also considered important, in terms of emotional processing, although Rogers’ interest in this was subsequent to his main theoretical statements (1957 and 1959).

There has been considerable interest in emotional processing in therapy within the person-centred movement, for example: Greenberg, Rice, and Elliott, (1993), Watson, (1996), Greenberg, (2002), Watson, et al., (2003), Goldman, Greenberg, and Pos, (2005), Missirlian, et al., (2005) and Watson and Bedard, (2006). Beyond person-centred therapy, emotional processing has also been found to be an important predictor of change in CBT (Castonguay, Goldfried, Wiser, Raue, & Hayes, 1996, Foa & Kozak, 1991), although there is some

evidence that suggests the actual (naturalistic) practice of CBT may discourage emotional processing (Malik, Beutler, Alimohamed, & Gallagher-Thompson, 2003).

Beyond person-centred notions of therapeutic relationships and emotional processing there are ideas about assimilation of problematic experiences (Stiles, 2002), configurations of self (Mearns, 1999), relational depth (Mearns & Cooper, 2005) and other concepts (Mearns & Thorne, 2000). In particular person-centred therapists view clients as 'active self healers' (Bohart & Tallman, 1999, Bohart, 2007); Rogers' 1959 idea of 'the actualising tendency'. Whilst Rogers' main theoretical statement was about the therapeutic relationship it is important to recognise that 'person-centred' isn't 'just' about congruent empathy and unconditional positive regard.

As noted above Rogers (1985) apparent volte face with quantitative research appeared to have influenced many/most of the subsequent generation of person-centred researchers, with some notable exceptions, many of whom are named as authors in the following section. The next section places this research in the context of what is known about outcomes and process-outcomes findings and methodologies.

3. Placing this research in the context of the literature

Having introduced person-centred psychotherapies this section now reviews what is known about outcomes, process-outcomes and research methodologies as related to the two main research questions: Does person-centred therapy work (outcomes study) and if it does work, does this have anything to do with the therapeutic relationship as defined by Rogers (process-outcomes study).

This section is arbitrarily divided into three sections; outcomes, process-outcomes and methodologies. This is for structural convenience, since many researchers have made contributions to outcomes, process-outcomes and methodologies in the same publication. Consequently there is some overlap between these areas in the following section and this done to avoid repetition, for example reporting the same study three times in three different sections. Generally studies were placed into one or other section on the basis of the focus this discussion would take. In general terms studies are placed in temporal order as this illustrated the evolution of methodologies over time. Exceptions are made when a study is linked to other studies, when it would be disruptive to stick to strict temporal order. For example Stiles and colleagues made huge contributions to process-outcomes studies over a number of years and these papers are reviewed together, rather than interrupting the flow by referring to other important works that occurred during this period.

The following three sections on outcomes, process-outcomes and research methodologies seeks to place the experimental part of this research in an appropriate context of what is already known.

3.1 Outcomes literature

This section reviews outcome literature as this related to the first of the research questions; the clinical effectiveness of the person-centred psychotherapies. This makes the case for this research, since researchers typically find the main schools of therapy have approximately equivalent outcomes. It was the intention of the outcomes part of this research to benchmark outcomes against a sample of studies from the literature and these studies are described.

Firstly the ‘anxiety’ studies are described, then the ‘depression’ studies and finally the ‘distress’ studies. These are in approximately temporal order. Along the way some methodological points are made, particularly as these arise in the third of these areas, the distress benchmark studies because these were naturalistic studies and this gave rise to a dialogue that helpfully articulated some of the key points to do with naturalistic research. Accordingly the distress benchmark studies section develops to discuss some methodological points as these related directly to this research.

The case for this research was further made by a consideration of the therapy workforce in the UK and the UK investment in therapy research. Despite recommendations for CBT, a large number of UK therapists are not trained in CBT. The recommendations for CBT are perhaps related to a massive financial investment in UK research into CBT.

This research was needed to make the case for non-CBT therapists providing therapy, in particular the person-centred therapists who make up the largest proportion of the BACP ‘workforce’. Furthermore, this research was needed because significantly less financial investment in the UK has been made in person-centred therapy, yet, as will be seen in the subsequent process-outcome section, there is a substantial literature to suggest the therapeutic relationship/alliance is related to outcome.

Finally in the literature review of outcomes, a brief description of outcomes management is provided. Through the course of this research the present author has come to practise in this way and this maybe of relevance to the interpretation of the results. Furthermore, outcomes management is shown to favourably impact therapy outcome.

3.1.1 Overview of outcomes literature

The preceding section on person-centred psychotherapies mentioned some of the key publications in the period that Rogers was researching non-directive and subsequently client-centred therapy, to maintain some sense of temporal order. These were Rosenzweig's (1936) 'Dodo bird' paper and Eysenck's (1952) claim that there was no evidence for the effectiveness of therapy. This section gives an overview of some of the outcomes literature. This research sought to 'benchmark' the outcomes from person-centred psychotherapy with some of the studies for depression, anxiety and distress. The benchmarking studies are reviewed in separate sections, subsequent to this overview.

Una Maguire (1973) critically examined studies of counselling effectiveness, referring back to what she referred to as Eysenck's 'unproven' claims. She found that many studies suffered with methodological weaknesses, problems of interpretation, problems with outcome measures, problems of over-interpretation of findings and concluded that there was little evidence for any positive effect from counselling beyond 'he would have grown out of it anyway' (p. 48).

Luborsky, Singer and Luborsky (1975) reviewed about forty controlled outcome studies and concluded that, in contrast to Eysenck's claims, psychotherapy was generally effective and that there was empirical support for Rosenzweig's point about the 'Dodo bird verdict'; different schools of therapy led to approximately equivalent outcomes.

Mary Lee Smith and Gene Glass (1977) reviewed nearly four hundred controlled studies of therapy and concluded there was 'convincing evidence of the efficacy of psychotherapy' (p. 752) and that the typical client was better off than 75% of untreated individuals. Smith and Glass looked at ten types of therapy (psychodynamic, Adlerian, eclectic, transactional analysis, rational-emotive, gestalt, client-centred, systematic desensitisation, implosion therapy and behaviour modification) and concluded that 'despite volumes devoted to the theoretical differences among different schools of psychotherapy, the results of research demonstrate negligible differences in the effects produced by different therapy types' (p. 760).

In 1999 Luborsky, et al. published their article on 'The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy'. These researchers had found empirical evidence of a 'researcher allegiance effect', wherein researchers comparing their own favoured therapy with other therapies tended to find their favoured therapy did best. NICE

did not control for ‘the researcher allegiance effect’ or for ‘the reviewer allegiance effect’ (Scriven, 1998). It was certainly the case that not all authors and interest groups agreed with NICE recommendations, see for example interested parties comments (NICE, 2009b) on the updated depression guidelines (NICE, 2009a).

In 2001 Teusch, Bohme, Finke, and Gastpar published research on 142 inpatients with personality disorders, together with other problems such as depression, anxiety and eating disorders. Significant improvements in depression, self esteem and social adjustment were noted at discharge and 12 month follow-up. Treatment was client-centred therapy with or without anti-depressants. Client-centred therapy was found to be superior to the ‘with medication’ condition for socially deviant, emotionally unstable/borderline and histrionic/narcissistic subgroups in the reduction of depression whereas medication enhanced outcome in the socially dependent subgroup.

In 2004 NICE published reports on anxiety (NICE, 2004a) and depression (NICE, 2004b), recommending, amongst other things, that the psychological treatment of choice was CBT and that there should be no more ‘counselling’ for anxiety on the NHS and that patients with ‘mild to moderate depression’ could be offered ‘counselling’ on a time-limited basis but not for ‘severe or recurring depression’. Part of the rationale for these recommendations was based on literature reviewed below. In terms of NICE recommendations for anxiety some of the journal articles cited as evidence for the recommendations included Borkovec and Whisman (1996), Gould, et al. (1997), Bryant, et al. (1998), Bryant, et al. (1999) and Barrowclough, et al. (2001) as evidence of the superiority of CBT over ‘counselling’. The NICE recommendations for depression included reference to Ward, et al. (2000) as evidence that ‘counselling’ could be an effective intervention for mild-to-moderate depression but not for ‘severe depression’ because Ward et al. did not have many patients with ‘severe’ depression. In considering these research questions it seemed important to research both depression and anxiety as NICE wanted research on these symptoms, with recognised measures. Whilst it might seem that the case had been made that person-centred therapy was an effective intervention for depression, based on Ward et al., this study had not satisfied the need for information about ‘severe’ depression. Furthermore, NICE (and the APA) were keen on replication in a number of studies, such that something akin to a meta-analysis could be performed. For example the APA’s Chambless and Hollon (1998) criteria required replication across independent research settings. These were factors taken into consideration in the construction of the research methodology for this research.

In contrast to the NICE recommendations, Elliott et al (2004) conducted a meta-analysis of research on experiential psychotherapies (including person-centred) and concluded there was evidence of effectiveness for depression, anxiety disorders, helping clients deal with traumatic and abusive events, marital distress and to a lesser extent there was also evidence of effectiveness for problems related to anger and aggression, especially domestic violence, severe client dysfunction, including schizophrenia and severe personality disorders plus various health related problems including psychosomatic problems, HIV and cancer (improved psychological wellbeing and substantially longer survival times in terminal illness). These authors also found evidence of comparable efficacy/effectiveness when compared with other schools of therapy. They analysed 127 therapy samples in 112 studies (6,569 clients) and found on average a 'large' $ES(d) = .99$ effect size (see below) which reduced to .86 when weighted by sample size due to two large studies with relatively low effects. In making judgements about the levels of empirical support for these therapies the authors used the Chambless and Hollon (1998) criteria.

For anxiety the mean pre-post effect size was 1.30 'large' and 'possibly efficacious' in terms of Chambless-Hollon criteria because the requirement for replication across independent research settings was not satisfied. Once researcher allegiance effects were controlled for there was no significant difference in outcomes for experiential therapies as compared with CBT.

In terms of 'trauma and abuse' the mean pre-post effect size was 1.15 'large' and the Chambless-Hollon criteria were satisfied for an efficacious and specific treatment.

For 'depression' the mean pre-post effect size was 1.18 'large' and the Chambless-Hollon criteria were fulfilled for a 'specific and efficacious' treatment.

For treatment of anger and aggression the mean pre-post effect size was .96 'large'.

For schizophrenia and severe, chronic dysfunction the mean pre-post effect size was .88 'large' for mixed inpatients, .80 'large' for schizophrenia, and 1.33 'large' for severe personality disorders and 'possibly efficacious' in terms of Chambless-Hollon criteria.

For health-related problems (cancer, HIV, rheumatoid arthritis and psychosomatic problems the mean pre-post effect size was .59 'medium' and 'possibly efficacious' in terms of Chambless-Hollon criteria.

The Elliott et al chapter appeared in Lambert (2004) and in the same volume, Lambert and Ogles' (2004) reviewed the efficacy/effectiveness literature and questioned whether one treatment was preferable to another. They concluded the available evidence was that differences between treatments were generally small and confounded by a number of different methodological difficulties (e.g. researcher allegiance) that meant in practice there was unlikely to be much 'real' difference in outcomes between bona fide treatments.

In 2004 Westen et al. published a critical review of the assumptions and findings used in randomised controlled trials (RCT) to establish psychotherapies as 'empirically supported'. They suggested that whilst RCT methods maybe used to draw accurate inferences such methods such may also be 'misused', particularly where money, power and prestige were at stake. Amongst many other criticisms these authors pointed to over-simplifications that they suggested risked ceasing to represent 'real world' psychotherapy in laboratory research because of research over-simplifications. They further suggested these over-simplifications could also be a result of researcher allegiance effects and some of the highlighted risk areas included 'the therapy', (e.g. brief, manualised therapy of fixed duration cf. how most clients access therapy in the community), 'the therapist' (risking either uncontrolled therapist effects or therapist allegiance effects), 'the client' (e.g. single diagnosis of a readily measured 'problem' with no comorbidity cf. most 'real world' clients), client selection processes (e.g. failure to accurately define client numbers included and excluded), the presentation and reporting of statistics (e.g. omitting long-term follow-up) and the 'transportability' of RCT findings into 'the real world' (high internal validity cf. low external validity, see below). One potential solution these authors pointed to was using alternatives to RCT designs, such as 'using practice as a natural laboratory' which is the foundation of this research study.

Unsurprisingly, what could be considered a long and detailed 'attack on the EST establishment' (the article is an unusually long 32 pages) led to much counter-argument, see for example Ablon and Marci (2004), Crits-Christoph, Wilson, and Hollon (2005), Weisz, Weersing, and Henggeler (2005) and the reply of Westen, Novotny and Thompson-Brenner, (2005).

In 2007 Siev and Chambless published research evidence that they claimed was a 'rebuttal' (p. 520) of previous claims in the literature in favour of the 'Dodo bird verdict'. Their meta-analysis compared the relative efficacy of cognitive therapy (CT) and relaxation therapy (RT) for generalised anxiety disorder (GAD) and panic disorder without agoraphobia (PD) found

that for GAD, CT and RT were equivalent and that for PD, CT performed significantly better, across all outcome measures.

Minami, Wampold, Serlin, Kircher, and Brown (2007) published benchmarks for psychotherapy efficacy in adult major depression based upon an extensive meta-analysis. Overall they found a mean effect size, $ES(d) = 1.85$ for completers and identified some moderators of this overall effect, e.g. samples starting with mean depression levels lower than that for the overall study would tend to have lower effect sizes. The mean effect size for non-completers was found to be $ES(d) = 1.70$ on a last observation carried forward basis and for a non-treatment control group the natural history of untreated depression was found to be a small improvement of mean effect size $ES(d) = .37$. Subsequently Minami et al. (2008) showed how their benchmark study could be used with non-central t methodology to assess the effectiveness of psychotherapy treatment for adult depression in a particular managed care environment.

In his 2008 book on 'essential findings in counselling and psychotherapy' Mick Cooper juxtaposed a list of 'empirically supported treatments' with a reiteration of the equivalence of schools 'Dodo bird' findings and made the point that there was evidence both 'sides' of the debate could point to.

In a 2008 update of Elliott et al. (2004), three Strathclyde-based colleagues Elliott, Freire and Cooper presented further evidence of empirical support for person-centred and experiential psychotherapies. At the same conference Hill and Brettle (2008) presented a systematic review of counselling in primary care and concluded that patients were generally highly satisfied with counselling they received in primary care and counselling was as effective as CBT with typical heterogenous primary care populations. The present author presented an early version of the outcomes part of this research (Weston 2008a) at the same conference and at a subsequent one in the same year (Weston 2008b).

During 2009 the draft revised NICE guidelines for depression were circulated (2009a). In contrast to the 2004 guidelines wherein Ward et al. (2000) was evidence for the effectiveness of person-centred therapy with mild-to-moderate (but not severe) depression, it was decided that 'counselling' should no longer be recommended by NICE, on the basis that there had been a large investment in IAPT, and instead warnings should be made that 'counselling' was of unknown efficacy. An early version of the depression findings from this study was submitted to NICE (2009b) and rejected as a legitimate form of evidence because it seemed

to the reviewers that only one therapist had provided the therapy, cf. NICE (2004a) guidelines on anxiety and reliance placed upon Barrowclough, et al. (2001) which had only one therapist in the ‘supportive counselling’ condition as evidence of the ineffectiveness of counselling. Furthermore NICE (2009b) rejected any evidence for depression outcome based upon CORE-OM on the basis that CORE-OM was not specific to depression, cf. Gilbody, et al. (2007). This suggested that research of the kind proposed by these research questions should incorporate a number of therapists and utilise diagnostic-specific measures of the kind recognised by NICE.

Having given an overview of some of the outcomes research the following sections review the outcomes studies used to benchmark the outcomes from the experimental part of this thesis; anxiety, depression and distress outcomes. These outcomes are described in this order as this mainly reflects the temporal order of these studies.

3.1.2 Anxiety outcomes

A number of the papers referred to in the NICE reviews of anxiety-related conditions are briefly reviewed below, to the extent that they impact on the research question of this thesis; what is the clinical effectiveness of the person-centred psychotherapies.

Borkovec & Whisman (1996) reviewed eight studies for generalised anxiety disorder in a meta-analysis and they found that CBT was superior to no treatment and pill placebo conditions. They found that CBT failed to demonstrate any superiority over ‘non-specific’ treatments such as the ‘non-directive therapy’ reviewed in their meta-analysis. The implication for this research was that it was possible that person-centred therapy could be an effective intervention for anxiety, perhaps on a par with CBT, suggesting the need for further research on anxiety outcomes from person-centred therapy.

Building on the work of Borkovec & Whisman (1996), Gould, Otto, Pollack and Yap (1997), reviewed CBT-based studies that included a control group in a meta-analysis for generalised anxiety disorder. They concluded that CBT ($ES(d) = .70$) was as effective as pharmacotherapy ($ES = .61$) and reiterated the findings of Borkovec & Whisman that CBT failed to demonstrate any superiority over non-specific treatments such as non-directive therapy. Again this suggested the need for further research on anxiety outcomes from person-

centred therapy and an assessment of the impact of the therapeutic relationship for anxiety outcomes.

Bryant, Harvey, Dang, Sackville, & Basten (1998) compared CBT with 'supportive counselling' following civilian trauma (road traffic collisions, industrial accidents). The same therapists provided the 'active treatment' condition as the 'attention control' condition, so that therapists were not blinded to treatment, research question, therapy allegiance, etc., cf. Mitchell, et al. (1977), Luborsky et al. (1999). The researchers found that CBT was more effective than 'supportive counselling'. In terms of the research questions for this thesis it was not clear to what extent this was a 'fair test' of the clinical effectiveness of person-centred psychotherapy and the impact of the therapeutic relationship.

Bryant, Sackville, Dang, Moulds and Guthrie (1999) compared CBT (prolonged exposure therapy with or without an 'anxiety management' component) with 'supportive counselling' following motor vehicle accident or non-sexual assault. As with the previous Bryant study the same therapists provided each of the different conditions cf. Mitchell, et al. (1977), Luborsky et al. (1999). These researchers found both 'active treatment' conditions were superior to supportive counselling, although there were comparable reductions in intrusive and arousal symptoms of PTSD across all groups. There was no difference between the CBT conditions of prolonged exposure with or without anxiety management. Again it was not clear to what extent this was a 'fair test' of the clinical effectiveness of person-centred psychotherapy and the impact of the therapeutic relationship and both Bryant studies suggested the need for further research on anxiety outcomes from person-centred therapy and an assessment of the impact of the therapeutic relationship for anxiety outcomes.

Barrowclough, et al. (2001) compared two CBT therapists with one 'supportive counselling' therapist in a trial of anxiety for CBT with older adults. Whilst the CBT condition was found to be significantly more effective than supportive counselling, the authors cautioned: 'It is not possible to eliminate the possibility that therapy effects were attributable to the qualities of the individual therapists rather than to differences in the treatment per se... this potential confound indicates some caution in interpretation of the results' (p. 761). (cf. Huppert, et al., 2001). It was not clear to what extent this was a 'fair test' of the clinical effectiveness of person-centred psychotherapy and the impact of the therapeutic relationship and this suggested the need for further research on anxiety outcomes from person-centred therapy and an assessment of the impact of the therapeutic relationship for anxiety outcomes.

Westen and Morrison (2001) conducted 'A multidimensional meta-analysis of treatments for depression, panic and generalised anxiety disorder: An empirical examination of the status of empirically supported therapies' based on studies in the period 1990 to 1998 for manualised therapies for these diagnoses. Whilst finding evidence of significant short-term improvements as a consequence of therapy these researchers contextualised these 'impressive' short-term effects by stating that there were insufficient numbers of clients who improved and stayed improved at clinically meaningful follow-up intervals. In particular they questioned the relevance of such studies 'for clinicians who cannot pick and choose their patients the applicability of these findings to clinical practice is largely unknown' (p. 884). Making their point about the sacrifice of external validity in favour of internal validity these authors called for researchers to state: exclusion rates, percent improved, percent recovered, percent who remain improved or recovered at follow-up, data on completers and intent to treat samples. These findings appeared to pave the way for the co-authored piece by Drew Westen that was a larger and more detailed critique of 'The empirical status of empirically supported psychotherapies: Assumptions, findings and reporting in controlled clinical trials' by Westen, et al. (2004), see below. The guidance to researchers offered by Westen and Morrison and Westen et al. was heeded in the conduct and reporting of this research, to the extent that it was possible. However, with the available resources it was not possible to conduct a follow-up exercise to check if any gains were maintained. Although Westen and Morrison make a strong case for naturalistic research that is high in external validity and it is not necessarily impossible to seek to control internal validity in naturalistic research (Cook & Campbell, 1979), and see below.

As mentioned Elliott, et al. (2004) reviewed the evidence for experiential therapies and found for anxiety the mean pre-post effect size was 1.30 'large' and 'possibly efficacious' in terms of Chambless-Hollon criteria because the requirement for replication across independent research settings was not satisfied. Once researcher allegiance effects were controlled for there was no significant difference in outcomes for experiential therapies as compared with CBT.

These were the anxiety benchmark studies chosen to compare this study with on the basis that these included a selection of the evidence that NICE used to make judgements about what therapies the NHS should offer. In addition the Elliott et al. (2004) meta-analysis was included as a comparator study because this considered the available evidence for experiential therapies in the treatment of anxiety. Clearly NICE can only review available

research and whilst no evidence is not evidence of ineffectiveness, the relative lack of evidence that NICE appeared to be pointing to for person-centred psychotherapies reinforced the need for research of the kind in this thesis. Even if NICE would not accept a naturalistic study such as this one, a naturalistic study was considered a necessary step in making the case for investment in an RCT of the kind that NICE might consider.

Writing about the NICE (2004) report on anxiety and in particular 'why counselling doesn't get a look in' Susanna Lawrence (2005) sought to describe the process NICE had used and the rationale for recommending against 'counselling' in favour of CBT. The process and findings of the NICE review of anxiety have been reviewed elsewhere (Weston 2005).

3.1.3 Depression outcomes

In their review of the evidence for counselling in primary care Rowland et al. (2000) found four controlled trials of non-directive counselling which taken together indicated that patients receiving counselling showed a modest and significant improvement in symptoms compared with usual GP care. These authors also found that patient satisfaction with counselling was 'high' and counselled patients were more likely to be considered 'recovered' than usual general practitioner (GP) care ('tentative evidence' for this, p. 215). Acknowledging the research base was limited these authors looked forward to further research in the area, citing in particular the work then underway by Ward et al (2000). Ward et al in their randomised controlled trial of non-directive counselling, cognitive-behaviour therapy (CBT) and usual GP care found that both forms of psychotherapy were equivalent in outcome at four months and superior to usual GP care. However, at twelve months all three forms of treatment had equivalent outcomes.

In their research published in 2003 Watson, Gordon, Stermac, Kalogerakos and Steckley compared process-experiential therapy with CBT in the treatment of depression, finding that outcomes were generally equivalent between therapies, in terms of depression, self-esteem, general symptom distress and dysfunctional attitudes; with clients randomly assigned to the process-experiential group reporting a significantly greater decrease in interpersonal problems.

As mentioned Elliott, et al. (2004) reviewed the evidence for experiential therapies and found for ‘depression’ the mean pre-post effect size was 1.18 ‘large’ and the Chambless-Hollon criteria were fulfilled for a ‘specific and efficacious’ treatment.

Missirlian, Toukmanian, Warwar, and Greenberg (2005) studied emotional arousal, client perceptual processing and the working alliance in experiential therapy for depression. Their results suggested that experiential therapy was effective for depression and that emotional arousal was necessary for client processing and therapeutic improvement.

Dimidjian, et al. (2006) conducted a randomised controlled trial of behavioural activation, cognitive therapy and antidepressant medication with adults in the acute phase of major depression and found that behaviour therapy and medication outperformed cognitive therapy, suggesting it was the behavioural activation component of CBT that was the active ingredient.

3.1.4 Distress outcomes and related commentary on methodologies.

Stiles, Barkham, Twigg, Mellor-Clark, and Cooper (2006) used the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM, Barkham, et al., 2001) to evaluate the effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service (NHS) settings to further evaluate psychotherapy’s ‘equivalence paradox’ (Dodo bird verdict). This was naturalistic research and patients (1309) received therapy at one of 58 NHS primary or secondary care sites without random allocation to treatment. The average pre-post effect size was $ES(d) = 1.36$, a statistically significant change and there was no significant difference between the three therapy conditions. The authors concluded that these ‘results tended to support the Dodo verdict for these three treatment approaches as practised routinely across a range of NHS settings’ (p. 562). The authors described some limitations on their findings, including limited specification of treatments, non-random assignment of patients to treatment groups, absence of a control group, missing data and restriction to a single self report measure. In terms of investigator allegiance these authors considered this paper to be balanced by the orientations of the individual authors.

A subsequent and larger replication was reported in 2007 by Stiles, Barkham, Mellor-Clark, and Connell. This time 5613 patients were studied and the overall pre-post effect size was

ES(d) = 1.39, again a statistically significant change. Again there was no statistically significant difference between the three different schools and the authors reported similar limitations to their study.

Clark, Fairburn, and Wessely (2007) published a commentary on Stiles et al. 2007 paper in which they criticised the study on a number of bases. These authors criticised this type of naturalistic outcome research on the basis that a) there were lots of 'missing cases' that could have been 'treatment failures', b) these findings could have been restricted to the 'easiest' clients in the research, c) these findings could have been restricted to the 'best' outcomes, d) these findings could have been simply a consequence of 'regression to the mean', e) these findings could have been simply 'natural recovery' and f) these findings could have been 'attributable to concurrently administered medications'.

Stiles, Barkham, Mellor-Clark, and Connell (2008) wrote a letter to the editor by way of a 'rejoinder' to Clark, et al. Stiles et al. re-stated that they were not 'drawing conclusions that were not warranted (p. 1), simply that they were reporting their findings as they were e.g. this was not an assessment of all therapy in the NHS, it was only an analysis of those clients who had re-completed questionnaires. Stiles et al. addressed what they summarised as Clark et al. three lines of argument a) that the equivalence finding was robust in its context, b) possible confounds were addressed to the extent that was possible with the available data, and c) randomised controlled trials have virtues and so do naturalistic studies, particularly in terms of realism (external validity).

In terms of the implications for this piece of research arising from these Stiles et al. papers, together with the Clark et al. commentary, there are two main implications and these are in terms of the sizes of outcomes and research methodologies. Firstly, this piece of research used the two Stiles et al. research studies as benchmarks with which to compare the outcomes from this research. Secondly, Clark et al. made some important points about possible threats to the internal validity of naturalistic research that would need to be considered as possible alternative hypotheses (Cambell & Russo, 1999) to the central hypothesis; that it was person-centred therapy that was responsible for any observed improvements in client-wellbeing.

Having reviewed the benchmark studies the rest of this section considers some methodological points related to the Stiles et al. and Clark et al. dialogue that arose in the context of reviewing outcomes for distress.

In contrast to NICE/APA reliance upon RCTs Westen et al. (2004) showed overly-focusing upon internal validity was at the expense of external validity; a very highly controlled RCT could have little relevance to the outside world. Victora, Habicht, and Bryce (2004) have also made the case for ‘moving beyond randomised trials’ in ‘evidence-based public health’.

Schwartz, Trask, Shanmugham, and Oswald Townsend (2004) examined the issues related to ‘conducting psychological research in medical settings: Challenges, limitations and recommendations for effectiveness research’. These authors provided what they termed ‘an alternative efficacy-effectiveness view’ (p. 501) in which they schematised the idea that high internal validity was directly related to low external validity and the opposite; low internal validity directly related to high external validity. These authors made the case for naturalistic effectiveness research and provided a process that they referred to as the ‘clinical research continuum’ (p. 502) wherein the act of providing healthcare may enable the identification of a problem, a potential solution could be piloted and if promising tested in an RCT, before being tested back ‘in the real world’ and this making a direct contribution to clinical practice (and the cycle begins again). In contrast the NICE/APA view that RCT was the ‘gold standard’ for research some authors described putative research situations where RCTs were not the method of choice, potentially problematic and in some cases simply unable to provide the required answers.

Iacoviello et al. (2007) identified a potential threat to validity of randomised controlled trials when they found that treatment preferences affected the therapeutic alliance. Iacoviello et al. randomly allocated patients with major depressive disorder to therapy, medication or pill placebo and assessed therapeutic alliance before treatment and subsequently finding ‘among patients initially preferring psychotherapy, those receiving psychotherapy experienced increases in their alliance over time, whereas those receiving active medication or pill placebo experienced decreases’ (p. 194). Patients who preferred medication had no differences in alliance development whatever treatment group they were allocated to. These authors cautioned: ‘because alliance is a robust predictor of outcome, treatment preferences may need to be considered in randomised controlled trial settings’ (p. 194) and recommended that naturalistic studies were warranted to augment those from RCTs.

Accepting that the naturalistic approach proposed in this research had a scientific basis, nevertheless this research would need to, in the words of Clark et al. (2007) carefully examine alternative explanations: a) whether there were lots of ‘missing cases’ that could be ‘treatment failures’, b) whether these findings could be restricted to the ‘easiest’ clients in the

research, c) whether these findings could be restricted to the 'best' outcomes, d) whether these findings could be simply a consequence of 'regression to the mean', e) whether these findings could be simply 'natural recovery' and f) whether these findings could have been 'attributable to concurrently administered medications'. Clearly this was a long list of alternative hypotheses to consider in relation to the proposed research. Whilst naturalistic research might appear 'easy' to conduct, compared with seeking to control for all plausible variables in an RCT, making valid scientific inferences from naturalistic research could be 'hard' in terms of the alternative hypotheses to consider. On the other hand, the low level of internal validity, if appropriately addressed, is compensated for by the high level of external validity; the research findings would be those of 'real world' therapy.

3.1.5 The therapy workforce in the UK and investment in therapy research

In the UK the largest organisation representing counsellors/psychotherapists is the British Association for Counselling and Psychotherapy (BACP, 2009). As at August 2007 there were 28,012 individual members (A. Couchman, personal communication, 3rd January 2008). Members indicated which theoretical models they were trained in and a member may have indicated they were trained in more than one model. As at 3rd January 2008 the total number of theoretical models that members had indicated they were trained in was 47,971; suggesting on average members were trained in around 1.7 models (A. Couchman, personal communication, 3rd January 2008). BACP members trained in CBT was 8.4%, one of the smaller groupings, the largest was person-centred, 30.4%, integrative 23.5%, psychodynamic 16.0%, humanistic 15.8%, gestalt 0.8% and 'other' 5.1%. There were a large number (perhaps 26,000 individual therapists) of non-CBT trained members of BACP. Combining this view of part of the 'therapist workforce' with the data on differential research funding for the different schools of therapy (below) suggested there may be a mismatch; do all non-CBT trained therapists need to be re-trained in CBT?

Given the research evidence in favour of the importance of the therapeutic relationship/alliance to psychotherapy outcome and the evidence in support of effective person-centred therapy outcomes, although not accepted by NICE, the present author wondered about the impact of Rogers' apparent discouragement of quantitative research (of the kind that NICE and the APA seek) and whether there was any difference in the amount of research funding the different therapeutic approaches received. Research on NHS funding

resulted in two presentations to international peer-reviewed conferences (Weston, 2008c and Weston & Weston, 2008). The author contacted the 71 research establishments that accounted for 99.5% of NHS funding of mental health research 2006-7 and asked (Freedom of Information Requests) each establishment to provide a list of how much money had been spent on researching the three main schools of therapy (cognitive/behavioural - CBT, psychodynamic - PDT, humanistic/person-centred - PCT) plus any 'other' psychotherapy research over the preceding ten years. Fifty four organisations responded in full (76.1% of 71), two gave part responses (three years and five years of data, 2.8% of 71), eight provided estimates or confirmed the author's assumptions (11.3% of 71, cumulatively 90.1%), six refused to give any data (8.5% of 71), and one did not reply (1.4% of 71). On this basis in the ten years to 2006-7, £18.5m was spent on researching psychological therapies by the NHS; 46.0% on CBT, 28.0% on PDT, 3.7% on PCT and 22.4% on 'other', including brief therapies, couples, groups, etc. This research was hampered by receiving only a part response from the South London and Maudsley NHS Trust, which was by far the biggest spender of NHS research money on mental health, accounting for 48.1% of the total in 2006-7; it should also be noted that this Trust received a comparable amount of money from non-NHS sources e.g. donations from private companies, organisations, overseas, etc. Combining this data with an inspection of the research reports published by the 71 organisations the author estimated that over the preceding seven years to 2006-7, £295.2m had been spent by NHS organisations researching psychological therapies; 96.6% on CBT, 1.7% on PDT, 0.2% on PCT and 1.4% on 'other'. Whilst there were difficulties with the methodology (e.g. missing data, estimations etc.) and the fact that there are other sources of research funding (e.g. Medical Research Council, Wellcome Trust, etc.) this was suggestive of differential research funding levels for the different schools of therapy. Irrespective of the precise numbers it would seem that comparatively little money has been spent on researching PCT, certainly compared with CBT, by the NHS.

3.1.6 Outcomes management.

During the early part of the twenty-first century there has been interest in using the instruments of research to 'manage outcomes'; using research questionnaires to check on client progress through therapy. This has also been referred to as 'outcomes assurance' or 'patient-focused research' and is arguably similar to the research approach that Rogers encouraged in 'towards a more human science of the person' (Rogers, 1985). Tracking client

response to treatment and changing therapy, if required, and or making progress through therapy part of the therapeutic dialogue has been shown to improve outcomes.

In 2010 Shimokawa, Lambert, and Smart published a meta-analytic and mega-analytic review of previous work that supported the approach to ‘enhancing treatment outcome. This had begun in 2001 when Lambert, Hansen, and Finch wrote about ‘Patient-focused research: Using patient data to enhance treatment effects’. These authors found that therapists who received feedback on patient progress, or lack of it, had better outcomes. The feedback helped identify ‘difficult cases’ and directed efforts towards patients who might otherwise have deteriorated. Patient progress was assessed by repeated (weekly) completion of outcome questionnaires. These authors found that ‘tracking patients on a weekly basis presented no more risks than did typical pre-test and post-test designs’ (p. 169).

The 2010 work built on six previous studies and assessed the potential value of three types of patient feedback interventions used in the quality assurance system: 1) giving progress feedback to therapists, 2) giving progress feedback to both therapists and clients, and 3) providing clinical support tools (CSTs) to help identify the causes of deterioration (e.g. therapeutic alliance, patient motivation or patient social support) and providing suggestions for resolving identified problems. All three items were effective in enhancing treatment outcome and items one and three in the preceding list were effective in preventing treatment failure. The authors conclude that ‘the accumulating evidence is substantial in favour of the routine use of progress feedback and clinical problem-solving tools’ (p. 309).

The fact outcomes were improved meant this system was making a contribution to therapy outcome and to that extent would appear likely ‘causative’ of at least some part of measured treatment outcome. Lambert and Shimokawa (2010) subsequently estimated the effect size for ‘collecting client feedback’ at $r = .23$ to $.33$ (see below). In the journal article, Shimokawa, et al. (2010) questioned the possible mechanism(s) of change and wondered if what they were observing was a function of the dose-response effect (Barkham, et al., 2006, Minami, et al., 2007) in that at least some of the effect appeared to come from keeping some clients in therapy longer since on average clients stayed in therapy for 0.9 sessions longer. This effect could be described in a multiplicity of ways e.g. a therapist effect from seeking to understand the client’s experience of therapy, an empathic effect from seeking to respond better to the client (itself a therapist effect), a client effect, perhaps from re-motivating the client, a ‘technique’ of therapy, or even a placebo effect (Kirsch, 2002, Kaptchuk, et al.,

2008). However, this approach was described, what was clear was that collecting client feedback and responding to this made a contribution to therapy outcome and it was thus likely one of many small-medium effects that made a contribution to an overall effect. In terms of the research questions for this thesis it seemed plausible that 'collecting client feedback', to the extent this was done in person-centred therapy could make a contribution to outcome and, depending upon the approach one took to defining factors, this could be described as an impact of the therapeutic relationship.

One issue that arises from this approach to 'outcomes management' is what to do if a client has clinical scores for both depression and anxiety, to avoid completing several questionnaires per session. With the rationale that anxiety treatments work better for those clients who are not depressed the answer appears to be to monitor depression outcomes first, ideally until the depression has become non-clinical, and then to monitor the anxiety; so 'treat' (and measure) depression before anxiety, see Newman, Stiles, Janeck and Woody (2006). Shimokawa et al., (2010) used the Outcome Questionnaire 45 (OQ-45), a broad measure of client distress, so this issue appeared not to arise for them. However, logically if a quality assurance system were expanded into other areas of therapeutic work it seemed plausible that a number of outcome questionnaires could be 'in progress' for one client seeking to work upon a number of therapeutic issues, e.g. depression, anxiety, post-traumatic stress disorder, etc. This idea of multiple outcomes required from the therapeutic work appeared to be an area for further work.

3.1.7 Concluding comments on outcome research.

This section has provided an overview of the outcomes literature and in general terms the argument appears to be between those who consider the main schools of therapy to have approximately equivalent outcomes and those who consider CBT to be superior. The former argue that once researcher allegiance is taken into account the latter's argument falls down. One of the aims of this research was to benchmark outcomes with a selection of research findings and these benchmark studies were briefly reviewed in separate sections, along with some methodological points along the way. The case for this research was also made in the context of the size of the 'person-centred' or 'non-CBT' therapy workforce in the UK and a comparable underinvestment in research in person-centred or non-CBT research in the UK. A description of outcomes management was provided as the present author practices in a way

related to this and this has later relevance. Arguably, outcomes management, client-centred research, and tailoring the therapy to the needs of the individual client, is as ‘person-centred’ as anything that Rogers did and maybe this approach will find some support in the person-centred community. The next section reviews the process-outcomes literature as it related to this thesis.

3.2 Process-outcome research literature

This section reviews the process-outcome literature as it related to the second of the main research questions; what is the impact of the therapeutic relationship. In addition to finding out what outcomes occurred with person-centred therapy it was important to consider the impact of the therapeutic relationship as this is considered foundational to the approach.

Reading Rogers’ theory it was unclear to the present author how a person could test Rogers’ theory, that the therapeutic relationship – congruent empathy and unconditional positive regard – caused outcomes. The literature review was needed to see how one might test Rogers’ theory and to see what researchers had found.

This section starts off with ‘the case against Rogers’ theory’ in the sense that in this period 1970s – mid 1980s reviewers criticised the early attempts at validating Rogers’ theory. Reviewers found methodological problems with the early work of Rogers and his colleagues. Perhaps it was this ‘case against Rogers’ that arose from the reviews described below that contributed to a decline in person-centred research, particularly quantitative research, in addition to the putative cause of this elaborated above re Rogers (1985).

In fact ‘the case against Rogers’ was simply that there were shortcomings in the research, not that the theory was wrong, perhaps the theory was ahead of its time in the sense that research methodologies had not evolved such that the theory could be adequately tested. Helpfully the reviews that made this ‘case against’ the early research pointed the way in how the research could be conducted. This research was needed to conduct the test implied by these reviews.

Further methodological inspiration was provided by a series of papers that involved Stiles and colleagues from 1986 – 1998 in a section entitled ‘abuse of the drug metaphor’ after one of the key papers in this series. During this time period Horvath and Greenberg (1989)

published the working alliance inventory (WAI) which led to a large number of studies on the impact of the working alliance. This is reviewed within the section ‘abuse of the drug metaphor’ in temporal order because methodological points were also made before and after this important publication.

These two sections (‘the case against Rogers’ and ‘abuse of the drug metaphor’) effectively provide the case for the process-outcome correlation research in the experimental part of this thesis. However, methodological development continues apace. The next section describes causation approaches as these make the case for an improved method of assessing the impact of the therapeutic relationship, beyond the scope of this research. The studies reviewed in this section go some way to support Rogers’ case for the therapeutic relationship as causative and demonstrate how such research could be conducted, together with some cautionary notes for both this process-outcome research and potential future research on the therapeutic relationship, as described by Rogers, as causative.

The following two sections further support Rogers’ case for the therapeutic relationship as causative and illustrate two further methodological approaches which have become important in psychotherapy causation research. The first of these is the ‘common factors’ approach developed by several authors and made ‘mathematical’ by Lambert and colleagues. The approach here was to seek to put percentages on the contributions different factors made to therapy outcomes. Whilst Rogers may have some claim to being correct about the causative nature of the therapeutic relationship he was wildly wrong that this was ‘necessary and sufficient’. There are many factors that go together for therapy success, or otherwise, and these are described in this section. Lambert’s approach was one of ‘cutting up the pie’ (Cooper, 2008) to describe the factors that contributed to outcome and whilst this was based on many research studies it was not a meta-analysis.

The next section is therefore about meta-analyses and the contribution these have made to understanding what contributes to therapy outcomes. This starts off with Shadish and Sweeney (1991) who made an important contribution to the idea of using meta-analysis to answer difficult questions about moderators and mediators of outcome, although unfortunately whilst they may have ‘pointed the way’, the research available to them did not result in very interesting findings. However, this ‘way pointing’, ultimately led to very interesting findings from the perspective of someone interested in the impact of the therapeutic relationship on outcomes and a preview of a soon to be published book is

described that goes some way to making the case for elements of the therapeutic relationship as predictors of therapy outcomes.

Recent studies have made some important findings with regard to the tenets of Rogers' theory, although mainly not by authors from a person-centred approach. These are reviewed to the extent that they point to some of the validity of Rogers' theory and are inspirational about what is possible for future research on the impact of the therapeutic relationship and to the extent that these more recent studies provide an interpretative context for these research findings.

The naturalistic approach to researching outcomes in the experimental part of this thesis needed to consider alternative causes mediators/moderators and two of these were alluded to above; antidepressant medications and dose-effect relationships. Since these are process variables these are reviewed in this section to the extent that they impact on the research questions – the clinical effectiveness of the person-centred approach and the impact of the therapeutic relationship. These are important variables that play a part in the interpretive context for this study, both in terms of the outcomes from person-centred therapy and the impact of dose-effect relationships on the process-outcome findings.

Following the review of process-outcome literature, to the extent it impacts on this thesis an introduction to the experimental research of this thesis is given, together with a brief description of some of the relevant methodological issues.

3.2.1 The case against Rogers' theory

Barrett-Lennard (1962) developed the Barrett-Lennard Relationship Inventory (BLRI) with which to test the therapeutic conditions as described by Rogers. He tested Rogers' theory with 42 college counselling centre clients seen by 21 therapists. Clients completed the BLRI at session 5 and post-therapy, average session length was 33 sessions. Although this research found a zero order correlation between therapeutic conditions and outcome this approach was criticised on methodological grounds. Gurman (1977, p.523) pointed out that pre- to post-change scores should be corrected for the initial level of the pre-score before testing for the perceived conditions-outcome correlation. Amongst other studies, Gurman also made the same criticism of the Wisconsin research published by Kiesler, Klein, Mathieu, and Schoeninger (1967) upon which Rogers (1967) had made his claim that 'the greater the

degree to which the conditions of therapy existed in the relationship, the greater... the evidence of constructive outcome' (p. 91).

In contrast Truax and Mitchell (1971) summarised research on empathy, warmth and genuineness to 1970 and concluded that:

...therapists or counsellors who are accurately empathic, non-possessively warm in attitude and genuine are indeed effective. Also, these findings seem to hold with a wide variety of therapists and counsellors, regardless of their training or therapeutic orientation, and with a wide variety of clients or patients, including college underachievers, juvenile delinquents, hospitalised schizophrenics, college counselees, mild to severe outpatient neurotics, and a mixed variety of hospitalised patients. Further, the evidence suggests that these findings hold in a variety of therapeutic contexts and in both individual and group psychotherapy or counselling (1971, p. 310).

This juxtaposition of this 'review of the evidence' with the already mentioned Gurman (1977) review and the radically different conclusions reached, points to the changes in methodologies and accepted practice over the period of this review. What may have been accepted practice in 1967 was not in 1977. This continual evolution of methodologies and accepted research practice inevitably continues, reinforcing the need to continue research effort and embrace new techniques cf. the apparent abandonment of quantitative research by some in the person-centred field as an interpretation of Rogers' 1985 paper.

Introducing 'effective psychotherapy: a handbook of research' Allen Bergin (1977) described the historical process in therapy research from emphasising 'technique' to 'relationship' and back again:

As research evidence and clinical experience have accumulated, the modal opinions of leaders in the field have shifted back and forth on these issues... Comparative studies continue to show little difference in the outcomes of diverse approaches, even though therapy, by itself, can be shown to have significant effects when compared to no-treatment. If this assessment is correct, then the viewpoint of Jerome Frank is most strongly supported by the evidence – namely that the effective factors are the same for all therapies and that these are the same as the common ingredients in all types of healing and influence processes that occur in all cultures (p. xv)

Bergin's reference to Jerome Frank is to Frank's 1973 work which elaborated upon the 'common factors' in therapy, see also a similar and later reference to Frank in research by Gomes-Schwartz (1978).

Mitchell, Bozarth, and Krauft (1977) reviewed the evidence for empathy, warmth and genuineness and pointed to some criticisms of the early research leading to Rogers' theories:

There were, however, early reservations about this position (Matarazzo, 1971, Meltzoff and Kornreich, 1970). They questioned the validity of many of the studies. They suggested that, in many cases the number of therapists was small and, in an unknown number of instances the therapists may have been aware of the particular hypotheses and even associated with the research effort. An additional problem of the utmost importance is the degree to which the therapists in these studies valued the interpersonal skills in question. No attempt has been made to determine this in the earlier outcome studies or in any of those cited in this chapter (p. 482).

Mitchell et al. go on to describe some of the methodological problems in researching these highly complex questions and concluded:

...the mass of data neither supports nor rejects the over riding influence of such variables as empathy, warmth and genuineness in all cases. The recent evidence, although equivocal, does seem to suggest that empathy, warmth and genuineness are related in some way to client change but that their potency and generalisability are not as great as once thought (p. 483).

Gurman (1977) reviewed 26 studies of perceived therapeutic conditions and outcome in individual therapy, 20 studies reported findings in favour of the conditions-outcome hypothesis, 3 reported mixed but supportive findings and 3 reported results that failed to demonstrate a relationship between outcome and conditions. Gurman concluded from these 26 studies that:

...there exists substantial, if not overwhelming, evidence in support of the hypothesised relationship between patient-perceived therapeutic conditions and outcome in individual psychotherapy and counselling (p. 523).

However, he also pointed out that a number of studies had methodological flaws, including the studies of Barrett-Lennard (1962) and Kiesler, et al. (1967) described above. He

suggested that the best test of the conditions-outcome hypothesis was to correlate outcome with conditions, controlling for initial severity. Of the 26 studies, Gurman described Lesser's (1961) paper as the only one to have used this method and this failed to find any significant relationship between conditions and outcome.

Lesser was completing a PhD at Michigan State University and required data from clients 'during the second week of June, 1958, regardless of whether or not the client had terminated counselling' (p. 332). Lesser used 'change in self-perceptions and in ideal-self perceptions' (p. 331) as his outcome measure, drawing on the work of Rogers and Dymond (1954). He found the 22 clients counselled by 11 client-centred therapists improved on the outcome measure such that 'the resulting t was 3.39 which is statistically significant beyond the .01 level of confidence' (p. 332). On this basis the therapy was effective in the terms defined, although whether or not there was an impact of the therapeutic relationship was much less clear; because 50% of the clients were still in progress at the time of outcome measurement and 'the client and counsellor ratings of counsellor empathic understanding were significantly higher for the continuing than for the terminated clients (.03 and .04 levels of confidence, using White's T)' (p. 333). Whilst it was Gurman's (1977) view that Lesser used the correct methodology, controlling for initial severity, it was not clear that this was a fair test of the impact of the therapeutic relationship because 50% of the 'end' scores were not in fact 'end' scores.

Lambert et al (1978) reviewed 'therapist interpersonal skills: process, outcome, methodological considerations and recommendations for future research' and concluded that:

...Rogerian hypotheses have been only modestly supported, that the lack of support is due both to the difficulties encountered in sampling and rating therapy sessions and to the failure of client-centred therapy to specify more precisely the times when specific conditions (such as empathy) might be most facilitative (p. 467).

These reviewers highlighted a number of methodological issues for research of this kind:

1. From whose point of view shall the process of therapy be assessed?
2. Is the medium of audiotape a representative and suitable one for judging facilitative conditions?
3. Should raters be experienced therapists or naïve persons trained specifically for the purpose?

4. Is it even necessary to train judges?
5. Does the sex of raters confound the rating process?
6. How should samples be collected: what are the effects of segment location?
7. Do the process scales measure separate uni-dimensional traits or a single dimensions?

Rogers' theory (1957, 1959) was that it was the client's perspective that was most important for perceiving the therapeutic relationship. Although the client's perspective was questioned in Rogers' theory because clients were, by definition according to his theory, 'incongruent' and subject to 'a discrepancy between the actual experience of the organism and the self picture of the individual, insofar as it represents that experience' (Rogers, 1957, p. 96).

For her doctoral dissertation Beverley Gomes-Schwartz (1978) studied effective ingredients in psychotherapy and prediction of outcome from process variables. Thirty five male college student clients were allocated to one of either four male psychiatrists (analytic therapists), four male psychologists (client-centred therapists, referred to as 'experiential therapists') or seven male college professors (alternative therapists). Using the Vanderbilt Psychotherapy Process Scale two raters blinded to treatment condition were able to identify from audiotapes significant differences in the the types of process underway in each therapy condition.

In Gomes-Schwartz (1978) research, sessions with analysts were significantly higher in both therapist and patient exploration than the other conditions, sessions with experiential therapists were significantly higher in both therapist and patient exploration than the alternative therapists; and both alternative and experiential therapists were found to exhibit greater therapist warmth and friendliness than the analysts. Yet there was no significant difference in overall outcomes nor any differences on any of the six outcome criteria. This finding appeared to support the Dodo verdict.

In terms of prediction of outcome from process variables, Gomes-Schwartz analysis met Gurman's (1977) idea that partial correlations were the appropriate measure, and even though there were significant differences in what the therapist was doing in sessions for the three therapy conditions, none of the process variables significantly predicted outcome. Instead: 'therapy outcome was most consistently predicted by the patient's willingness and ability to become actively involved in the therapy interaction – a dimension of therapy process that did *not* distinguish among the three treatment groups' (p. 1031, emphasis original).

Gomes-Schwartz interpreted these findings as supporting ‘Frank’s (1973) theory of “non-specific factors” as determinants of outcome’ (p. 1032). She wondered about ‘one option for maximising the effectiveness of psychotherapy would be to select only those patients who evidence a capacity to actively participate in a therapeutic interaction’ (p. 1032). This appeared to have some similarity with Rogers and Dymond’s (1954) findings about ‘client characteristics’ that were ignored in Rogers’ theory, Gomes-Schwartz found that ‘patients who were not hostile or mistrustful and who actively contributed to the therapy interaction achieved greater changes than those who were withdrawn, defensive, or otherwise unwilling to engage in the therapy process’ (p. 1032).

Patterson (1984) conducted a non-numerical ‘review of reviews’ of the evidence to support Rogers’ theory. In contrast to Lambert et al. (1978) review, Patterson concluded that:

Considering the obstacles to research on the relationship between therapist variables and therapy outcomes the magnitude of the evidence is nothing short of amazing. There are few things in the field of psychology for which the evidence is so strong. The evidence for the necessity, if not the sufficiency, of the therapist conditions of accurate empathy, respect or warmth, and therapeutic genuiness is incontrovertible (p. 437).

Patterson’s main point seemed to be that ‘all reviewers are biased’ and made four points about the sources of this bias:

1. Reviewers are biased in the selection of the studies they review.
2. Reviewers apply critical standards to research they are biased against and are more lenient with findings they like.
3. Reviewers emphasise studies they favour and de-emphasise studies they dislike.
4. Reviewers are selective in what they report about outcome measures, mentioning results they like and not those they don’t.

Watson (1984) reviewed the empirical status of Rogers’ hypothesis of the necessary and sufficient conditions for effective psychotherapy and concluded that:

Though there is a substantial amount of research on Rogers’ hypotheses of the necessary and sufficient conditions for effective therapy, none of the studies meet all of the conceptual and methodological criteria for rigorous research on this topic. Researchers have not carefully followed the logic of the hypotheses in designing studies and interpreting the results. A central shortcoming is the inattention to major conceptual criteria: employinig

client ratings of the therapist provided conditions, including all the hypothesised conditions and addressing the issue of causality. A large number of studies have used judge ratings of the therapist provided conditions which are irrelevant to the hypotheses as Rogers stated them, and neglected client perceptions of the relationship, which are essential to a test of the hypotheses. The studies that have focused on client perceptions of the relationship typically have not included all of the hypothesised conditions, thereby not testing the hypotheses as propositions of a set of necessary and sufficient conditions. Moreover, studies of client perceptions have not addressed the issues of the hypothesised conditions as causes of outcome. After 25 years of research on Rogers' hypotheses, there is not yet research of the rigour required for drawing conclusions about the validity of this important theory (p. 40).

3.2.2 Abuse of the drug metaphor

Stiles, Shapiro and Elliott (1986) sought to examine an apparent contradiction between apparently equivalent outcomes between therapies with demonstrably different techniques, hypothesising that this could be a consequence of: a) differential outcomes awaiting better methods and measures to be revealed, or; b) different techniques share a common core of processes, or; c) gross simplifications of therapy outcome and process studies failing to observe differential effectiveness at the micro-process level. These authors concluded that the case that 'psychotherapy works' had been established (p. 175) and called for 'more fine-grained thinking' (p. 176) in both outcome and process studies in order to tease out *how* therapy works and any differences between schools at the micro-process level.

In 1988 Stiles wrote an article 'Psychotherapy process-outcome correlations may be misleading' the conclusion of which was a 'discrediting' of the correlational approach to demonstrating causal links between psychotherapy process and outcome (p. 33). Reviewing process-outcome findings to date, in particular Orlinsky and Howard's (1986) recent meta-analysis of process-outcome findings, Stiles concluded research linking process to outcome had been 'disappointing' with a 'meagre yield'. Stiles noted that researchers, including himself, had 'interpreted the absence of significant process-outcome correlations as indicating that a process component is therapeutically inert' (p. 28). However, upon reflection, Stiles recognised that if therapy was perfectly responsive to client needs '...then all clients would have the same outcome. With no outcome variation, the correlation would be zero, despite variation in process.' (p.30). Furthermore Stiles pointed to one of the

fundamental features of psychotherapy training and supervision, adapting therapy to meet the needs of individual clients. Pointing out that standardising delivery of therapy so that all clients were treated the same would be ‘absurd’, ‘psychotic’ or ‘incomprehensible’ (p. 29). Stiles’ argument was that therapy is responsive to client needs such that clients receive more or less of a component dependent upon their needs and as such ‘to the degree that therapists are appropriately responsive to client requirements, process-outcome correlations underestimate the process-outcome relationship’ (p.30).

Given this conclusion, that correlating process components with outcomes was thrown into question, Stiles pointed out that there were alternatives to process-outcome research and these included incorporating responsiveness in process measures, case studies, microanalysis of microprocesses and the analysis of similar ‘events’ in therapy (e.g. Gestalt two-chair procedure) and closely examining subsequent change as a consequence of these similar events.

Stiles subsequently published a joint paper with Shapiro (1994) that would further elaborate these points and stir up some controversy and debate.

In 1989 Horvath and Greenberg published an article about the development and validation of the working alliance inventory’ (WAI) a pan-theoretical three part (bonds, goals and tasks) conceptualisation of the therapeutic alliance. As part of the development and testing of the WAI these authors had determined that the WAI strongly correlated with the empathy scale of the BLRI, sharing some 48-52% of variance, although these authors’ initial tests suggested the WAI had better outcome predictive validity than the empathy scale. The development of the WAI was seemingly quite a turning point in psychotherapy research as it appeared to ‘give back’ the notion of the therapeutic relationship/alliance as a legitimate area for research across therapy schools without being necessarily associated with one particular school. It may also have been a better conceptualisation of the relationship ‘common factors’ than that achieved by Rogers and collaborators. A number of research studies subsequently used the working alliance as a process-outcome correlate and some of these are described below, in general terms the results have been to support the correlation of working alliance with outcome.

Earlier, Stiles et al. (1986) had pre-empted the publication of the work on the therapeutic alliance and had described ‘interest in the therapeutic alliance emerged out of a growing dissatisfaction during the 1970s with the “therapeutic conditions” concept’ (p. 173). They

described how in factor analytic studies measures of participation from both client and therapist loaded onto the same factor, citing the work of Gomes-Schwartz (1978) amongst others in supporting this assertion, and this suggested a 'pattern of mutual facilitation' (p. 173). Stiles et al. expected that the work on the therapeutic alliance could 'subsume' earlier research on the Rogerian 'conditions' and outcome because 'client reports of therapist, warmth, empathy and genuineness can be construed as measuring therapeutic alliance' (p. 174). The extent to which therapeutic relationship (Rogers, 1957) and therapeutic alliance (Horvath and Greenberg, 1989) are the same and or different was unclear, as too the extent to which the BLRI and WAI overlap or map separate domains. It seemed plausible that at least two positions were theoretically possible and not necessarily incompatible: the therapeutic relationship was more than just the alliance, and the alliance was more than just the relationship. It was unclear to what extent the respective psychological measures were 'fine-grained' enough to discern differences and similarities, to use the phrase employed by Stiles et al.

Stiles et al. (1986) pointed out at least two drawbacks of the approach based on the therapeutic alliance: Firstly, correlations with outcome might be a confound of early outcome, i.e. because a client feels improved s/he rated the alliance more favourably. It was subsequently shown that this potential criticism was unfounded and the alliance was shown to contribute directly to outcome independent of any early treatment effect, e.g. Klein, et al. (2003). Furthermore it was also subsequently shown that the therapeutic relationship (BLRI) was also predictive of outcome independent of any early treatment effect, Zuroff and Blatt (2006). The second drawback anticipated by Stiles et al. was similar to that described above, that 'the alliance construct is really only a conceptual umbrella for uniting a number of client and therapist contributions; the exact operation of these constituent factors remains to be clarified' (p. 174). Their concern here was that both the 'therapeutic alliance' and the 'therapeutic relationship' it was perhaps seeking to replace were described at too high a level of abstraction. The difficulty here was, what was an appropriate level of abstraction to define 'behaviour' or 'process', and whether or not measures of behaviour, process or outcome were sufficiently well defined and embraced within an appropriate statistical and methodological context for meaning-making to take place.

In 1994 a special feature of the Journal of Consulting and Clinical Psychology was published to 'contain' Stiles and Shapiro's (1994) paper 'Disabuse of the drug metaphor: Psychotherapy process-outcome correlations'. An introduction was provided in which Frederick Newman

(1994) described the Stiles and Shapiro article as ‘either an editor’s dream or an editor’s nightmare’ (p. 941) because the arguments they presented were both ‘attractive’ and causing some ‘provocation’ yet could not be accepted in ‘normal science’ because their argument required acceptance of a null hypothesis, contrary to convention, see for example Field (2005, pp, 22-6)

Stiles and Shapiro built upon Stiles’ 1988 paper about the potentially misleading nature of psychotherapy process-outcome correlations. They described an experiment within which ‘five theoretically relevant, reliably measured verbal process components were compared with the rate of change in three standard symptom intensity measures across the brief treatment of 39 (mainly depressed) psychotherapy clients. The expected significant process-outcome correlations were not found’ (p. 942). Stiles and Shapiro argued that the drug metaphor, more is better, was misleading because psychotherapists (and clients) were responsive to client requirements for process components. The responsiveness model suggested that process components could not be related to outcome because the process component would be varied to respond to client needs and was therefore moderated in a feedback loop to match client needs. Therefore, these authors argued, what would be expected would be an absence of correlation with outcome, if a process component was meeting client need in a responsive way, and this required acceptance of the null hypothesis. Unfortunately the scientific paradigm of hypothesis testing meant the null hypothesis could never be accepted, it was only the experimental hypothesis that could be accepted, at a defined level of probability.

In addition to the ‘containment’ of Stiles and Shapiro’s paper from a preceded introduction by Newman there were two following articles putting the case against Stiles and Shapiro. Firstly Silberschatz (1994) argued that process-outcome correlations were perfectly appropriate scientific tests provided that the process measures were ‘adequately conceptualised’ and gave an example of such. Secondly, Sechrest (1994), argued that, amongst other things, Stiles and Shapiro had inadequate statistical power and could have used better statistical analyses. One of the more interesting arguments that Sechrest deployed was that, given the responsiveness paradigm, ‘one would have to validate the efficacy of each psychotherapist seperately’ (p. 952) and this does become a later theme in therapy research e.g. Lambert, et al., 2002 and Wampold and Brown, 2005.

Stiles (1994) was given an opportunity on behalf of himself and Shapiro to respond to Silberschatz and Sechrest in the 'special edition' and Stiles summarised the contra-arguments as 'more complex measures should solve the problem' and 'more complex analyses should solve the problem', respectively. Stiles accepted that in correctly specified research 'correlational designs are entirely suitable for many questions in psychotherapy research' (p. 955). To some extent the argument put forward by Stiles and Shapiro could be termed a 'wake up call' to researchers to think carefully about the hypothesis they were testing, the methods and analyses employed and to incorporate the phenomenon of responsiveness into an understanding of process-outcome correlations.

In 1996 Hayes, Castonguay, and Goldfried published a further article commenting upon Stiles and Shapiro's 1994 paper and their assertion that the yield of process-outcome correlation research had been meagre because of incorrectly specified testing. Hayes et al. presented findings from research into anxiety and depression which they claimed were good examples of process-outcome correlation research, unlike Stiles and Shapiro's 'unfair' test. These authors accepted Stiles and Shapiro's points about overreliance on the drug metaphor and the importance of the responsiveness phenomenon. They moved the thinking on by pointing out that it was likely that the responsiveness phenomenon was more applicable for moment-to-moment process measures than for 'when the process variable is a potent intervention, change process or common factor of therapy that is based on a solid theory of change' (p. 913). These authors concluded that 'the study of change in psychotherapy is complex and requires multiple measures of process and outcome, and multiple methods of inquiry. The process-outcome correlation paradigm is only one of these methods and one that has contributed significantly to the advancement of the field' (p. 913).

In response Stiles (1996) accepted some of the points that Hayes et al. had made and moved the thinking on further by considering four reasons why process components might be in short supply in a therapy/research setting, i.e. lack of resources to provide the component, ignorance that the component was important to provide, failure to adequately evaluate the process component and failure to recognise that a process component may really be a 'subgoal' of outcome e.g. an adequate therapeutic alliance could be construed as an early index of outcome. Right back to Rogers and Dymond (1954) it had been apparent that 'process' and 'outcome' were not so easily distinguishable as they might first appear, e.g. for some clients the ability to form an alliance could be an important outcome along the way to achieving what might be considered more traditional 'outcomes'. Stiles (1996) argued that

‘more is better only when clients aren’t already getting enough, and this is particularly unlikely to be the case for robust therapy interventions that are embedded in clinical theory and practice’ (p. 918).

Subsequently, Stiles, Agnew-Davies, Hardy, Barkham, and Shapiro (1998) presented a complex and sophisticated research of relations of the therapeutic alliance with psychotherapy outcome. Five factor based aspects of the alliance (bond, partnership, confidence in the work, openness and client initiative in therapy) were correlated with residual gains on six outcome measures at three post-therapy measurement points (end of treatment, 3 month follow-up and 12 month follow-up) for 79 clients of five therapists who each provided either eight or sixteen sessions of psychodynamic therapy or CBT, according to random assignment. A number of alliance dimensions were found to be statistically linked to therapy outcomes and therapist perceptions of alliance were found to be more closely correlated with outcome than client perceptions. Some aspects of alliance were correlated with some aspects of outcome and these researchers commented that ‘we have the impression that techniques for measurement and analysis of the alliance have outstripped theory, so that we and other alliance researchers are faced with differentiated results that we do not understand’ (p. 800). Building upon previous publications these authors concluded that the alliance was a consequence of a responsive process, not a number of behaviours that could be counted, and therefore suited to correlational research, unlike process-outcome correlations that looked at behaviours. This research was evidence that strengthening the alliance was good for outcome but not something that could be ‘arbitrarily and unilaterally’ increased like, say, the rate of transference interpretations.

3.2.3 Causation approaches

The presence of therapeutic empathy was found to exert a differential effect on outcome (more empathy, better outcome) in CBT by Burns and Nolen-Hoeksema (1992). These authors created a structural equation model of therapeutic empathy and recovery from depression with CBT. They wrote that this was the first report that they were aware of that had documented the causal effect of empathy on recovery when controlling for the simultaneous causal effect of depression on therapeutic empathy and concluded that therapeutic empathy had a direct effect on clinical improvement, estimated at $r = .26$. In the context of the research questions it would seem plausible that observed effectiveness of the

person-centred psychotherapies were likely to include some contribution from 'empathy' as an impact of the therapeutic relationship; especially as empathy is a fundamental part of the approach. In contrast Burns and Nolen-Hoeksema also found that adherence to homework also predicted outcome, $r = .14$ from patient ratings of homework adherence and $r = .25$ from therapist ratings. Giving clients homework was not a part of Rogers' theory. Yet Rogers was an empiricist, it was unclear whether Rogers would have amended his theory to take account of findings such as these. Perhaps a person-centred approach would have been to discuss putative homework tasks with clients if they wanted to or asked to. Although it is not clear how clients would know to ask for this without being informed of this possibility.

One possible conception of 'homework adherence' upon outcome was that observed homework adherence was really a subgoal of therapeutic alliance/relationship; clients who worked well with their therapist wanted to please them by doing the homework they were asked to do. This possibility occurred to Burns and Nolen-Hoeksema who sought to test this 'alternative hypothesis' as best they could with the data available. Using SEM they found that empathy contributed directly to outcome when controlling for initial depression severity, homework compliance and other factors. They found that the effect of empathy was direct upon outcome and did not operate by facilitating homework compliance. Instead homework compliance had a direct effect on outcome independent of the perception of empathy from the CBT therapist. These authors speculated as to the cause of homework compliance on outcome, noting that CBT theory was that 'self help assignments may directly reduce depression by teaching patients to cope with dysfunctional attitudes and behaviour patterns as hypothesised by cognitive and behaviour therapists' (p. 447). Alternative hypotheses these authors noted were that 1) adherence to homework might be as a consequence of an unmeasured variable such as motivation that caused recovery or 2) that improvement might motivate homework completion. Problems with this kind of causation research include those of temporal precedence and external unmeasured variables that might also be causative. However, what was clear from this research was that both perceptions of empathy and something to do with homework had independent and direct effects on recovery from depression.

Subsequently Burns and Spangler (2000) asked 'Does psychotherapy homework lead to improvements in depression in cognitive-behavioural therapy or does improvement lead to increased homework compliance?'. Building upon the previous data from Burns' 1992 joint paper these authors sought to address temporal precedence and unmeasured variables and

concluded that homework adherence contributed directly to depression outcome. These authors noted the difference between a large effect (a net gain of 14-16 BDI points, from doing homework versus not) and a small correlation (they estimated the effect of homework adherence at a correlation coefficient $r = .19$ to $.21$ depending upon the measures used).

To the present author, and perhaps to others, the Burns and Spangler (2000) paper is impressively complex, filled with complicated diagrams and somewhat esoteric statistics. In a subsequent commentary Kazantzis, Ronan, and Deane (2001) pointed out, whilst Burns and Spangler was a commendable piece of work, fundamentally they had incorrectly concluded causation from correlation (Field, 2009). In particular that a number of alternative hypotheses had not been examined e.g. 1) homework compliance could have been a proxy dependent variable, previous research had shown patient improvement had led to patient over-estimation of homework compliance; 2) possible measurement error in homework compliance ratings could have explained the findings; 3) homework compliance could have been a confound of therapist competence, perhaps better therapists administer homework better, etc. Kazantzis et al. concluded that Burns and Spangler had 'provided a service to the field' (p. 1081) by their paper; and for those who would seek to assess the impact of the therapeutic relationship, a related endeavour, this was a cautionary note.

Whilst Kazantzis et al. (2001) praised the efforts of Burns and Spangler (2001) they also pointed out that an appropriate test of causation would include a prospective test (rather than a retrospective correlational study) of homework's effects incorporating appropriate measures of homework completion and therapist competence. In addition to these experimental and statistical points Burns and Spangler themselves pointed out the combined difficulties of relatively small sample sizes available to psychologists, measurement error and lots of small-medium sized correlational effects that could lead to non-significant findings in research of this kind. In the context of considering the impact of the therapeutic relationship these are related concerns, together with those of only making appropriate inferences from findings.

In a subsequent review article Thase and Callan (2006) wrote about 'the role of homework in cognitive behaviour therapy of depression'. They cited the meta-analysis of Kazantzis, Deane, and Ronan (2000) who found 'strong evidence that interventions that included homework were more effective than interventions that did not (weighted $r = .36$)' (p. 168). The meta-analysis suggested that homework-outcome effects with depression outcome tended to be larger than those with anxiety. Thase and Callan pointed out that the process-outcome

correlation was likely more complicated than it first appeared because of the possible presence of moderators of homework adherence. In their analysis putative moderating variables were put into three groups: task, therapist and patient variables. Firstly, they considered the impact of the perceived and objective difficulty of the task. Length of homework task had been found to be one aspect of its difficulty and written tasks tended to be perceived as more difficult. Secondly, therapist variables included individual differences in enthusiasm for homework, the regularity with which a therapist gave homework assignments plus the therapist's skill at describing the homework. Finally, client variables included personality disorder, symptom severity, homework preferences, perfectionism, fear of failure, fear of displeasing the therapist, how patients attributed their depression i.e. their own fault, bad luck, etc.

A further interpretation of homework could be that this is comparable to a 'dose-effect' relationship (Barkham, et al., 2006, Minami, et al., 2007) and see later. Clients who do homework could be considered by analogy to have an extra non-therapist session compared with those who do not. It might be necessary to control for the nature of the task by giving out non-therapy homework and even this could still be considered part of a 'dose-effect' because it was at the request of the therapist.

The point here is an important methodological one. Therapists giving out homework is a relatively simple system, compared with the whole therapeutic enterprise, of which it is a subset. Yet, seeking to understand the impact of homework has proven difficult. Hopefully the difficulties inherent in a relatively small part of therapy point to the potentially much larger difficulties in understanding the impact of the therapeutic relationship in person-centred psychotherapy. In therapies that support homework this is an additional aspect of the therapeutic relationship. One question for person-centred therapists, in considering the clinical effectiveness of the person-centred psychotherapies, is whether with convincing evidence for an independent impact on outcome, person-centred therapy should adopt, or at least test, homework as an evidence-based intervention. At present the approach of this author is to discuss possible homework tasks with a client if they indicate a wish to do something outside of therapy to help themselves, e.g. addressing agoraphobia by phased exposure. This comes from within the frame of reference of the client and can thus be considered 'client-centred' (Rogers, 1940, p. 163). A more challenging scenario could be a situation in which homework tasks were an evidence-based intervention and a person-centred

therapist refused to mention the possibility of these out of their own (therapist-centred) sense of ‘methodological purity’; potentially a difficult ethical issue for some.

3.2.4 Common factors

In 1992 Norcross and Goldfried published their ‘Handbook of psychotherapy integration’ which featured a chapter by Lambert (1992) on ‘Psychotherapy outcome research: Implications for integrative and eclectic therapists’. Perhaps following his detailed critique of research on Rogers’ theory (Lambert et al., 1978) and his review of the effectiveness of psychotherapy (Lambert, et al., 1986) Lambert put forward an idea of ‘common factors’ spanning school-based approaches to psychotherapy being responsible for observed changes as a consequence of therapy. This research resulted in an estimation of the percentage contribution four therapeutic factors made to psychotherapy outcome (Lambert, 1992). Lambert explained this estimation was based upon many decades of empirical research dealing with ‘a large range of adult disorders and a variety of research designs, including naturalistic observations, epidemiological studies, comparative clinical trials and experimental analogues’ (pp. 97-98). Lambert made clear that no ‘statistical procedures’ were used in these estimations and the percentages appear ‘more precise than is warranted’. A central premise of this kind of analysis was that ‘there is little evidence to suggest the superiority of one school or technique over another’ (p. 103). This view has certainly not been shared by all authors; see for example Siev & Chambless (2007). Nevertheless, Lambert’s ‘common factors’ approach has received some support and the percentages reported in 1992 had remained unchanged for ten years (Lambert & Barley, 2002) and were as follows:

- 40% Extratherapeutic factors e.g. spontaneous remission, fortuitous events, social support, etc.
- 30% Common factors, variables found in most therapies e.g. empathy, warmth, acceptance, encouragement of risk taking, client and therapist characteristics, confidentiality of the relationship, therapeutic alliance, process factors etc.
- 15% Expectancy e.g. placebo effect, client’s knowledge that they are being treated, client’s belief in treatment technique and rationale, etc.
- 15% Techniques, factors specific to a particular therapy e.g. biofeedback, hypnosis, systematic desensitisation, etc.

(Lambert & Barley, 2002)

In a correspondence with the author (Lambert, personal communication, 7th October 2006) an update was provided that estimated the relationship between outcome and other variables as:

- 40% Extratherapeutic factors
- 35% Common factors
- 20% Therapist effects
- 5% Techniques

The author of these estimates cautions there is no direct way of making these estimates (Lambert, personal communication, 7th October, 2006). Although widely referred to, for example by Cooper (2008), and perhaps reasonably widely accepted, it is important to note that these 'percentages of improvement in psychotherapy patients as a function of therapeutic factors' are perhaps 'illustrative', 'indicative' or 'suggestive' rather than a direct consequence of some statistical technique e.g. meta-analysis. Nevertheless there was some direct evidence, as outlined above, for the therapeutic relationship/alliance/empathy as a contributor to outcome, e.g. Horvath and Greenberg (1989), Burns and Nolen-Hoeksema (1992).

In his 2001 book Bruce Wampold revisited the psychotherapy literature and made the case for equivalence of outcomes between schools and for factors common to different schools accounting for similar outcomes.

Sachse and Elliott (2002) reviewed process-outcome research on humanistic variables. In addition to looking at microprocess research findings and implications for practice these authors summarised macro level humanistic process-outcome research. The approach they took to this was look at what they termed 'core' humanistic therapeutic variables (empathic understanding, acceptance/affirmation, and genuineness/congruence), together with 'other' therapist process variables linked to the humanistic approach (therapeutic alliance, directiveness and process-directive methods e.g. two-chair dialogue), processes central to humanistic therapies (self exploration and experiencing) together with other client process variables (client role involvement, client passivity, client openness and positive versus negative affective reactions to therapy). This was not a structured meta-analysis as such, although see Elliott's later co-authored work on empathy (Bohart, Elliott, Greenberg, & Watson, 2002, and Elliott, Bohart, Watson, & Greenberg, 2010), and was rather more a presentation of some of the findings from humanistic and related research. Whilst Gurman

and others were critical of, say for example, work by Barrett-Lennard (1962) these authors presented his work as an empirical result that supported the significance of therapist empathic understanding as a predictor of outcome. These authors concluded that on balance understanding the client empathically was generally associated with positive outcome. Similarly Elliott and Sachse described therapist acceptance of the client as a constructive response although noted great variance in the results and called for more research. In terms of congruence these authors found mixed results, some positive, some finding no connection with outcome, and concluded that congruence was a 'potentially effective therapist condition' (p. 88) and that further research was required. In particular they referred to the 'abuse of the drug metaphor' work of Stiles and Shapiro and pointed out that more congruence may not necessarily lead to better outcomes and in fact 'more congruence' could be detrimental to some clients.

Orlinsky, Rønnestad, and Willutzki (2004) reviewed what they called 'fifty years of psychotherapy process-outcome research'. A large number of process factors were reviewed and the evidence presented as a 'tally count' of studies that found positive, negative or zero impact outcome, often from the perspective of different process raters e.g. patient, therapist or independent rater. A tally count such as this was not quite the same as a meta-analytic synthesis of research findings. These authors found 62 studies in favour of a positive impact for empathy on outcome, 53 studies finding no significant impact and no studies with a negative impact on empathy. In terms of congruence they found 23 studies with positive impact on outcome, 36 with no impact on outcome and 1 with a negative impact on outcome. Orlinsky et al. found 87 studies with a positive impact on outcome for 'therapist affirmation (versus negativity) toward patient', 63 with no significant impact and 4 with a negative impact. In terms of 'therapeutic bond', the name these authors gave to 'therapeutic alliance' they found 87 studies with a positive impact on outcome, 44 with no discernible impact and 1 with a negative impact. Many other process-outcome variables were also presented

In the same volume Clarkin and Levy (2004) considered the impact of client variables, a subject that Rogers appeared to want to avoid, and concluded amongst many other conclusions, that clients with a co-morbid personality disorder were unlikely to benefit in therapy as much as those without.

3.2.5 Meta-analytic approaches

Shadish and Sweeney (1991) brought together outcome meta-analyses, questioning the Dodo bird verdict, and research on third variables that link outcomes and process, mediators and moderators. Their argument was that whilst there was plenty of research showing on average clients receiving psychotherapy do better than clients not receiving psychotherapy a more 'intelligent' view was required about 'when, where, why and how therapy works' (p.883). Their approach was to look at mediators and moderators of outcome and they found differences in effect sizes between behavioural and non-behavioural therapy outcomes based on variables that they found influenced effect sizes. Their conclusion was that Dodo birds are not very smart and shouldn't be allowed to hand out prizes because a more intelligent view of the research would find differences between therapy outcomes. Whilst the approach taken by these authors was certainly complex the findings were not entirely convincing in uncovering the mediators and moderators of *therapy* and seemed to point more to the problems of doing research, and the mediators and moderators of *research findings*, than about therapy outcome itself, e.g. finding that behavioural studies that used measures with low reactivity yielded lower effect sizes than non-behavioural studies that used measures with low reactivity perhaps says more about research methods than practice. These authors claimed not to have solved the problems of therapy research but perhaps to have provided some directions for others to follow.

Martin, Garske, and Davis (2000) examined the impact of the therapeutic alliance with outcome. They identified 79 studies (58 published, 21 unpublished) and found that the alliance was moderately related to outcome, sample size weighted $r = .22$ (see below). These authors concluded that if a proper alliance was established between a patient and a therapist that the alliance might be therapeutic in and of itself, regardless of other psychological interventions and that the alliance was predictive of outcome, whatever the mechanism underlying the relationship.

Norcross (2002) published the results of an APA task force he established 'to identify, operationalise and disseminate information on empirically supported therapy relationships' (p. v). The main conclusions of the task force were that some elements of relationships were 'demonstrably effective' in therapy and these were the therapeutic alliance, cohesion in group therapy, empathy and goal consensus and collaboration. Horvath and Bedi (2002) were the contributors for the chapter on the working alliance and they published a meta-analysis of the 90 studies on the relationship between the working alliance and outcome and concluded

that the effect of the alliance on outcome in most therapy situations was somewhere between $r = .21$ and $r = .25$. Bohart et al. (2002) contributed the chapter on empathy and their 'best estimate' for the size of the effect for empathy was $r = .32$, a medium effect size. This effect size was reported as being of a similar order to previous meta-analyses for the therapeutic alliance-outcome correlation.

The Norcross task force also concluded that some elements of relationships were 'promising and probably effective' in therapy and of particular relevance to the person-centred approach were the chapters on positive regard (Farber & Lane, 2002) and congruence (Klein, Kolden, Michels, & Chisholm-Stockard, 2002). These authors provided 'tally counts' of studies showing positive, negative or no impact on client outcome, rather than a meta-analytic synthesis of an effect size. The other elements receiving this same task force 'rating' were feedback, repair of alliance ruptures, self disclosure, management of counter transference and quality of relational interpretations.

The overview in the Norcross volume (Lambert & Barley, 2002) described above concluded that 'this review would lend some support to the person centred concepts of facilitative conditions and their proposed influence on client progress' (p. 23).

The findings of Bohart et al. (2002) could be seen in the light of those of Simpson, Orinda, & Ickes (2003) who examined 'when empathic accuracy hurts and when it helps' in the field of marital interactions. They found that what they termed 'relationship threatening behaviour' was associated with greater pre- to post- declines in perceived subjective closeness when the behaviour had greater perceived empathic accuracy on the part of the perceiver and trained observers. It was important to note from this study that 'empathy' was not unidirectional, it was not the case that 'more is always better'. To some extent this was the point made by Stiles (1988), about therapy being 'systematically appropriately responsive'. In this case more empathy was hurtful, not therapeutic. Beyond the more or less argument was also the manner in which the empathy was 'used', to hurt rather than to help, or in the context of person-centred therapy, 'to understand'. It was not simply the amount of empathy but also what went with it, perhaps what might be put in Rogerian terms as a lack of positive regard; pointing to the necessity of considering all the relationship factors together (Watson, 1984). In terms of the present research and the impact of the therapeutic relationship this pointed to the importance of considering all the factors together and of measuring these in a manner that captured the sense of 'appropriate responsiveness', not

simply counting the number of empathic statements, or similar. After the further debate that arose from Stiles and Shapiro (1994) this appeared to be where Stiles had got to with the publication of Stiles, et al. 1998, cf. the drug metaphor, that an appropriate measure of relationship/alliance should encapsulate 'appropriate responsiveness'.

In a preview (Norcross 2010) of a soon to be published book on 'evidence-based therapy relationships' (Norcross, 2011), Norcross and Lambert (2010) reiterated the findings from the literature that: a) psychotherapy is effective, with typically 75-80% of patients who enter therapy showing benefit, and that; b) irrespective of therapy-type, 'the therapy relationship makes substantial and consistent contributions to patient success in all types of psychotherapy studied (for example, psychodynamic, humanistic, cognitive, behavioural, systemic)' (no pagination specified). Meta-analyses were conducted for key elements of the therapy relationship and the results of this are summarised, Table 1.

Table 1

Table 1 : Summary of meta-analysis of key elements of the therapy relationship based on Norcross, 2010.

Element of therapy relationship	r ^a	Reference
Alliance in individual psychotherapy	.28	Horvath, Del Re, Fluckiger, & Symonds, 2010
Alliance in child and adolescent psychotherapy	.19	Shirk & Karver, 2010
Alliance in couple and family therapy	.26	Friedlander, Escudero, Heatherington, & Diamond, 2010
Cohesion in group therapy	.25	Burlingame, Theobald McClenon, & Alonso, 2010
Empathy	.30	Elliott, Bohart, Watson, & Greenberg, 2010
Goal consensus	.34	Shick Tryon & Winograd, 2010
Collaboration	.33	Shick Tryon & Winograd, 2010
Positive regard and affirmation	.27	Farber & Doolin, 2010
Congruence/genuineness	.22	Kolden, Klein, Wang, & Austin, 2010
Collecting client feedback	.23-.33	Lambert & Shimokawa, 2010)
Repairing alliance ruptures	.24	Safran, Muran, & Eubanks-Carter, 2010
Countertransference	-.16	Hayes, Gelso, & Hummel, 2010
Managing countertransference	.56	Hayes, et al., 2010

Note: ^a effect size expressed as partial correlation coefficient.

Norcross and Lambert (2010) made the point that alliance and cohesion are composed of multiple relationship elements, there is evidence for the effect of more specific relationship elements (empathy, goal consensus, collaboration, positive regard and affirmation, congruence/genuineness) and specific relationship behaviours that promote the relationship (collecting client feedback, repairing alliance ruptures, managing countertransference). Countertransference had a small negative impact on client outcome and the management of countertransference had a large positive impact on outcome. The extent to which these effect sizes interact or are 'additive' is unclear from the preview of the forthcoming book. For example Rogers would recognise empathy ($r = .30$), positive regard (.27) and congruence (.22), although it is doubtful that taken together these would constitute $r = .79$, or perhaps not much more than the highest of these, $r = .30$. This also highlights a point about the specificity of the definition of the different relationship elements and the extent to which they do, or do not, overlap. For example Horvath and Greenberg (1989) found a correlation between the alliance and empathy. It is also unclear to what extent these relationship elements are differentially important for different outcomes e.g. is empathy more important for depression than anxiety? Norcross and Wampold (2010) summarised the evidence for adapting the relationship to the characteristics of the individual patient. They found statistically significant evidence from their meta-analysis for adapting to the reactance level of the patient(defiant/compliant), readiness for change, culture, coping style (internaliser or externaliser), religion or spiritual belief and preferences in terms of therapy school, treatment format (individual, family, group), relationship style, treatment length, etc. These authors conclude that 'psychotherapists can create a new, responsive psychotherapy for each distinctive patient and singular situation – in addition to his/her disorder' (no pagination specified) cf. comment by Sechrest (1994) 'one would have to validate the efficacy of each psychotherapist separately' (p. 952).

3.2.6 Some Recent studies

Klein, et al. (2003) looked at therapeutic alliance in depression treatment controlling for prior change and patient characteristics. These authors found that early alliance significantly predicted subsequent improvement in depression symptoms, controlling for nine effects that might affect measures of the alliance. Measures of the alliance and depressive symptoms were independent and based on different methods, avoiding shared measurement error, cf. Watson (1984)

Cloitre, Chase, Miranda, and Chemtob (2004) considered the related contributions of the therapeutic alliance and negative mood regulation to the outcome of a two-phase treatment for childhood abuse-related post traumatic stress disorder (PTSD). They found the strength of the early alliance predicted improvements in PTSD and that this relationship was mediated by participants improved capacity to regulate negative mood states. These authors concluded that 'the therapeutic relationship may be an especially "active ingredient" in the remediation of childhood abuse-related PTSD' (p. 414-5). The rationale for this 'active ingredient' appeared to these authors to be that once a good relationship was established patients were better able to tolerate distress, in the relationship, when exposed to difficult feelings associated with the trauma and of the 'apparent importance of the therapeutic alliance for achieving negative affect regulation' (p. 415).

A large naturalistic study of 6,146 clients seen by 581 therapists found about 5% of the variance in outcomes was due to therapist variability (Wampold & Brown, 2005). The authors noted there was a wide range of estimates for the percentage of outcome variance due to therapist effects in the literature and identified reasons for this (e.g. statistical methods used, heterogeneity of patients in sample, etc.), concluding that within the clinical trial literature about 8% of outcome variance was due to therapist effects. The authors found a similar level of effect in their naturalistic sample, 7.8% of variance due to therapist effects, and this reduced to 5.5% when differences in initial severity of patients seen by therapists was controlled for. When initial severity was greater, variability due to therapists was greater; the 5.5% figure was for 'average severity' patients. The authors explained their relatively low figure of about five per cent of the variance in outcomes being due to the therapist in their naturalistic sample, compared with eight per cent for clinical trials, by the greater variability amongst patients in usual practice, meaning there was less variation that could be accounted for by therapist effects. The five per cent figure also took account of sampling error and was therefore a relatively 'pure' estimate for therapist effects; on a comparable basis the authors estimated the proportion of variance explained by the type of treatment delivered at 1-2%. Subsequent research involving Bruce Wampold (Baldwin, Wampold, & Imel, 2007) with another naturalistic sample suggested the outcome variance explained by the therapeutic alliance was 3%. Note the 3-8% percentages for 'variance due to therapist effects' are very much lower than the 20-30% for therapist effects/ common factors posited in the 'percentages contributing to psychotherapy outcome' models (Lambert, 1992, Lambert & Barley, 2002,

Lambert, 2006). There was a difference between the contribution a factor makes to outcome and the outcome variance a factor explains; the latter is likely to be smaller than the former.

Wampold and Brown also found that more effective therapists in one time period were also similarly effective in a subsequent time period; less effective therapists were also predictable from one period to another. This study found patients of more effective therapists received more benefit from concurrent medication than those of less effective therapists, although the authors caution that this finding was based on a relatively small subset of the overall database. In addition to prospectively enhancing the effectiveness of medication the therapeutic relationship may also influence adherence to a medication regime (Hays, Ordway, & Di Matteo, 1992, DiMatteo, et al., 1993), and this 'influence' can be 'taught' to prescribers (Qureshi, Hatcher, Chaturvedi, & Jafar, 2007).

Zuroff and Blatt (2006) reanalysed data from the National Institute for Mental Health (NIMH) Treatment of Depression Collaborative Research Programme (TDCRP, Elkin, et al., 1989) and found that all four treatment arms – CBT, interpersonal psychotherapy (IPT), anti-depressant medication (imipramine) plus clinical management and pill placebo plus clinical management – had outcomes that were differentially affected by a Rogerian measure of the therapeutic relationship at the start of therapy (better relationship/better outcome). The Barrett-Lennard Relationship Inventory (BLRI, Barrett-Lennard, 1962), the Rogerian measure of the therapeutic relationship, see below, completed at the second treatment session had predicted a composite measure of clinical improvement across all treatment conditions (Blatt, Zuroff, Quinlan, & Pilkonis, 1996). The 2006 more rigorous analysis controlled for early clinical change and relationship variables and found that 'across CBT, IPT and the two medication with clinical management conditions, the perceived quality of the early therapeutic relationship, adjusted for early clinical improvement, predicted the rate of decrease in maladjustment subsequent to the measure of the relationship' (p. 137). In fact a composite score derived from summing the empathy, regard and congruence component of the relationship inventory was used (D. C. Zuroff, personal communication, 29th January 2008). It was this composite score that suggested 'the perceived quality of the therapeutic relationship early in the treatment process contributes directly to multiple dimensions of outcome during brief treatment of depression, including symptom reduction, improved global adjustment and EAC (Enhanced Adaptive Capacities)' (p. 137). The higher quality relationship also predicted lower levels of maladjustment throughout the 18 month follow-up period. Re-running the analysis using each of the four subscales of the BLRI (regard,

empathy, unconditionality or congruence) individually, instead of the composite factor, showed each subscale contributed to outcome to a similar extent, none of the subscales stood out as having a unique role (D. C. Zuroff, personal communication, 19th March 2008).

Hawley, Ho, Zuroff, and Blatt (2006) examined the relationship of perfectionism, depression and therapeutic alliance during treatment for depression. The strength of the alliance significantly predicted longitudinal perfectionism change and perfectionism significantly predicted the rate of depression change during therapy. Reflecting upon the language and concepts Rogers was seeking to convey the conclusions of these authors have some similarity to something Rogers might have written:

Perfectionistic patients often hold maladaptive beliefs about themselves involving harsh self scrutiny, overly critical evaluations of their behaviour, and unrealistically high standards of performance, associated with themes of guilt and inferiority. They often believe that others will be overly critical of their behaviour, having high expectations for their performance that must be met to gain approval and avoid rejection. Once a strong alliance has been established, the therapist's accepting, nonjudgemental and supportive attitudes and behaviours can provide an environment that allows the patient to challenge this maladaptive belief system. Within a collaborative therapeutic framework, the patient becomes capable of disclosing personal information without fear of being rejected or criticised by the therapist. As the content and structure of the patient's mental schemata shift toward more realistic and adaptive beliefs, symptom alleviation occurs as the underlying vulnerability improves. A successful psychotherapy intervention can be seen as providing a collaborative setting in which maladaptive schemata are challenged while working to develop a more realistic, differentiated and integrated belief system' (p. 939-40).

Strauss, et al. (2006) examined a non-randomised trial of cognitive therapy for avoidant and obsessive-compulsive personality disorders. They found stronger early alliances and rupture-repair episodes predicted more improvement in symptoms of personality disorder and depression and speculated that the therapeutic relationship was a 'corrective experience to disconfirm maladaptive schemata' (p. 342). These authors placed their findings in the context of the importance of the early alliance in working with clients with chronic problems such as depression (Klein, et al., 2003) and childhood abuse-related PTSD (Cloitre, et al., 2004). Furthermore, Strauss et al checked, as these authors had, that the therapeutic alliance was not simply associated with early symptom change.

In contrast to Beck (1976), who was also a co-author of Strauss, et al. (2006), Strauss et al. considered that the therapeutic relationship could be causative in its own right, not simply a non-specific convenient means of implementing causative CBT techniques. Rogers 1957 paper was about the relationship conditions for 'constructive personality change', not simply symptom relief. These authors cited the ideas of Beck, Freeman, Davis, and Associates (2004) that the therapeutic alliance could be particularly important in Axis II populations and pointed to their own data which suggested rupture-repair episodes could be therapeutic if handled properly. Unusually for journal articles which were ordinarily written in the past tense Strauss et al. look forward to 'the next phase of Adele M Hayes's treatment development research focuses on identifying therapist strategies associated with better and worse early alliances and rupture outcomes to improve treatment retention and treatment outcomes in this prevalent and challenging population' (p. 344). In the context of the research questions, if person-centred psychotherapy were clinically effective and there was an impact of the therapeutic relationship on outcome it seemed plausible that person-centred psychotherapy might be able to make some useful contribution to research of this kind. This perhaps reinforced the idea that research questions of this kind could have some validity in terms of making a contribution to therapeutic literature.

Lutz, Leon, Martinovich, Lyons, and Stiles (2007) looked at therapist effects in outpatient psychotherapy using a three level growth curve approach. They concluded that 8% of the total variance and 17% of the variance in rates of patient improvement could be attributed to therapists and that 'the contention that... therapist differences tend to be larger in naturalistic studies than in controlled trials finds some support in our findings... therapist effect on outcome was approximately double that typically observed in the clinical trials' (p. 36).

Spinhoven, Giez, van Dyck, Kooiman, and Arntz (2007) published a paper about the therapeutic alliance in therapy for borderline personality disorder. Two conditions, schema-focused and transference-focused therapy were found to have significantly different ratings of the alliance. The authors sought to make sense of this finding by hypothesising that the higher ratings in the schema-focused condition possibly reflected the effort in this condition to 'connect to the patient by adapting an unthreatening and supportive attitude and to develop mutual trust and positive regard' (p. 112) and citing the work of Beck, Freeman, Davis, & Associates (2004) as the foundation for this theory, cf. Beck (1976). The researchers found that negative ratings of the alliance from both therapists and patients predicted drop-out and increasingly positive ratings of the alliance predicted improvement and these authors

concluded that the development and maintenance of the alliance during the first year of therapy was crucial, keeping clients in therapy so that they might benefit. The authors speculated that the therapeutic alliance and techniques might interact and may serve to facilitate a change process.

In 2009 Crits-Christoph, et al. published work on the alliance in motivational enhancement therapy (MET) and counselling as usual (CAU) for substance use problems. Whilst these authors expected a significantly higher alliance for the MET condition, given its 'extensive focus... on empathy, acceptance and positive regard' (p. 1132) in fact there was no statistically significant difference in alliance between the two conditions. Like Baldwin, et al. (2007) these authors found it was the therapist-to-therapist variability in average alliances that predicted outcome, not the patient-to-patient variability. In fact these authors found that scoring the alliance at extremes, compared with other patients, was associated with relatively poorer outcomes and that the relationship between alliance and change in drug use was not – linear but quadratic; scoring the alliance at an extreme was associated with poorer outcome. Crits-Christoph, et al. concluded that since the alliance was found to play an important role in outcome that training 'therapists in ways that might enhance the alliance are justified' (p. 1133). It appeared possible that the person-centred approach with its emphasis on the alliance/relationship may have some contribution to make to therapy literature and practice.

Perhaps related to the findings of Crits-Christoph, et al. (2009), South, Turkheimer, & Oltmanns (2008) looked at personality disorder and marital functioning. They found low levels of marital satisfaction and high levels of verbal aggression associated with more extreme scores for personality disorder, especially borderline and dependent personalities. These authors found the processing dynamics in people with personality disorders led to misunderstandings, misconceptions, poor communications and unhappy relationships. They suggested personality traits were important in understanding why relationships thrive or falter. It was plausible to consider that there may be some similar phenomenon in therapy, both in terms of clients' experience and assessment/recording of the therapeutic relationship/alliance with comorbid personality disorder. Something of this kind maybe what went on in the different ways in which clients reported on the therapeutic alliance of the kind observed in Baldwin et al. (2007) and Crits-Christoph (2009). Given the intention of this research to measure outcomes and the known impact of the presence of a co-morbid personality disorder in moderating outcomes (Clarkin & Levy, 2004) it seemed important to consider the impact of a comorbid personality disorder in order to contextualise the outcomes

study. Furthermore, given the plausibility that presence of a personality disorder could affect both the measurement and the impact of the therapeutic relationship it appeared important to bear these factors in mind.

3.2.7 Anti-depressant medication

It was mentioned above that as naturalistic research in order to make a legitimate claim that any observed effect was due to the person-centred therapy this research would need to rule out any alternative explanations. One of the alternative explanations mentioned by Clark et al. (2007) that would need to be considered was that pre-post changes in outcome scores could be due to concurrently administered medications. Researching clients with depression, anxiety or non-specific distress it seemed plausible that at least some of these clients could be taking concurrently administered medications and it would be important to know who was taking what medication, as this could provide an alternative explanation for observed changes. In particular some of the clients in the research could be taking anti-depressants as these are routinely prescribed for depression, co-morbid depression/anxiety, anxiety or other psychological/psychiatric difficulties and it was important to consider what possible effects anti-depressant medication might have on the study.

The efficacy/effectiveness of anti-depressants has been questioned. Kirsch and colleagues have conducted a number of meta-analyses of drug company research, including those studies submitted to the US Food and Drug Administration, and their wide range of publications includes for example: Kirsch & Sapirstein (1998), Kirsch, Moore, Scoboria, and Nicholls (2002), Kirsch (2002), Kirsch, Scoboria, and Moore (2002), Moncrieff and Kirsch (2005), Moncrieff (2006) and Kirsch, et al., (2008). In general terms these authors found that pill placebo was apparently powerful in relieving depression as this typically matched 70-80% of the efficacy of anti-depressant medication. In some cases the remaining 20-30% of incremental effect from the 'active treatment' condition was less than that required for reliable and clinically significant change (Jacobson & Truax, 1991).

Taken together these articles pointed to a 'minimal' (i.e. non-clinically significant) role for anti-depressants in recovery from depression and have themselves been controversial and given rise to many comments, for example: Salamone (2002), Antonuccio, Burns, and Danton (2002), Brown (2002), Greenberg (2002), Hollon, DeRubeis, Shelton, and Weiss (2002), Moerman (2002), Munoz (2002), Rehm (2002) and Thase (2002).

In addition to inquiring about NHS investment in therapy research by theoretical model, the present author also inquired about NHS spending on anti-depressants (Freedom of Information requests). In the 16 years to 2006-7, the period when Selective Serotonin Reuptake Inhibitors (SSRIs) became available, the NHS in England spent £3,115m on prescriptions for six anti-depressants (Fluoxetine/Prozac, Paroxetine/Seroxat, Sertraline/Lustral, Venlafaxine/Efexor, Nefazadone/Serzone and Citalopram/Cipramil) in primary and secondary care settings (Weston, 2008 and Weston & Weston, 2008). It seems in the UK a comparatively large amount of money has been spent for a comparatively small benefit directly attributable to anti-depressants.

Kirsch and colleagues have wondered about what they term ‘placebo effects’ and have sought to understand the components of these effects; finding that a ‘placebo’ intervention with ‘augmented’ human contact is more powerful than one without human contact (Kaptchuk, et al., 2008). For these researchers ‘augmented’ human contact was a component of ‘the placebo effect’. There is perhaps an issue of nomenclature: one researcher’s ‘placebo effect’ is maybe another researcher’s ‘therapeutic relationship effect’.

In the field of prescription medications the implications of therapist effects were highlighted in stark terms at the APA conference (Brown, 2007) who pointed out that in the TDCRP, comparison of the antidepressant and pill placebo legs of the trial showed 9.1% of the outcome variance was due to the psychiatrist and 3.4% due to the medication; who prescribed accounted for 2.7 times more outcome variance than what was prescribed. The top third of psychiatrists achieved better outcomes with placebo than the bottom third achieved with the antidepressant (McKay, Imel, & Wampold, 2006).

Whilst it may have appeared that pill placebo was apparently powerful in relieving depression it could be that at least some of the effect was really a therapist effect, since typical drug trials include some ‘clinical management’, a regular review of symptoms and medication with a psychiatrist. This could account for at least some of the apparently large effects of pill placebos in depression RCTs.

Nevertheless, whilst anti-depressants might appear to have only small effects on outcome it was important to consider in this research whether any apparent effect of person-centred therapy could be due to concurrent medications.

3.2.8 Dose effect relationships

Medications are often considered to be ‘dose-dependent’ and this concept may also be applied to psychotherapy where there can also be dose-effect relationships (Barkham, et al., 2006) although see papers by Stiles starting from 1986 and reviewed above re. ‘abuse of the drug metaphor’. The number of sessions a client has can influence their outcome and the evidence appears to suggest, in the absence of other limits, clients will tend to regulate their own therapy ‘dose’ by the use of a client-internalised ‘good enough’ level of outcome (Barkham, et al., 2006).

In the context of ‘outcomes management’ some researchers have sought to understand the dynamics of change, e.g. Stulz, Lutz, Leach, Lucock, & Barkham (2007) looked at 192 clients session-by-session for the first six sessions and identified five ‘change groups’ that clients could be fitted into, and Stulz, Thase, Klein, Manber, & Crits-Christoph (2010), identified three change groups for 504 clients with chronic depression. In addition to providing reference groups to monitor client progress against by way of outcomes management, analyses of this kind prospectively provide a way for identifying which type of treatment is best suited to what type of patient and for process-outcome research to better understand linear cause and effect (cf. Burns & Spangler, 2000). The ‘sudden gains’ literature (e.g. Tang, De Rubeis, Hollon, Amsterdam, & Shelton, 2007) could be considered an early form of this type of analysis. Hopefully in future this type of research will help in the understanding of the impact of the therapeutic relationship, perhaps by monitoring the development of the relationship along with symptom change as a function of time. Stiles, et al. 1998 discovered a complex interplay of symptom changes correlated with alliance at different stages. Subsequently, Svarthberg, Seltzer, Choi, & Stiles (2001) investigated cognitive change before, during and after short-term dynamic and non-directive therapies in a preliminary growth modelling study and found that ‘patients in both conditions changed significantly after pretherapy evaluation and diagnostic interviews as well as during the second half of therapy’ (p. 201). Both conditions led to cognitive changes and this suggested cognitive changes may not be specific to CBT but may also be an impact of the therapeutic relationship and perhaps a mechanism of clinical effectiveness in person-centred psychotherapy.

In the section on ‘the case against Rogers’ mention was made of Lesser (1961). Gurman (1977) stated that the approach Lesser had taken was an appropriate one, and this was further supported by the ‘abuse of the drug metaphor’ work. It seemed plausible that Lesser’s

research may have been affected by a dose-response effect in that the clients at Lesser's college had to have their outcomes measured to fit in with his PhD, rather than when they had finished therapy. The previously mentioned meta-analysis of Minami, et al. (2007) quantified a dose-effect relationship with the Hamilton Rating Scale for Depression (HRSD) which they quantified as +.053 ES(d) per week.

It was important for this research to consider dose-response effects for both the outcomes and process-outcomes part of the experimental work.

3.2.9 Concluding comments on process-outcome research

This section has reviewed the process-outcome literature to the benefit of both the findings therein with regards to the possible impact of the therapeutic relationship on outcomes and the research methodologies to test this. The comments made by reviewers of the early research on Rogers' theory were helpful in pointing the way for this research e.g. Gurman (1977), Lambert et al. (1978), Watson (1984). The subsequent 'abuse of the drug metaphor' papers clearly helped the whole field, as well as this research. The findings reviewed in the causation approaches, common factors, meta-analytic approaches and recent studies sections served to both inspire the author to attempt the process-outcome research, the evidence appeared to point in the direction that the therapeutic relationship as defined by Rogers could likely correlate with outcome and provided further methodological information, not necessarily to do the methods employed therein but factors to consider. Two factors to consider for both the outcomes and process-outcomes work were the impact of anti-depressant medication and dose-effect relationships.

The next section looks at the design of the experimental part of this research, in terms of an overview of design issues and separate sections on measuring outcomes and measuring process-outcome correlations.

3.3 Research Design.

3.3.1 Overview of design issues.

The ‘gold standard’ research design favoured by NICE and APA in making evidence-based recommendations is the randomised controlled trial (RCT). Some have argued that RCTs are not always the best or most appropriate research methodology. RCTs have been criticised for not being practical or ethical for evaluating public health interventions (Victora and Bryce, 2004). Whilst the Consolidated Standards of Reporting Trials (CONSORT) statement has improved the quality of RCT reporting there has been nothing comparable for non-randomised designs, until recently; the TREND statement (Transparent Reporting of Evaluations with Non-randomised Designs sought to remedy this (Des Jarlais, Lyles, Crepaz, & TREND Group, 2004). This research is presented in accordance with TREND guidelines and the publication guidelines of the APA (American Psychological Association, 2001).

The present research was an attempt to look at whether person-centred therapy had any helpful impact (outcome study) and whether any outcome might be related to the therapeutic relationship as defined by Rogers (process-outcome study). Without a budget to do a randomised controlled trial (RCT) this research was structured as an evolving non-controlled naturalistic study of person-centred therapy as practised by the author, in the author’s private practice (PP), and colleagues at the University of East Anglia University Counselling Service (UCS). The research at the UCS provided the opportunity, for a subset of the clients in this research, to do a single group waitlist own-control quasi-experiment; with a short un-standardised wait period (Cook & Campbell, 1979). Additionally data was collected on medication usage by clients and the impact of concurrent medication could also be assessed. This research could be considered a pilot study in preparation for a more structured RCT, in particular finding out where potential research problems might be and perhaps building the case for investment in an RCT. This is in contrast to the huge and highly structured endeavour that Rogers and colleagues undertook that took many years, had hundreds of people involved, cost many millions of pounds in today’s money and was only partially successful in research terms (Rogers, 1967, Weston, 2005).

As described above, Rogers was a keen supporter of quantitative research, although reading some of his work and that of his associates it is not clear that there was a thorough grounding in the disciplines of such. It seems plausible that the piece Rogers (1985) wrote ‘towards a more human science of the person’, that sought to encourage research beyond only

‘quantitative research’ may have played an influence in discouraging proponents of the ‘person-centred approach’ from doing quantitative research. It was perhaps because of the influence of this that when the present author sought to recruit person-centred therapists to participate in the quantitative research reported herein some responded in terms approaching ‘it’s the devil’s work’ and at the very least ‘disrespectful to clients to ask them to fill in questionnaires’.

Firstly, this section looks at issues surrounding the measurement of outcomes and secondly issues surrounding the measurement of process-outcome correlations are explored.

3.3.2 Measuring outcomes.

At its most simple the study of outcomes requires a comparison between clients before and after therapy (Hill & Lambert, 2004). Quantitative instruments are frequently used for these comparisons, often based around diagnostic criteria for particular illnesses (American Psychiatric Association, 2000). Quantitative instruments may be based on the observations of clients, therapists or observers: all of the quantitative instruments used in this research were client completed questionnaires. Rogers’ hypotheses required the client be the source of the outcome ratings, although this is not without problems, for example clients who like their therapist may favourably evaluate outcome, shared method variance, and so on (Watson, 1984). There are strengths and weaknesses associated with the source of the outcome evaluation (Hill & Lambert) and for this research the most theory-specific perspective was that of the client. Additionally there were insufficient resources to use further outcome assessment perspectives such as therapists, judges, clinical experts, significant others, etc.

To be effective questionnaires must fulfil certain criteria (Hill & Lambert), e.g. content validity – ensure the questionnaire does in fact measure what it sets out to measure (Beck & Steer, 1993, pp. 10-12); criterion validity – the questionnaire discriminates between those with and without a particular condition (Beck, Steer, & Brown, 1996, pp. 34-35); construct validity – the degree to which the measure has the theoretically expected relationship with other variables (Dekeyser, Prouty, & Elliott, 2008, p. 47) e.g. a depression measure should correlate with other depression measures (Beck et al., 1996, pp. 25-28); test-retest stability – clients receive similar scores for similar underlying conditions (Beck et al., 1996, p. 25);

reliability – the questionnaire has internal consistency, measured using coefficient alpha, to ensure all items on the questionnaire are measuring a related phenomenon (Field, 2005, pp. 666-676). The suggestion is that coefficient alphas in the range .7 to .8 are acceptable levels of reliability (Field, 2005, p. 668). Assuming an appropriate questionnaire has been used for pre and post observations the appropriate statistical test would be a repeated measures analysis of variance (Field, 2005, pp. 427-482) to see if there was a statistically significant change.

A simple pre and post outcome test takes no account of what change might occur ‘naturally’, without an intervention, the so-called ‘natural history’ of a particular clinical issue (Minami, et al., 2007). On this basis it can be helpful to compare treated and untreated samples of clients, to identify any marginal benefit an intervention may have over and above the natural history. Assuming that clients of both treated and untreated cells start with similar levels of severity an appropriate statistical test to see if there was a statistically significant difference in outcome scores for the two cells would be an independent samples t-test (Field, 2005). An RCT randomises clients to treatment and no-treatment conditions to provide an opportunity to discover whether the treatment has any marginal benefit over natural history. This assumes that randomisation ensures that clients in both conditions are similar, although this assumption may not be borne out, for example the treatment preferences of clients in randomised controlled trials may impact the therapeutic alliance and hence outcome in randomised controlled trials (Iacoviello, et al., 2007). These authors recommended patient treatment preferences need to be considered and controlled for in RCT settings and concluded that ‘because of the potential of preferences on how alliance develops, naturalistic psychotherapy studies in which the impact of such factors is minimised may be warranted to augment findings from RCTs’ (Iacoviello, et al., 2007, p. 197). An alternative to an RCT design is to precede treatment with a wait period and to compare changes under both conditions. Statistically an advantage of this approach is that variation between clients is minimised (they are the same people under different conditions), this is termed a ‘repeated measures design’ (Field, 2005) and the appropriate statistical test would be a repeated measures analysis of variance. A disadvantage of this approach is that people who may ‘get better’ in the wait phase of the research may not participate in the treatment phase and be lost to the analysis.

In a clinical trial setting outcome research is referred to as ‘efficacy research’ and in a naturalistic setting outcome research is referred to as ‘effectiveness research’ (Hill &

Lambert, p. 115, Beutler, et al., 2004, p. 227). As this research was carried out in naturalistic settings it was referred to as ‘effectiveness’ research.

Rogers and colleagues (Rogers & Dymond, 1954) struggled with the ethics of denying patients treatment by randomising them to a ‘no treatment condition’ and used a wait-list control method. In this research, where there was a ‘wait’ this was measured to provide a wait-list control, otherwise clients were treated as they arrived for treatment in the usual run of things.

3.3.3 Measuring process-outcome correlations.

Elliott (2010) recently reviewed ‘change process research’, research that seeks to understand the processes whereby change takes places in psychotherapy. He described four types of change process research: process-outcome, helpful factors, sequential process and significant events. Building upon previous work (Cook & Campbell, 1979, Haynes & O’Brien, 2000) Elliott attempted to assess the strengths and weaknesses of each of these four methodological groups in providing direct causal evidence for change in psychotherapy. It was important to note that quantitative process-outcome research was not the only approach to causal inference and this has strengths and weaknesses.

Compared with the causal inference criteria Elliott identified, he considered that process-outcome research could show covariation, one of the criteria described by Haynes and O’Brien. With optimal research design process-outcome research could demonstrate temporal precedence of putative cause on subsequent effect, could consider alternative causes (see below), and could demonstrate construct validity of cause and effect. However, on its own process-outcome research could not provide a plausible explanation for the proposed cause nor necessarily provide evidence that was directly relevant to clinical practice. In summary process-outcome research provides a part of the evidence linking putative causal process to outcome and not the whole of the required evidence. It was therefore important to place the findings from such research into an appropriate context and draw carefully considered appropriate conclusions about the generalisability of any findings.

Elliott concluded that process-outcome research may have been overused and that the other methods identified provided a necessary complement to the strengths and weaknesses of process-outcome research; an argument for ‘systematic methodological pluralism’. However,

‘at the same time I must confess a continuing fondness for the process-outcome design in spite of the controversy over its use, particularly if practical self-report measures of process (e.g. client ratings of the alliance) are used and temporal precedence is carefully considered’ (p. 132).

4. Introduction to this study

This section introduces the research that follows. Firstly, the need for this research is described, in the context of the literature previously reviewed. Secondly, an overview of the hypotheses is provided, before, thirdly, the main hypotheses are stated in testable terms.

4.1 The need for this research

This research was needed for a number of reasons that are outlined below. Broadly these reasons are to do with the need for further outcome research with person-centred psychotherapy and to do with the need to see if the therapeutic relationship as defined by Rogers had anything to do with the outcomes of person-centred psychotherapy. As with the literature just reviewed, some overlap is inevitable in highly enmeshed concepts, however, for the sake of clarity these reasons are described in separate groupings.

a) Outcomes research

There was a need to research the clinical effectiveness of person-centred psychotherapy.

Reviewers of therapy want evidence of effectiveness when they seek to make evidence-based recommendations e.g. American Psychological Association, Division of Clinical Psychology, Task Force on Promotion and Dissemination of Psychological Procedures (1995), Chambless and Hollon (1998), NICE (2004a, 2004b, etc.). These reviewers found insufficient evidence of empirical support for person-centred psychotherapy, in particular for depression and anxiety. Depression and anxiety are common problems affecting very many people with a large economic cost NICE (2004a, 2004b). There was a need to research the effectiveness of person-centred psychotherapy for depression and anxiety.

Reviewers of the effectiveness of psychotherapy want outcome measures to be diagnostic-specific (NICE 2009a) and there was a need to research the effectiveness of person-centred psychotherapy for depression and anxiety with diagnostic-specific measures e.g. BDI-II and BAI used in this study are widely accepted diagnostic-specific measures. In addition it was reasonably standard practice in the UK to use CORE-OM in psychotherapy research (e.g.

Stiles, et al., 2006, Stiles et al., 2007) and there was a need for this research to use this measure too for comparability.

Whilst there was some evidence for the effectiveness of person-centred psychotherapy (e.g. Elliott et al., 2004) this has not been accepted by all reviewers, e.g. NICE (2009a). Whilst some academics (D. Cramer, personal communication, 24th February 2011) have argued that the case for person-centred therapy has already been made by research such as Ward, et al. (2000) this has not led to recommendation by NICE (2009a). In addition to single studies reviewers also want replication (Chambless and Hollon, 1998) and ideally the contribution of several/many studies to a meta-analysis, or similar. There is perhaps something of a ‘weight of evidence’ argument here that suggested that the more studies the better and there was a need for this research to contribute to the evidence-base.

Ideally reviewers of therapy research want evidence from RCTs, or other ‘well-controlled studies’ (NICE, 2004b) cf. Westen, et al. (2004), Victora, et al (2004), Schwartz, et al. (2004). The present author did not have the resources to conduct an RCT, however a naturalistic study of the kind described in this thesis could make the case for investment in an RCT. Furthermore, in the absence of RCT evidence, the NICE hierarchy of evidence was to accept evidence from ‘well controlled trials’, behind RCT evidence but ahead of ‘expert view’. The extent to which this research was ‘well controlled’ is a judgement for the reader, and see below.

The fact of doing this research has to some extent already been influential in as much as the present author has been asked to do some visiting lectures, made conference presentations, etc. There was a need for this research to demonstrate that person-centred therapy could be subject to outcomes research cf. some therapist’s views re. Rogers (1985). There was a need for this research so that the present author, and hopefully others, would make the collection of outcomes data a routine part of practice (Evans, et al., 2002) and the presentation of quantitative person-centred research a possibility. There was a need for this research for the present author to find out how to do outcomes research. In addition to the idea of a pilot for a possible RCT there was also the need for this research to find out how to/how not to routinely monitor outcomes and conduct outcomes research.

There was a need for this research to pave the way for other subsequent outcomes research on other less studied aspects of human difficulties e.g. PTSD, OCD, disordered personality processes, panic, suicidality, self harm, etc.

Outcomes management has been shown to moderate the effect of therapy (Shimokawa, et al. 2010) and there was a need for this research to test whether some aspects of outcomes management could be applied to person-centred psychotherapy.

Critics of naturalistic research have pointed to rival hypotheses that could explain outcome, rather than the psychotherapy intervention, e.g. Clark, et al. (2007) pointed to a number of alternative hypotheses re. Stiles, et al. (2007). There was a need for this research to examine the validity of these rival hypotheses and related ones to attempt to rule these in or out, as appropriate, in an attempt to make the outcomes observed as a consequence of person-centred psychotherapy ‘well controlled’.

Some of the known moderators/alternative change processes of the effects of psychotherapy were identified from the literature review. Three in particular were identified from the literature as important to consider for a study of this kind and these were a) the presence of a co-morbid disordered personality process as this was known to reduce the size of outcome effects (Clarkin & Levy, 2004); b) the impact of the number of sessions clients have, the dose-response effect (Barkham, et al., 2006, Minami, et al., 2007), especially as some of the clients at the research sites would have limits on the number of sessions they could attend, and; c) the impact of concurrently administered medications, as these were intended to improve conditions such as depression and anxiety, although some research suggested this may not be the case in practice (e.g. Kirsch, et al., 2002).

Supporters of naturalistic research established the TREND guidelines to support the CONSORT guidelines (RCT reporting) and there was a need for this research to utilise the TREND reporting guidelines to uphold the quality of the research.

Combining the views of the critics and supporters of naturalistic research there was a need for this research to report pre-post outcomes, LOCF outcomes and reliable change (deteriorated, improved and recovered) percentages. There was a need for this research to compare the findings with other comparable studies, although the present author was recommended to keep this relatively straight-forward cf. non-central t methodology, see below.

In addition for the outcomes part of this naturalistic research there was a need for this research cf. Stiles, et al. (2006), Stiles, et al. (2007), Clark, et al. (2007) to consider a) the impact of ‘missing cases’ on pre-post effect sizes and reliable change percentages, b) the effect of regression to the mean on pre-post effect sizes, c) the impact of concurrently

administered medications, d) the prevalence of disordered personality processes, e) whether these outcomes were simply the 'best' or 'easiest' clients, and; f) the impact of time. The impact of time manifests itself in a number of ways and there was a need for this research (where possible) to consider: a) wait-controlled effect sizes for those clients who acted as their own wait list control, b) the impact of time on progress during wait and treatment, to control for different length wait and treatment periods and to consider dose-response effects, and c) to compare the clients who acted as their own wait list control with those who did not to check if the findings from the wait list control subset were applicable to the rest of the sample.

The present author was unaware of any comparable 'well controlled' naturalistic outcomes study of person-centred psychotherapy and this was one of the things that made this research unique.

Providers of psychotherapy want to know that the research reports used by reviewers to make judgements about the effectiveness of 'counselling' used correctly specified forms of the treatment cf. Bryant et al. (1998), Bryant et al. (1999). There was a need to research the effectiveness of person-centred psychotherapy provided by therapists trained in the approach. In particular a large part of the 'therapy work-force' in the UK were trained in person-centred psychotherapy. There was a need for this research to help consider whether these people should re-train to offer CBT, perhaps as part of IAPT.

There was a need for this research so that person-centred psychotherapy could participate in the evolution of research methodology. The literature review showed that many of the early research efforts that established the approach used techniques, methodologies and measures that contemporary researchers would not now recognise as 'acceptable' nor pass peer-review. Research methodologies evolve and it is not possible to stand still with the sense of 'all the research that was needed is now done' because research techniques are continually being refined. There was a need for this research so that the present author and hopefully others could be encouraged to participate in the evolution of person-centred psychotherapy outcomes research.

b) Process-outcomes research

In addition to outcomes evidence researchers and reviewers want to see a rationale for the treatment effect and that ideally this rationale/process effect is evidence-based such that there is a ‘convincing causal explanation of the process’ (Elliott, 2010, p.123).

The literature review showed that the early tests of Rogers hypotheses were largely unsuccessful (e.g. Rogers, 1967) and where there was any apparent ‘success’ these findings were largely rejected by subsequent researchers/reviewers e.g. Gurman (1977), Lambert, et al. (1978), Watson (1984). Some subsequent research has established process-outcome correlations for some of the elements of Rogers’ theory e.g. empathy, see Elliott, et al. (2010). Watson made the case that a thorough test of Rogers’ theory should test all the relationship elements simultaneously and for ‘congruent’ as well as ‘incongruent’ clients. Some subsequent research has established a causative effect for some of the relationship elements e.g. Zuroff and Blatt (2006). However, Zuroff and Blatt tested a factor composed of three of the four relationship elements in non-person-centred therapies for depressed and therefore presumably ‘incongruent’ clients.

There was a need for this research as a simultaneous test of all of the Rogerian relationship elements with both ‘congruent’ and ‘incongruent’ clients. There was a need for this research to see if any observed effect of person-centred psychotherapy had anything to do with the therapeutic relationship.

The present author was unaware of any comparable methodologically balanced study of the impact of the therapeutic relationship as defined by Rogers on depression, anxiety or distress outcomes as a consequence of person-centred psychotherapy and this was one of the things that made this research unique.

The impact of the therapeutic relationship is foundational to the person-centred approach. Whilst training the present author made the observation that ‘theory was something for the therapist to believe in when the going got tough’. If therapists are to ‘believe’ in their theory the theory must be supported by evidence. This research was needed for the present author, and perhaps for some other person-centred psychotherapists, to support the theory and in so doing to support clients ‘when the going gets tough’.

The research review showed that there have been lots of studies with non-person-centred therapies that have found process-outcome correlations with a number of relationship

elements, including the working alliance. As previously described person-centred therapy would do well to participate in the evolution of research methodologies and this research was needed to hopefully encourage other person-centred researchers to do further change-process research. In addition, this research was needed to identify how to/how not to do research of this kind in practice, perhaps acting as a pilot for further work in this area.

The literature review found many recent journal articles from the early part of the twenty-first century whose authors were engaged in highly sophisticated methodological and statistical techniques that were shedding new light on previously un-examined areas, some of them predicted by person-centred theory/practice, mainly not carried out by researchers from the person-centred approach. Clearly there is a risk that person-centred psychotherapy will get left behind if its proponents don't know how to do the research. This research was needed to hopefully inspire others to become involved in these developing areas.

Some of the insights being developed, for example the potentially causative nature of the therapeutic alliance for clients with disordered personality processes (Strauss et al., 2006) led to at least one of the authors to focus on 'identifying therapist strategies associated with better and worse early alliances and rupture outcomes to improve treatment retention and treatment outcomes in this prevalent and challenging population' (p. 344). This looks like something person-centred therapists should know about and could make a contribution to, but not if they are not participating in quantitative research. Again, this research was needed to encourage other person-centred therapists into the field of quantitative research, where given the encouraging research evidence for 'relationship elements' (Norcross, 2010) there would seem to be quite some contribution that could be made to the field.

In addition to the need for this research to consider the effect of potential moderators on outcomes (e.g. dose-effect relationships, medication status, disordered personality process, etc.) this research needed to consider the effects of these on putative process-outcome correlations for the therapeutic relationship with depression, anxiety and distress outcomes.

Gurman (1977) stated that Lesser's (1961) study was the only study, in his review of the impact of Rogerian conditions on outcome, that controlled for the pre-test variable and whilst this research used an appropriate methodology it was a null finding. The review by the present author identified that Lesser's study may have been subject to what is now termed dose-effect (Barkham, et al., 2006) and that this potentially confounding variable could have overlapped with the experimental effect, potentially reducing the statistical size of the effect

that would otherwise have been attributable to the experimental effect (Field, 2009, pp. 397-399). Considering issues such as these made the present author consider that it might be possible to identify a statistically significant effect for the process-outcome correlation where previously this had not been identified. This was especially encouraged by the findings in the literature review e.g. alliance-outcome correlations (Horvath, et al., 2010), putative impact of client-variables on ratings of therapists (Baldwin, et al., 2007, Crits-Christoph, et al., 2009), etc.

In addition to considering the effect of some other variables this research needed to consider the impact of outlier and influential cases in the process-outcome correlations. Again this was an encouragement that a statistically significant finding could be made where none had previously been found.

This research was needed to consider whether the effect of therapy on outcome had anything to do with the therapeutic relationship as defined by Rogers, whilst considering the effect of other potentially moderating variables and extreme cases and this was one of the things that made this research unique.

4.2 Overview of hypotheses to be tested

In summary, the purpose of this research was to examine the clinical effectiveness of person-centred psychotherapy as offered at the University of East Anglia (UEA) University Counselling Service (UCS) by a group of therapists (including the author) and by the author at the author's private practice (PP). In addition to examining the clinical effectiveness, evidence was also sought as to the role of the therapeutic relationship in predicting outcome. Rogers' theory (1957, 1959) was that the therapeutic relationship as defined by client perception of the therapist's provision of congruent empathy and unconditional positive regard was responsible for outcome.

This study was primarily designed to look at outcomes by comparing pre-therapy and post-therapy responses to self-completion questionnaires and to look at the impact of the therapeutic relationship as scored by the client as a putative predictor of outcome.

In general terms the hypotheses were that:

- A. Outcomes - Comparing where a client started from, to where a client was at the end of therapy; on average clients would improve such that measures of depression, anxiety and distress would show statistically significant improvements ($p < .05$).
- B. Process – Any observed change in depression, anxiety or distress symptoms, as measured, would be at least partially predicted by the therapeutic relationship, as measured, to a statistically significant extent ($p < .05$).

Beyond these hypotheses further analyses were required to seek to control for other variables in both the outcomes and process-outcomes parts of the study.

Once appropriate psychological measures were identified for the constructs to be tested the precise hypotheses were specified in terms of the psychological instruments to be used and their scoring protocols. The method section gives further information about each of the psychological instruments used (their scoring protocol, reliabilities, clinical cut-offs, etc.) and the rationale for the nature of the precise hypotheses (clinical cut-off scores). However, in

order to state in precise terms the hypotheses being tested in this section, in summary the psychological instruments used were as follows:

- Beck Depression Inventory (BDI-II) a client completed standard measure of depression (Beck, et al., 1996).
- Beck Anxiety Inventory (BAI) a client completed standard measure of anxiety (Beck & Steer, 1993).
- Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) a client completed measure of subjective distress widely used in psychological therapy services, including within the NHS (Barkham, et al., 2006).
- Personal Beliefs Questionnaire (PBQ) a client completed measure of beliefs associated with disordered personality process (Beck, et al., 2004) thought to have some predictive ability with diagnostic status (Beck, et al., 2001).
- Barrett-Lennard Relationship Inventory (BLRI) a client completed measure of the therapeutic relationship (Barrett-Lennard, 1962) as defined by Rogers (1957) and used by Rogers at Wisconsin (1967).

The outcome hypotheses were chosen because, as described above, NICE has produced clinical guidelines for depression and anxiety and CORE-OM distress is a widely used psychotherapy outcome measure. The predictor hypotheses were chosen because, as described above, these are related to the theory of the person-centred approach.

The precise hypotheses to be tested were defined as follows in the next section.

4.3 Specific hypotheses to be tested

A. Outcome Hypotheses

A.1 Depression Outcomes

H_0 Null hypothesis – mean client BDI-II scores do not improve (amongst clients who start therapy with a clinical level of depression ($BDI-II \geq 14$) and have a subsequent measurement of their depression).

H_1 Experimental hypothesis – mean client BDI-II scores improve, $p < .05$ (amongst clients who start therapy with a clinical level of depression ($BDI-II \geq 14$) and have a subsequent measurement of their depression).

A.2 Anxiety Outcomes

H_0 Null hypothesis – mean client BAI scores do not improve (amongst clients who start therapy with a clinical level of anxiety ($BAI \geq 8$) and have a subsequent measurement of their anxiety).

H_1 Experimental hypothesis – mean client BAI scores improve, $p < .05$ (amongst clients who start therapy with a clinical level of anxiety ($BAI \geq 8$) and have a subsequent measurement of their anxiety).

A.3 Distress Outcomes

H_0 Null hypothesis – mean client clinical CORE-OM scores do not improve (amongst clients who start therapy with a clinical level of distress (CORE-OM score ≥ 10) and have a subsequent measurement of their distress).

H_1 Experimental hypothesis – mean client CORE-OM scores improve, $p < .05$ (amongst clients who start therapy with a clinical level of distress (CORE-OM score ≥ 10) and have a subsequent measurement of their distress).

B. Predictor Hypotheses

B.1 Prediction of Depression Outcomes

H_0 Null hypothesis – BLRI scores do not predict subsequent BDI-II scores at a statistically significant ($p < .05$) level whilst controlling for start BDI-II scores.

H_1 Experimental hypothesis – BLRI scores predict subsequent BDI-II scores at a statistically significant ($p < .05$) level whilst controlling for start BDI-II scores.

B.2 Prediction of Anxiety Outcomes

H_0 Null hypothesis – BLRI scores do not predict subsequent BAI scores at a statistically significant ($p < .05$) level whilst controlling for start BAI scores.

H_1 Experimental hypothesis – BLRI scores predict subsequent BAI scores at a statistically significant ($p < .05$) level whilst controlling for start BAI scores.

B.3 Prediction of Distress Outcomes

H_0 Null hypothesis – BLRI scores do not predict subsequent CORE-OM scores at a statistically significant ($p < .05$) level whilst controlling for start CORE-OM scores.

H_1 Experimental hypothesis – BLRI scores predict subsequent CORE-OM scores at a statistically significant ($p < .05$) level whilst controlling for start CORE-OM scores.

Beyond these formal statements of the hypotheses further analyses were required to control for the effects of other variables on both outcomes and process-outcomes and to place the hypothesis testing in an appropriate context.

4.4 Concluding comments on the introduction to the research.

This section introduced the research that follows. The case was made for this research within the context of the literature review in the preceding section. The need for this research was expressed in terms of both the outcomes and process-outcomes parts of the research. Beyond, a need to perhaps re-orientate person-centred psychotherapy research in the direction of more quantitative research that the present author perceived, at summary level this research was needed to:

1. Add to the evidence-base for outcomes as a consequence of person-centred psychotherapy, especially in terms of distress (CORE-OM), to put this on a comparable base with other psychotherapy outcome studies and through the use of diagnostic-specific measures for depression (BDI-II) and anxiety (BAI).
2. Conduct this naturalistic outcomes research in such a way as to make this a 'well controlled' study in its context by a thorough examination of rival hypotheses. This was one of the things that made this research unique to the knowledge of the present author.
3. Examine whether there was statistically significant prediction of outcomes, whilst controlling for the pre-test measure, from the prior measurement of all four of the Rogerian 'therapeutic conditions' with both congruent and incongruent clients, this was one of the things that made this research unique to the knowledge of the present author.
4. Conduct this naturalistic process-outcomes research on the impact of the therapeutic relationship as defined by Rogers, in such a way as to make this a 'well controlled' study in its context by a thorough examination of some potentially confounding variables and extreme cases, this was one of the things that made this research unique to the knowledge of the present author.

Given the formal hypothesis testing structure to this research, items 1 and 3 in the above list would be reported as 'results' (the formal hypotheses) and items 2 and 4 (the control aspects of the analyses) as 'further results'. The outcome results (1) would thus be put into context by

the consideration of rival hypotheses (2) and the process-outcome results (3) put into context by the consideration of potentially confounding variables and extreme cases (4).

The next section describes the Method, this section is followed by the Results section and the Further Results section. Finally a Discussion is provided.

5. Method

In general terms the method of examining the clinical effectiveness of the person-centred approach and the impact of the therapeutic relationship was to look at client self-reported changes using validated outcome questionnaires, comparing pre-therapy with post-therapy scores, and to look at the client's score of the therapeutic relationship, using a validated self-completion questionnaire, as a putative predictor of outcome.

In order to test the hypotheses and rival hypotheses the research literature was consulted, in particular the methodologies literature, which as well as those items mentioned above also included a range of different studies, not referenced here, plus some key methodological texts, for example Cook and Campbell (1979), House (1980), Snow and Wiley (1991), Cambell and Russo (1999), Bickman (2000), and Mcleod (2003). In addition the following professional guidelines about research reporting and conduct were consulted: American Psychological Association (2001), British Association for Counselling and Psychotherapy (BACP, 2004a) and BACP (2004b).

The design of the method evolved, this evolution is described below together with the rationale for the evolution and the details of the questionnaires used. The research started by using a general measure of clinical distress (CORE-OM) as an outcome measure, then added in firstly a measure of depression outcome (BDI-II) and secondly a measure of anxiety outcome (BAI). When the research was extended to the UCS site some clients waited for therapy and the length of the wait was measured, together with changes in the outcome measures, so that clients could act as their own control. Clients were also asked about concurrent medications so that some attempt could be made to control for these. A process measure was added, the Barrett-Lennard Relationship Inventory, a measure of the therapeutic relationship as defined by Rogers. Subsequently the outcome measures were rationalised such that depression (BDI-II) and anxiety (BAI) outcomes were measured, together with the therapeutic relationship. The final change to the method was to add in a measure of personality disorder (PBQ) as a measure of prevalence of personality disorders amongst the population being assessed for depression and anxiety outcomes, since presence of a co-morbid personality disorder was known to impact outcomes (Clarkin & Levy, 2004).

5.1 Participants

5.1.1 Clients

The client sample was of 321 clients, 137 clients from the author's Private Practice (PP) and 184 clients from the University Counselling Service at the University of East Anglia (UCS). Clients were invited to take part in this naturalistic research at their first meeting with a therapist. PP clients were invited to take part in the research at their first session with the author and UCS clients were invited at their exploratory session with a qualified therapist. At the UCS all clients were offered an exploratory session as part of the usual way of working, this was an opportunity for the client to ask any questions they may have about counselling and for the qualified therapist to assign a priority (urgent, not urgent) to the client and judge whether that client might be suitable for a trainee (suitable for trainee or experienced counsellor required). All therapists working at the UCS, either as pre-qualification students or as post-qualification staff, were invited to participate. Only post-qualification staff conducted exploratory sessions. Not all of the UCS therapists doing exploratory sessions were participating in the research, only those who were participating invited clients to take part in the research. There were no formal protocols to support decision-making for the judgements about priority and suitability for trainee.

This was intended to be a naturalistic study of bona fide clients, being seen for therapy, as opposed to a 'laboratory study' so there were no inclusion or exclusion criteria; all clients attending for their initial session were simply invited to take part in the research. For the analyses certain inclusion and exclusion criteria were used (e.g. depression outcomes for those clients starting with a clinical level of depression, etc.) and these are specified below (sections 6 and 7).

The study was started with the intention that the methodology evolve during the study and no formal sample size calculation was conducted at the start.

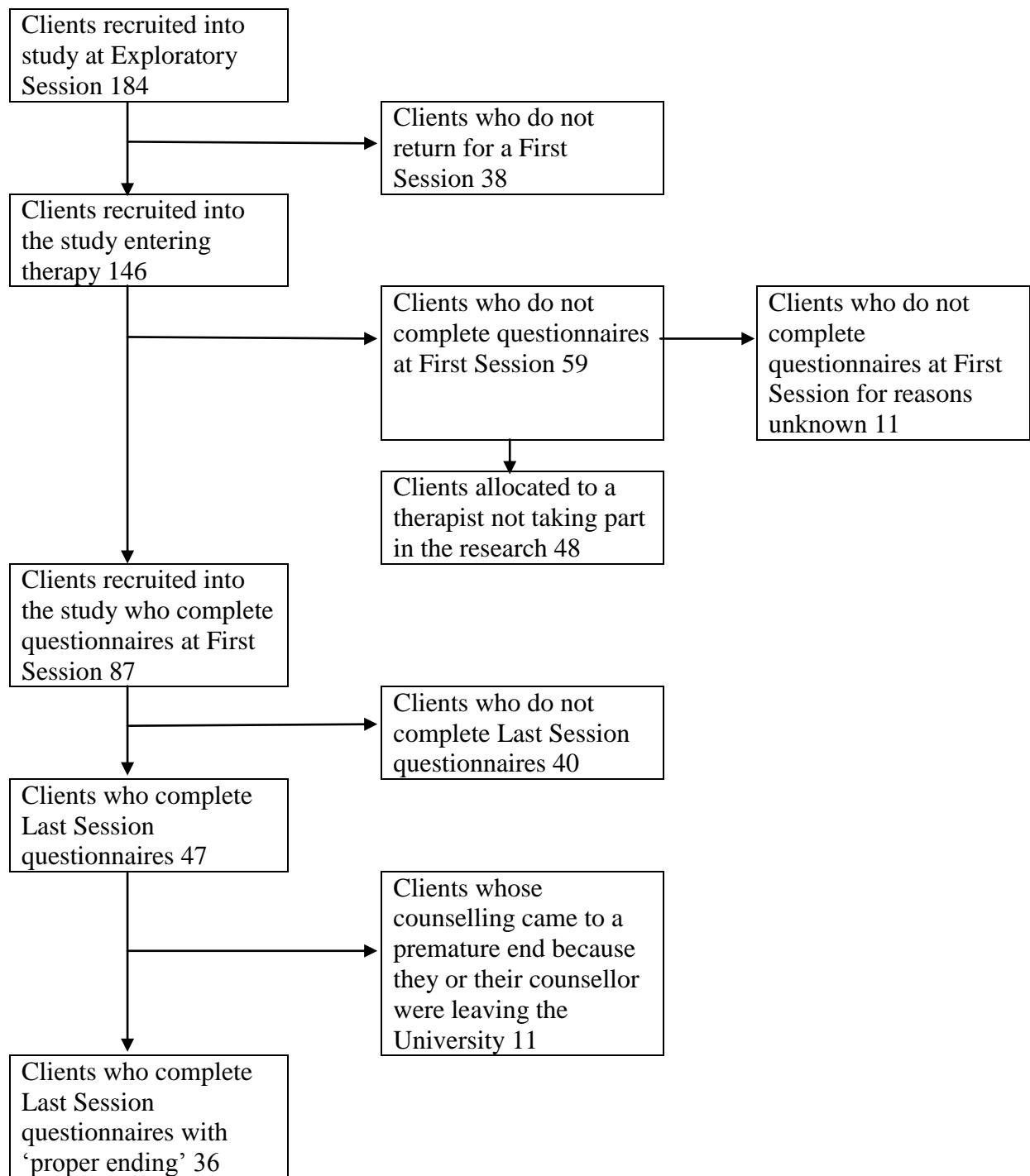
This research was approved by the University of East Anglia Research Ethics Committee (Appendix 1: Research Ethics Committee submission), as was the preceding Masters Research. Appendix 1 shows the detail of the recruitment and information/consent procedures, together with the handouts used by clients and therapists to opt into the research.

As part of the Ethics Committee review and discussion with the UCS therapists the methodology changed slightly from that described in the original Ethics Committee submission and the methodology was as reported in this document, rather than in the appendix.

a) University Counselling Service (UCS)

A flow diagram for participation at each stage of the research at the UCS is provided (Figure 1). Of the 184 clients at the UCS who opted into the research at the exploratory session, 38 did not return for a first counselling session and so 146 UCS clients entered into therapy. Of these 146 clients, 59 did not complete the first session paperwork. Of the 59 clients who did not complete first session paperwork, 48 of these were allocated to a therapist not taking part in the research, as part of the usual way of working at the UCS; A further 11 clients did not complete the first session paperwork for reasons unknown. It was possible that some clients may have decided not to continue in the research, or their therapist decided not to ask for the paperwork to be completed, i.e. therapist subsequently opted out of actively doing the research. Therefore 87 clients 'in the research' (146 minus 59) began counselling at the UCS, of these 47 completed all three outcome questionnaires at their last counselling session (a 48th client completed just one outcome questionnaire at the last session). The 40 clients in the research who did not complete all of the last session paperwork may have had satisfactory endings, although because the paperwork was not completed it is not possible to know definitively if this was the case. Some of the client record cards at the UCS suggested at least some of these 40 clients had what the therapist described as 'mutually agreed and satisfactory' endings. The 47 clients who completed all of the last session paperwork included at least 11 clients whose counselling came to a premature ending because they or their counsellor were leaving the University at the end of term, thus circa 36 clients had 'proper endings' and completed the outcome measures at the various stages. Note that 'types of ending' were recorded by therapists and may not accord with client perspectives on this.

Figure 1: For University Counselling Service clients, flow diagram of participation at each stage of research.



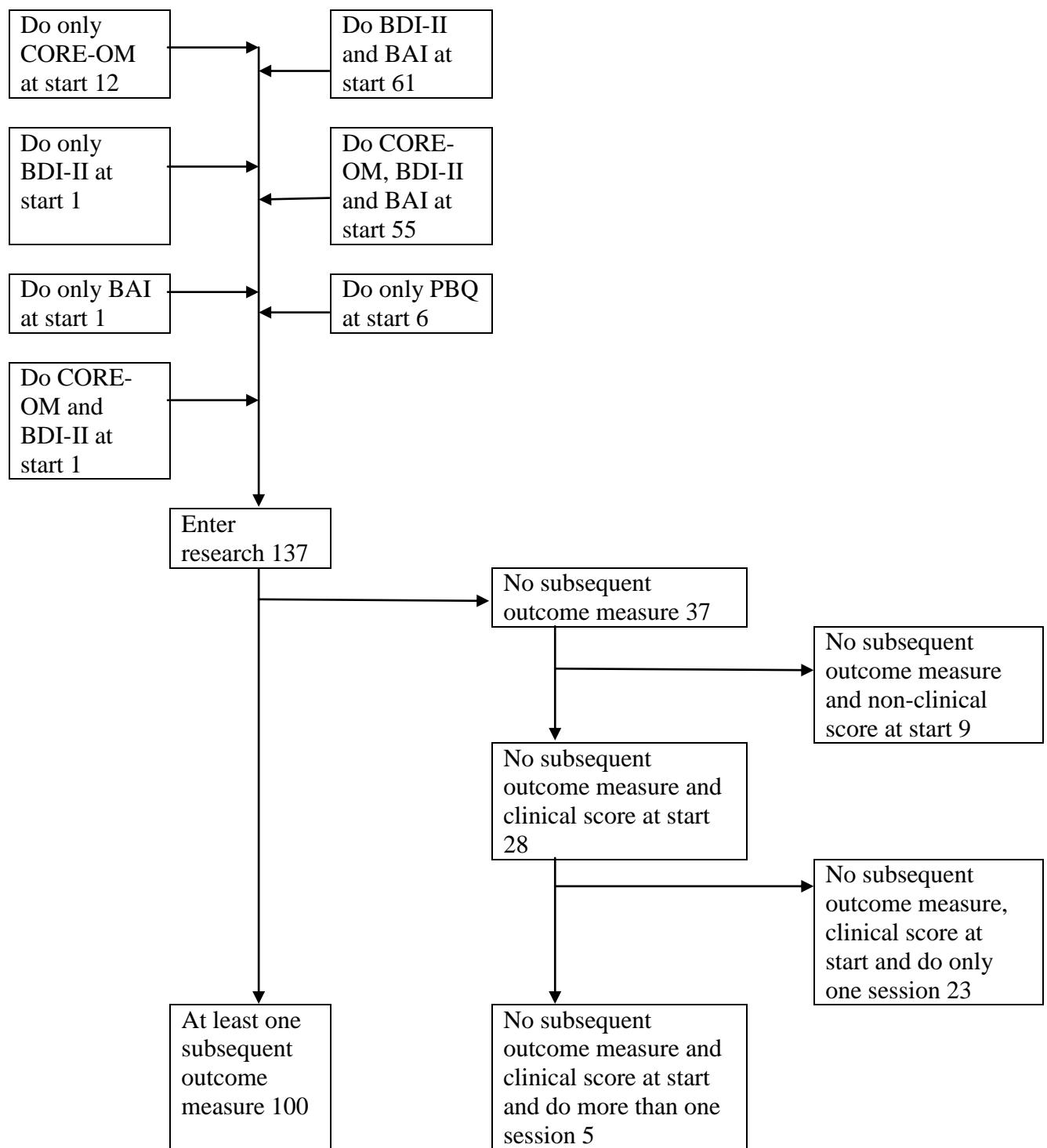
b) Private Practice Clients (PP)

A flow diagram is provided for participation in the research at the Private Practice (Figure 2). Of the 137 clients at the PP who opt into the research at the start of therapy the following numbers of clients completed the following outcome questionnaires at the start of therapy:

- 12 completed only a CORE-OM questionnaire at the start of therapy, all completed a subsequent CORE-OM questionnaire
- 1 completed only a BDI-II questionnaire at the start and completed a subsequent BDI-II questionnaire
- 1 completed only a BAI questionnaire at the start and completed a subsequent BAI questionnaire
- 1 completed both a CORE-OM and BDI-II questionnaire at the start and subsequently re-completed both questionnaires
- 61 completed both BDI-II and BAI questionnaire at the start, subsequently:
 - 22 completed both BDI-II and BAI
 - 13 completed BDI-II only (this was partly because of 'measure depression before anxiety' (Newman, et al., 2006))
 - 3 completed BAI only
 - 23 did not complete either a BDI-II or a BAI:
- 55 completed CORE-OM, BDI-II and BAI at the start, subsequently:
 - 22 completed CORE-OM, BDI-II and BAI
 - 1 completed CORE-OM only
 - 5 completed BDI-II only (this was partly because of 'measure depression before anxiety' (Newman, et al., 2006))
 - 1 completed BAI only
 - 6 completed CORE-OM and BDI-II
 - 6 completed BDI-II and BAI
 - 14 did not complete either CORE-OM, BDI-II or BAI
- 6 completed only PBQ at the start

Of the 137 clients who started in the research there were 100 who completed at least one subsequent outcome measure and 37 who had no subsequent outcome measurement. Of these 37 clients with no subsequent outcome measure there were 9 clients who had non-clinical scores on each measure they completed at the start and 28 clients who started with at least one clinical score and had no subsequent outcome measurement; 23 clients did only one session. In total there were five clients who started with a clinical score on at least one outcome measure and who did more than one session and had no outcome measurement; one client did two sessions and four did three sessions each.

Figure 2: For Private Practice clients, flow diagram of participation at each stage of research.



5.1.2 Therapists

All therapists had received some formal training as a person-centred practitioner. Therapists were invited to sign up to the research and complete some demographic information about themselves, they also took a unique and confidential therapist reference number with which to mark the questionnaires of clients they had seen. Whilst 27 therapists signed up to do the research not all of the client questionnaires had therapist reference numbers on them, so it was not possible to know whether there were more than 27 therapists who saw clients. For the 321 clients who opted into the research 38 clients were not allocated a therapist because they did not attend a first session and for 49 clients it was not known who the therapist was. However, all BLRI forms and subsequent outcome measures had therapist reference numbers on them. So whilst it was not known who the therapists were for all of the clients, where it was important to identify the therapist, this information was available.

Clients were seen by 27+ therapists, although 54.3% of the 283 clients who were allocated a therapist were seen by one therapist (the author), another therapist saw 7% of clients, another 4% and the rest saw 1-5 clients each; eleven therapists saw only one client each. At the UCS 27+ therapists saw clients 'in the research', although only 18 therapists had clients complete outcome questionnaires at their last session. Of these 18 therapists, 12 were studying for a diploma in person-centred counselling and 6 were post qualification, with an average of 7 years post qualification experience (range for post-qualification experience 2-20 years). All but two of these 18 therapists completed information about themselves, so some demographic information is available for 16 therapists. The two therapists who did not complete demographic information about themselves each saw one client through to their final session and were both trainee therapists. For the 16 who provided demographic information about themselves, 12 of the therapists were female. At the start of the research the average age of the therapists was 43 years (range 28-63 years); 10 were married (5 single, 1 separated/divorced) and 6 had at least one child. Most therapists were British (12) with 4 from Europe, Asia or America.

There was no formal check on adherence to treatment approach and no treatment manual; however, given this was a person-centred counselling service, partly staffed by students on placement from a person-centred diploma course, the UCS expected that person-centred therapy would be offered to clients, as defined by Rogers (1957, 1959). To some extent the BLRI was a check on adherence since Rogers' theory was that the therapist provide

congruent empathy and unconditional positive regard and the BLRI was a measure of the client's perception of these 'conditions'.

5.2 Measures

5.2.1 Outcome measures

a) Depression (BDI-II)

A standard measure of depression was used (Beck et al 1996) that has been widely used in research (Minami, et al., 2007). Clients respond to 21 questions by circling one of four response options for each question, each response carries a score of 0, 1, 2 or 3 points. The maximum possible score is 63 and the authors defined total scores as: 0-13 minimal depression ('non-depressed'), 14-19 'mild depression', 20-28 'moderate depression' and 29-63 'severe depression'. The authors reported a test-retest stability of .93 ($p < .001$) and an internal consistency of .92 (coefficient alpha). For the 303 clients who completed a BDI-II questionnaire at the first time of asking the coefficient alpha was found to be .900 for the 21 items of the BDI-II, this was slightly lower than that reported by the authors.

At the first time of completion for these 303 clients the mean BDI-II score was 22.41 (SD 10.68, range 0-56) with a distribution that was significantly different from normal, $K-S D(303) = .07$, $p = .002$. Using the method recommended in the literature to calculate reliable change (Jacobson & Truax, 1991) this suggested a Standard Error of measurement for the BDI-II questionnaire of 2.83 ($S_E = S_1\sqrt{1 - r_{test-retest}}$, where $S_1 = 10.68$ and $r_{test-retest} = .93$). The 'spread of the distribution of change scores that would be expected if no actual change had occurred' (p. 14) was given by S_{diff} and this was calculated from the Standard Error ($S_{diff} = \sqrt{2(S_E)^2}$, where $S_E = 2.83$) to give a value of 4.00. For 95% confidence of reliable change 4.00 was multiplied by 1.96, the z-value corresponding to $p < .05$, ($S_{diff} \times 1.96$, where $S_{diff} = 4.00$) to give a value of 7.83; being the number of BDI-II points change required for 95% confidence that reliable change had occurred in a client's pre and post BDI-II scores. Table 2 shows the relevant data and calculations for all measures.

Whilst the distribution of BDI-II scores at the first time of completion was significantly different from normally distributed the decision was made not to transform this data for the main parts of the analysis. This was to retain the ability to compare outcomes from this study with other studies, the benchmarking objective, and also to retain simplicity. It was found that the BDI-II scores at the first time of completion could be transformed to a distribution not significantly different from normal by adding four to each score (there were some zero values) and taking the square root of the resultant sum. This gave a distribution not significantly different from normal, $K-S D(303) = .05$, $p = .083$. The transformed data had a mean of 5.03 (SD 1.07). Working this through gave a reliable change index of .786 for the transformed data. If $\sqrt{(BDI-II \text{ start} + 4)} - \sqrt{(BDI-II \text{ end} + 4)} > .786$ then reliable change had occurred, e.g. If BDI-II at the start was 19 and subsequently 12 this was a reliable change at better than 95% confidence because $\sqrt{(19+4)} - \sqrt{(12+4)} = .796$ which is $> .786$. This was not very convenient for clinical use and for the reasons stated the data was not transformed and this point was borne in mind for the following analyses.

Table 2: Calculation of reliable change criteria (Reliable Change Index, RCI) for 95% confidence in each outcome measure: Data from this research.

Measure	n	r ^a	Mean of first measure	$S_1 = \text{Standard Deviation of first measure}$	$\text{Std Error } S_E = S_1 \sqrt{1-r}$	$S_{\text{diff}} = \sqrt{2(S_E)^2}$	$RCI = S_{\text{diff}} \times 1.96$	RCI/Mean (%)
BDI-II	303	.93	22.41	10.68	2.83	4.00	7.83	34.9%
BDI-II transformed	303	.93	5.03	1.07	0.28	0.40	0.79	15.6%
BAI	301	.75	16.61	10.14	5.07	7.17	14.06	84.6%
BAI transformed	301	.75	4.02	1.22	0.61	0.86	1.69	42.2%
CORE-OM	251	.90	17.02	6.10	1.93	2.73	5.34	31.4%

Note: ^a Test-retest stability. ^b BAI subscale. ^c CORE-OM subscale.

b) Anxiety (BAI)

The Beck Anxiety Inventory (BAI) (Beck & Steer, 1993) is a standard measure of anxiety that has been widely used in research (NICE, 2004a) and with a similar scoring pattern to BDI-II. The maximum possible score is 63 and the authors defined total scores as: 0-7

minimal anxiety ('non-anxious'), 8-15 'mild anxiety', 16-25 'moderate anxiety' and 26-63 'severe anxiety'. The authors suggested the scores for women may be an average of 4 points higher than for men (p. 5) although this author used the bandings suggested by the authors as above throughout this work.

The authors reported a coefficient alpha of .92 and a test-retest stability of .75 ($p < .001$). For the 301 clients who completed the BAI on at least one occasion for this research, coefficient alpha was .894, lower than that reported by the authors.

At the first time of completion the mean BAI score was 16.61 (SD 10.141, range 0-54) with a distribution that was significantly different from normal, $K-S D(301) = .10$, $p < .001$. Using the method described above (Jacobson & Truax, 1991) this suggested a Standard Error of Measurement for the BAI questionnaire of 5.07 ($S_E = S_1\sqrt{1 - r_{test-retest}}$), where $S_1 = 10.14$ and $r_{test-retest} = .75$), S_{diff} was calculated from the Standard Error ($S_{diff} = \sqrt{2(S_E)^2}$), where $S_E = 5.07$) to give 7.17. For 95% confidence of reliable change 7.17 was multiplied by 1.96, ($S_{diff} \times 1.96$, where $S_{diff} = 7.17$) to give a value of 14.06; the number of BAI points change required for 95% confidence that reliable change had occurred in a client's pre and post BAI scores (Table 2).

Whilst the distribution of BAI scores at the first time of completion was significantly different from normally distributed the decision was made not to transform this data for the main analysis, for the reasons given above. It was found that the BAI scores at the first time of completion could be transformed to a normal distribution by adding one to each score (there were some zero values) and taking the square root of the resultant sum. This gave a distribution not significantly different from normal, $K-S D(301) = .05$, $p > .2$. The transformed data had a mean of 4.02 (SD 1.22). Working this through gave a reliable change index of 1.69 for the transformed data. If $\sqrt{(BAI \text{ start} + 1)} - \sqrt{(BAI \text{ end} + 1)} > 1.69$ then reliable change had occurred. This was not very convenient for clinical use and for the reasons stated the data was not transformed and this point was borne in mind for the following analyses.

c) Distress (CORE-OM)

Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM, Barkham, et al., 2006) is a measure of distress used widely within NHS settings (Stiles et al. 2006, 2007). Clients respond to 34 questions by ticking option boxes, each response carrying a score of 0, 1, 2, 3 or 4. The maximum possible score is 136. The authors defined what they term 'clinical scores' by taking the total score for the questionnaire, dividing this by 34 to get an item mean score and multiplying this mean score by 10 to get what the questionnaire's authors referred to as the 'clinical score'. The authors defined what they term 'clinical scores' as falling into the following bands: 0-6 'healthy', 6.2-9.7 'low' (scores of 9.7 and below are considered 'non-clinical'), 10.0-14.7 'mild', 15.0-19.7 'moderate', 20.0-24.7 'moderate-to-severe' and 25.0-40.0 'severe' (Barkham, et al., 2006).

Developers of the questionnaire reported a coefficient alpha of .94 and a test-retest stability of .90 (Barkham, et al., 2001). The 251 clients who completed CORE-OM in this research had a coefficient alpha of .93 at the first time of completion. The mean CORE-OM score was 17.02 (SD 6.10, range 1.47-33.53) with a distribution not significantly different from normal K-S D(251) = .04, $p > .2$. Using the method described above (Jacobson & Truax, 1991) gave a Standard Error for the CORE-OM of 1.93 ($S_E = S_1\sqrt{1 - r_{\alpha}}$, where $S_1 = 6.10$ and $r_{\text{test-retest}} = .90$), S_{diff} was calculated from the Standard Error ($S_{\text{diff}} = \sqrt{2(S_E)^2}$, where $S_E = 1.93$) to give 2.73. For 95% confidence of reliable change 2.73 was multiplied by 1.96, ($S_{\text{diff}} \times 1.96$, where $S_{\text{diff}} = 2.73$) to give a value of 5.34; the number of CORE-OM points change required for 95% confidence that reliable change had occurred in a client's pre and post CORE-OM scores (Table 2). This compared with the questionnaire's authors (Barkham, et al., 2006) finding that changes greater than 5 in the clinical score provide evidence of reliable change.

d) Personality Disorder (PBQ)

Whilst not strictly used as an outcome measure. The Personality Beliefs Questionnaire (PBQ, Beck, et al., 2001) was used to assess the prevalence of disordered personality processes amongst clients. The PBQ was developed as a self report questionnaire to discern probable presence of a personality disorder based on what the authors described as 'dysfunctional beliefs' and as a possible outcome measure for therapy with clients with personality disorder(s). It is not unusual to seek to determine presence of a personality disorder based on

client's beliefs (Arntz, Dreesen, Schouten, & Weertman, 2004). There are eleven personality disorders, usually considered in the following three clusters (American Psychiatric Association, 2000):

- Cluster A: Paranoid, Schizoid, Schizotypal.
- Cluster B: Anti-social, Narcissistic, Histrionic, Borderline.
- Cluster C: Avoidant, Dependent, Passive-Aggressive, Obsessive-Compulsive.

The PBQ was designed to test for presence of ten personality disorders, all but 'schizotypal'. The PBQ consists of an equal number of items (14) representing beliefs thought to be associated with avoidant, dependent, passive-aggressive, obsessive-compulsive, anti-social, narcissistic, histrionic, schizoid and paranoid personality disorders. Together this makes up 126 items ($14 \times 9 = 126$) and together with the nine personality disorder subscales a tenth subscale, thought to correspond to a borderline personality subscale is composed of 14 items from the PBQ avoidant, dependent, histrionic and paranoid domains representing themes of dependency, helplessness, distrust, fears of rejection/abandonment/losing emotional control and extreme attention-seeking behaviour (Butler, Brown, Beck, & Grisham, 2002). Each of the 126 'dysfunctional beliefs' is assessed by the client as 'I don't believe it at all' (score 0), 'I believe it slightly' (1), 'I believe it moderately' (2), 'I believe it very much' (3) and 'I believe it totally' (4). The maximum score on each subscale is thus 56 ($14 \times 4 = 56$).

Permission was sought and gratefully received from the lead author (A. T. Beck, personal communication 20th December 2007) to amend some of the item wording for a UK English-speaking client group.

The developers of the questionnaire (Beck, et al., 2001) provided mean z-scores for patients with a diagnosis of each personality disorder and instructed that the primary diagnosis was of the personality disorder with the highest z-score in the case of that a client had more than one z-score equivalent to that of the diagnosed patients.

The developers of the questionnaire reported a coefficient alpha of .89 amongst a group with mean score 18.8 (SD 10.9); a mean score of 25.6 (z-score .62) for clients with a diagnosis of avoidant personality disorder and a mean score of 11.3 (z-score -.69) for those without avoidant personality.

The developers of the questionnaire reported a coefficient alpha of .90 amongst a group with mean score 18.0 (SD 11.8); a mean score of 27.8 (z-score .83) for clients with a diagnosis of

dependent personality disorder and a mean score of 12.2 (z-score -.49) for those without dependent personality.

The developers of the questionnaire reported a coefficient alpha of .88 amongst a group with mean score 19.3 (SD 10.5) and a mean score of 15.3 (z-score -.38) for those without passive-aggressive personality.

Unfortunately the questionnaire's authors had insufficient client numbers with a diagnosis of passive-aggressive personality disorder to estimate a mean clinical score. In the absence of a mean clinical score the approach taken was that if a client had a clinical z-score on one of the other subscales and the PBQ-PA z-score was higher then this was counted as a passive-aggressive personality, in line with the authors instructions (Beck et al. 2001).

The developers of the questionnaire reported a coefficient alpha of .90 amongst a group with mean score 22.7 (SD 11.5); a mean score of 26.3 (z-score .31) for clients with a diagnosis of obsessive-compulsive personality disorder and a mean score of 16.8 (z-score -.51) for those without obsessive-compulsive personality.

The developers of the questionnaire reported a coefficient alpha of .81 amongst a group with mean score 9.3 (SD 6.8); a mean score of 11.4 (z-score .31) for clients with a diagnosis of anti-social personality disorder and a mean score of 8.1 (z-score -.18) for those without anti-social personality. Note the relatively small range of 3.3 points between the mean clinical and non-clinical populations.

The developers of the questionnaire reported a coefficient alpha of .84 amongst a group with mean score 10.0 (SD 7.6); a mean score of 18.4 (z-score 1.10) for clients with a diagnosis of narcissistic personality disorder and a mean score of 7.1 (z-score -.38) for those without narcissistic personality.

The developers of the questionnaire reported a coefficient alpha of .87 amongst a group with mean score 14.0 (SD 9.3) and a mean score of 11.3 (z-score -.29) for those without histrionic personality.

Unfortunately the questionnaire's authors had insufficient client numbers with a diagnosis of histrionic personality disorder to estimate a mean clinical score and the approach taken in this was research was that as described above for the passive-aggressive subscale.

The developers of the questionnaire reported a coefficient alpha of .81 amongst a group with a mean score of 16.3 (SD 8.6) and a mean score of 15.1 for those without schizoid personality.

Unfortunately the questionnaire's authors had insufficient client numbers with a diagnosis of schizoid personality disorder to estimate a mean clinical score and the approach taken in this research was that as described above for the passive-aggressive subscale.

The developers of the questionnaire reported a coefficient alpha of .93 amongst a group with mean score 14.6 (SD 11.3); a mean score of 20.4 (z-score .51) for clients with a diagnosis of paranoid personality disorder and a mean score of 8.4 (z-score -.55) for those without paranoid personality.

The developers of this subscale from the PBQ questionnaire reported a coefficient alpha of .89 amongst a group with mean score 15.8 (SD 10.5); a mean score of 23.9 (z-score .77) for clients with a diagnosis of borderline personality disorder and a mean score of 9.0 (z-score -.65) for those without borderline personality (Beck, et al., 2001, Butler, Brown, Beck, & Grisham, 2002).

5.2.2 Process measure

The Barrett-Lennard Relationship Inventory (BLRI) was developed by Barrett-Lennard (Barrett-Lennard, 1962) in conjunction with Carl Rogers and used at the Wisconsin project 'to measure the conditions of therapy as perceived by the individual' (Rogers, 1967, p. 32). Permission was granted by Barrett-Lennard for use of the BLRI in this study (G. T. Barrett-Lennard, personal communication, 20th January 2006) and his help was gratefully received in turning the inventory into a tick-box format (G. T. Barrett-Lennard, personal communication, 23rd March 2006), as illustrated (Figure 3) to make it quicker to complete.

Figure 3: Extract of Barrett-Lennard Relationship Inventory as used in this research: Tick box format as agreed with Barrett-Lennard.

Today's date:	Counsellor code:	Session: Pre-second	Client code:
---------------	------------------	---------------------	--------------

Barrett-Lennard Relationship Inventory - Form OS-40

© Godfrey T Barrett-Lennard, PhD

Below are listed a variety of ways that counsellors may feel or behave in relation to clients. Please consider each numbered statement with reference to your present relationship with your counsellor. Whilst you have only recently met, we are interested in your initial view of your counsellor. Mark each statement in the answer columns on the right, according to how strongly you feel that it is true, or not true, in this relationship. Please be sure to mark every one. Please indicate your answer by placing a tick in the box that most closely matches your response, using the following key:

-3: No(!), I strongly feel that it is not true,	+3: Yes(!), I strongly feel that it is true
-2: No, I feel it is not true	+2: Yes, I feel it is true
-1: (No), I feel that it is probably untrue, or more untrue than true	+1: (Yes), I feel that it is probably true or more true than untrue

DISAGREE			AGREE		
No	No	No	Yes	Yes	Yes
-3	-2	-1	+1	+2	+3
<input type="checkbox"/>					

1. My counsellor respects me...

The BLRI used in this study was the 40 item version with four subscales each of ten questions designed to probe a client on their perception of their therapist's regard, empathy, unconditionality and congruence with or for the client. Completion of the 40-item version at session five was recommended by Barrett-Lennard in the context of the proposed research (G. T. Barrett-Lennard, personal communication, 20th January 2006). The 40-item version (Barrett-Lennard, 1978) was developed as a more 'economical' version to make data collection more 'manageable' (Barrett-Lennard, 1998, p. 284) and has been used by other researchers to satisfactory results e.g. Goldman, Greenberg, and Angus (2006).

The scoring key of the BLRI meant that clients were forced to choose to agree or disagree with a statement about their therapist; there was no 'neither agree nor disagree' option. There were both positive and negatively worded response items. The scoring was designed such that aspects of what Rogers considered a 'good' relationship (Rogers, 1957) were scored with a maximum score of 6, or 5, or 4; aspects of a 'bad' relationship were scored with a minimum score of 0, or 1, or 2, such that there is no mid-score of 3 available. The maximum score a perceived therapeutic relationship could have was 240 (40 x 6 = 240) and the minimum score was 0 (40 x 0 = 0). A review of the research using the BLRI (Gurman, 1977) found the mean internal reliability coefficients across 14 studies to be .91 and the mean test-retest stability

across 10 studies to be .90, although it should be noted these studies were of different variants of the BLRI. For the 118 clients who completed a BLRI in this study the alpha coefficient was .88 with a mean score of 189.71 for the client's perception of the therapeutic relationship (SD 21.61, range 126-240), scores were not significantly different from a normal distribution $K-S(118) = .04, p > .2$. Initially the BLRI was completed at the start of the fifth session; however, this was changed to the end of the first session/start of the second session during the research, so as to increase the numbers of clients completing the BLRI before therapy ended and this reduced the time to establish a relationship, such that the BLRI was scored based upon 'initial impressions'. The mean session at which the BLRI was completed was 4.26 (SD 2.4, median session 5). The effect of earlier versus later BLRI completion was assessed using t-tests at different session cut-offs:

Session Three

There were 93 clients who completed their BLRI at the third session or later and 25 clients who completed their BLRI before the third session. The mean BLRI score for later completion was 192.16 (SD 20.8) and 180.60 (SD 22.6) for earlier completion. This difference was significant $t(116) = 2.42, p = .017$ and represented a small effect $r = .22$.

Session Four

There were 79 clients who completed their BLRI at the fourth session or later and 39 clients who completed their BLRI before the fourth session. The mean BLRI score for later completion was 192.53 (SD 20.3) and 184.00 (SD 23.3) for earlier completion. This difference was significant $t(116) = 2.04, p = .043$ and represented a small effect $r = .19$.

Session Five

There were 60 clients who completed their BLRI at the fifth session or later and 58 clients who completed their BLRI before the fifth session. The mean BLRI score for later completion was 189.78 (SD 20.8) and 189.64 (SD 22.6) for earlier completion. This difference was not significant $t(116) = .04, p = .971$.

It seemed that on average clients tended to score relationships lower in the first 1-3 sessions than they did in later sessions. This suggested there could be merit in measuring both 'initial impressions' of relationships (session 1-3) and more 'mature' relationships (session 4 onwards). These were likely measuring slightly different phenomena i.e. initial judgements

about whether to work with a particular therapist or judgements about what it was like to work with a particular therapist.

The client demographics for the samples upon which coefficient alpha was estimated for each questionnaire used are shown in Appendix 2.

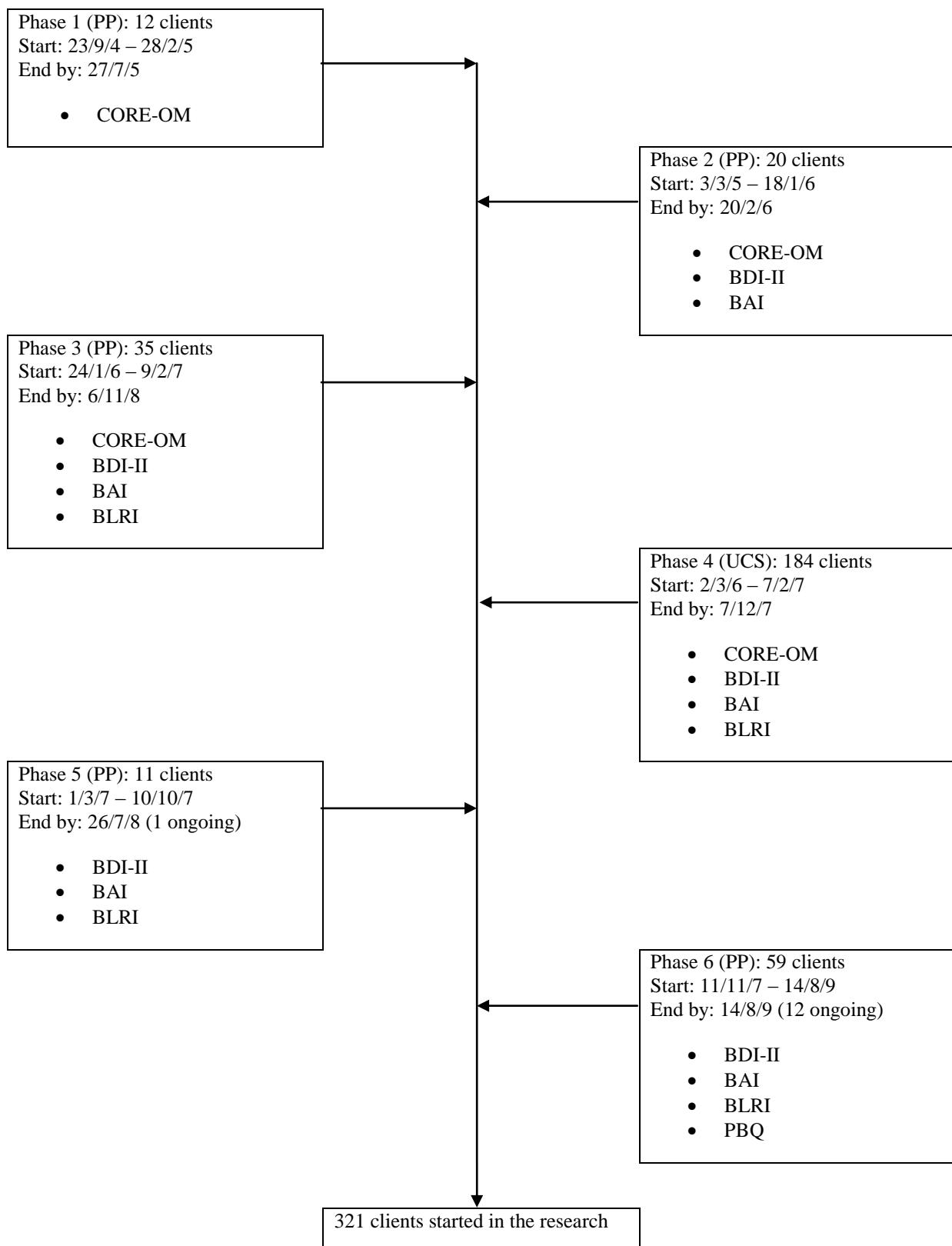
5.3 Research protocol

The precise methodology evolved through the course of the study. There were six distinct phases and these were when the data collection protocols were slightly different. These differences are described below in a description of the six different phases whose demarcations were changes to the data collection protocols. There were six distinct phases (Figure 4):

- Phase 1 – began during the author's Masters Research (Weston, 2005); in which pre-therapy and post-therapy distress scores (CORE-OM) were compared for a single therapist (the author). There were 12 clients in this phase that began in therapy between 23/9/4 and 28/2/5, the last of whom finished on 27/7/5. This author was concerned that CORE-OM may not be recognised as a diagnostic specific measure and so further diagnosis specific measures were sought; this concern was borne out in the NICE review of depression where NICE rejected CORE-OM as an indicator of depression (NICE, 2009a) although there was some published evidence to support the use of CORE-OM for depression diagnosis (Gilbody, et al., 2007).
- Phase 2 – in which pre-therapy and post-therapy scores were compared for a single therapist (the author) using up to three different outcome measures (CORE-OM, BDI-II and BAI). There were 20 clients in this phase that began in therapy between 3/3/5 and 18/1/6, the last of whom finished on 20/2/6. In addition to examining outcomes the author chose to look at the evidence for the role of the therapeutic relationship in predicting outcome and a relationship measure was sought.
- Phase 3 – as Phase 2 together with a measure of the therapeutic relationship (BLRI). There were 35 clients in this phase that began in therapy between 24/1/6 and 9/2/7, the last of whom finished on 6/11/8. In addition to researching outcomes and process for a single therapist there was an opportunity to widen the research to include therapists at the institution hosting the author's PhD.

- Phase 4 – in which pre-therapy and post therapy scores were compared using three outcome measures (CORE-OM, BDI-II and BAI) for a number of therapists (including the author) at the University of East Anglia University Counselling Service (UCS). This phase overlapped with phase 3. There were 184 clients who began in this phase between 2/3/6 and 7/2/7, the last of whom finished on 7/12/7. Experience to date had shown a certain amount of duplication between CORE-OM and the BDI-II/BAI combination, so the decision was made to reduce the paperwork burden on clients by dropping the CORE-OM for future phases.
- Phase 5 – at the Private Practice (PP) in which pre-therapy and post-therapy scores were compared for a single therapist using two outcome measures (BDI-II and BAI) together with an interim measure of the therapeutic relationship (BLRI). There were 11 clients who began in this phase between 1/3/7 and 10/10/7. All but one of these clients had completed by 26/7/8, with one client ongoing at the time of writing. One of these clients also completed a CORE-OM, in addition to BDI-II and BAI as an Employee Assistance Provider (EAP) requirement. The author was concerned that a typical RCT excludes clients with personality disorders and to date there had been no assessment of the prevalence of personality disorders within the sample. Four clients in this phase completed a PBQ after starting in therapy when this was introduced on 11th November 2007.
- Phase 6 – as Phase 5 together with a self-completion measure for personality disorders (PBQ). There were 59 clients who began in this phase between 11/11/7 and 14/8/9. Four of these clients also completed a CORE-OM, in addition to BDI-II and BAI, as an EAP requirement. During this phase it was decided to move BLRI completion forward to the end of the first session, so as to increase the numbers of clients who completed a BLRI. The BLRI was found to mediate outcome when completed after one session (Zuroff & Blatt, 2006). This was accomplished by giving the client a BLRI and a PBQ to take away and complete after the first session and post back in a pre-paid envelope. All but 12 of the clients who began in this phase had completed by 14/8/9, with 12 clients from this phase ongoing at the time of writing.

Personality disorder questionnaires were completed by a subset of the overall sample and data for personality disorders was included as a co-morbid personality disorder is known to impact outcomes (Clarkin & Levy, 2004).

Figure 4: Research protocol: The six phases of development of the research protocol.

Note: PP = Author's private practice; UCS = University Counselling Service; CORE-OM = Clinical Outcomes in Routine Evaluation, Outcome Measure; BDI-II = Beck Depression Inventory Second Edition; BAI = Beck Anxiety Inventory; BLRI = Barrett-Lennard Relationship Inventory; PBQ = Personal Beliefs Questionnaire (Beliefs associated with Personality Disorder).

5.4 Treatment duration

Treatment duration was variable, with an average of 7.1 sessions per client (SD 10.2, range 0-102), although this average included a large number of clients who did not start in counselling or continue in the research. The numbers of clients banded by numbers of sessions are shown (Table 3).

Table 3: For all clients in the research, numbers of sessions this episode.

Number of sessions this episode	Number of clients	Cumulative %
0 (Exploratory only)	38	11.8
1-6	180	67.9
7-12	58	86.0
13-18	21	92.5
19-24	10	95.6
25-36	7	97.8
37-54	4	99.1
55-102	3	100.0
Total	321	100.0

At the UCS clients were offered up to an initial six sessions, with the option to continue beyond this if both client and therapist agreed; the average number of sessions was 5.5 (SD 6.4, range 0-37). The numbers of UCS clients banded by number of sessions are shown (Table 4).

Table 4: For University Counselling Service clients, numbers of sessions this episode.

Number of sessions this episode	Number of clients	Cumulative %
0 (Exploratory only)	38	20.7
1-6	94	71.7
7-12	30	88.0
13-18	11	94.0
19-24	7	97.8
25-36	3	99.5
37-54	1	100.0
55-102	0	100.0
Total	184	100.0

At the Private Practice the average number of sessions was 9.2 (SD 13.4, range 1-102). There were 29 clients who attended for only one session. The numbers of PP clients banded by number of sessions are shown (Table 5).

Table 5: For Private Practice clients, numbers of sessions this episode.

Number of sessions this episode	Number of clients	Cumulative %
0 (Exploratory only)	-	-
1-6	86	62.8
7-12	28	83.2
13-18	10	90.5
19-24	3	92.7
25-36	4	95.6
37-54	3	97.8
55-102	3	100.0
Total	137	100.0

Within the PP, 44 clients were paid for by their employer with varying set limits to the number of sessions, e.g. 4, 6, 10 or 12 sessions, sometimes with the option to extend if the employer agreed. For this group the average number of sessions was 6.3 (SD 3.8, range 1-18). The numbers of clients by sessions this episode, where the employer was paying for therapy are shown (Table 6).

There was a significant difference between the number of sessions for the 44 PP EAP clients (mean 6.3 sessions, SD 3.8) and the other 93 PP clients (mean 10.6 sessions, SD 15.9), $t(111.95) = 2.45$, $p = .016$, a small effect $r = .23$.

Table 6: For Private Practice clients paid for by their employer, number of sessions this episode.

Number of sessions this episode	Number of clients	Cumulative %
1	5	11.4
2	2	15.9
3	2	20.5
4	3	27.3
5	5	38.6
6	15	72.7
7	1	75.0
8	1	77.3
9	1	79.5
10	3	86.4
12	4	95.5
15	1	97.7
18	1	100.0
Total	44	100.0

5.5 Resulting sample

The sample was of 321 clients starting in the research. The mean age at the start of counselling was 31.2 years (SD 12.0) with 65.7% female. Marital status was 62.3% single, 26.2% married, 10.6% separated/divorced and 0.9% widowed. Parental status was 66.7% non-parent, 9.3% with one child, 17.8% with two children and 6.3% with three or more children. Ethnic status was requested from clients in free-format and this led to a variety of modes of completion for this data, the most consistent interpretation of this data was to state

the client's nationality and this was predominantly British, 88.2%, with clients from Europe (3.1%), America (2.8%) plus clients from Asia, Africa and Australia. Medication information was interpreted from the client's information sheet by a Member of the Royal College of General Practitioners (MRCGP) and allocated to different groups of psychologically relevant medication. Most clients (83.5%) were taking no psychologically relevant medication at the start of counselling, with 11.2% taking antidepressants only, with the remaining 5.3% of clients taking anxiolytics, sedatives or anti-psychotics, some in combination with anti-depressants; for 3 clients it was unknown whether they were taking any relevant medication at the start. The best available information was that 87.5% of clients were not taking relevant medication at the end of counselling with 12 clients still taking or phasing out anti-depressants at the end and two clients still taking anti-psychotics. Appendix 3 summarises the demographic information for the overall sample of 321 clients, the PP sample of 137 clients, the UCS sample of 184 clients and the subsets of the UCS sample; the 38 clients who do only an exploratory session, the 146 clients who enter therapy, the 59 clients who do not do the first session paperwork, the 87 clients who enter therapy, the 40 clients who do not do the last session paperwork, the 47 clients who do complete the last session paperwork and the 36 clients who complete the last session paperwork and have a 'proper' ending, not prematurely terminated by either the client or therapist leaving UEA.

The protocol differed slightly between UCS and PP. At UCS the protocol was that clients should be invited to complete outcome questionnaires at the end of therapy, accordingly clients with non-clinical scores at the start of counselling also completed an outcome questionnaire at the end of therapy. At the PP it was usually only clients who started off with clinical scores at the start of therapy who completed a subsequent outcome questionnaire. Furthermore the protocol at the UCS was for clients to re-complete outcome questionnaires at the end of therapy whereas at the PP clients were monitored through therapy by outcome questionnaire completion at a frequency determined by the author in conjunction with the client. Hence there was a reduced level of 'missing clients' at the PP cf. UCS. This difference in protocol gave rise to a different pattern of post-measure completion and is one difference between the two parts of the sample.

The three outcome questionnaires mainly used in this research were the BDI-II, BAI and CORE-OM. The PBQ was added at a much later date and played a more minor role in the research as an estimate of prevalence.

In terms of study samples the TREND guidelines (Des Jarlais, et al., 2004) recommended providing information on study samples (TREND item 1), numbers of sessions provided (TREND item 4), sample sizes (TREND item 7), flow of participants through each stage of the study, including numbers completing each stage of the study and those who did or did not complete subsequent measurement (TREND item 12), dates defining periods of recruitment and follow-up (TREND item 13), baseline demographic and clinical characteristics of participants in each study condition (TREND item 14), numbers of participants in each analysis of each condition, including analysis of those not completing subsequent measurement (TREND item 16) and results for each study condition, the estimated effect size and a confidence interval for the effect size (TREND item 17).

5.6 Participant flow and demographic characteristics for each sample analysed

Naturalistic research has been criticised e.g. Clark, et al. (2007). This research sought to address such criticisms and one such way to address these was to report the research in line with the TREND Guidelines (Des Jarlais, et al., 2004), the naturalistic equivalent to the CONSORT Guidelines. One of the requirements of the TREND Guidelines, described above, was to make clear what the participant flow was and the demographic characteristics for each sample analysed and data was prepared to comply with the TREND reporting requirements and is included as Appendix 3.

In Appendix 3, for each of the main outcome measures, there is a participant flow diagram and a summary of demographics for each sample subsequently reported upon.

The main samples analysed were as follows:

- Depression (BDI-II)
- Anxiety (BAI)
- Distress (CORE-OM)
- Clients in wait-control analysis (depression – BDI-II, anxiety – BAI and distress – CORE-OM)
- Clients in hypothesis testing (outcomes A1 to A3 and prediction B1 to B3)

The protocol for clients at the UCS was to complete outcome measures at Exploratory, First and Last Session. The intention was that each client would then act as their own wait control; outcome measure at start of wait compared with outcome measure at end of wait, although each client had a session with a qualified therapist as their exploratory so had received some counselling input. Changes during wait could then be compared with changes during 'active treatment'; outcome measure at first session compared with last session. The demographic characteristics of the clients in the 'wait time analysis', comparing change during wait with treatment are shown in Appendix 3.

It is important to note a key difference between the outcome and process-outcome samples for the analysis. Analysis of outcomes for clients with clinical scores, by definition, *excludes* clients with sub-clinical scores, whereas the predictor samples *include* clients with sub-clinical scores. This is because a test of person-centred theory (Rogers, 1957, 1959) must include clients who are 'well' (congruent clients in addition to incongruent clients) within the sample (Watson, 1984, p. 37)

5.7 Analytical approach

SPSS 15.0 and 18.0 for Windows was used for the analysis. Outcome hypotheses were assessed using repeated measures analysis of variance to assess statistical significance of any changes (Field, 2005), so that this was consistent with when there were three measurement points (see below) and, using reliable change criteria, proportions of clients with reliable change were counted. Predictor hypotheses were assessed using multivariate regression (Field, 2005). Further results are provided to support the findings, to address some of the criticisms in the literature of naturalistic research and further understand the findings. Further results used independent groups t-test, Wilcoxon signed-rank test, regression analysis, one way analysis of variance, repeated measures analysis of variance with between-subjects factors (Field, 2005) and repeated measures analysis of covariance (M. Adams, 28th August 2008).

6. Results

In addition to the results of the hypotheses testing this section provides a wider picture to the context of the hypotheses and samples tested.

6.1 Outcomes

Naturalistic research has been criticised for seeming to pick the best outcomes only to report (Clark, et al., 2007); a criticism that has been challenged (Stiles, et al., 2008). For completeness, for the main outcome hypotheses, effect sizes are shown on a Last Observation Carried Forward basis (LOCF) and also where a subsequent outcome measurement was made. This section seeks to provide some results in addition to the direct reporting of the outcomes hypotheses. To the extent that it was possible, with the data available, some of the criticisms of uncontrolled naturalistic research (Clark, et al., 2007) are addressed in section ‘7. Further Results’.

The demographic characteristics of the clients tested for each of the hypotheses are shown (Appendix 3). The findings for each of the outcome hypotheses are shown (Table 7) and this section also presents some additional results that support and fill out the picture shown by the hypothesis testing.

Table 7: Summary of outcomes hypotheses.

Hypothesis	n	Start		Subsequent		ES(d)	95% CI	p	Hypothesis conclusion
		Mean	SD	Mean	SD				
Depression	111	26.9	9.6	12.7	9.1	1.48	1.28-1.68	<.001	Accept experimental
Anxiety	91	19.7	9.6	8.6	7.1	1.15	.95-1.35	<.001	Accept experimental
Distress	79	18.4	5.0	9.5	5.2	1.80	1.53-2.06	<.001	Accept experimental

Note: ^a ES(d) effect size calculated by mean score at start minus subsequent mean score divided by start standard deviation.

6.1.1. Depression (BDI-II) outcomes

a) Effect sizes

There were 205 clients who completed a BDI-II at their first therapy session, and for these clients, starting with any level of BDI-II score, on a Last Observation Carried Forwards (LOCF) basis mean client BDI-II scores improved significantly $F(1, 204) = 118.42, p < .001$, $ES(d) = .70$ (95% CI .57 to .82). Of these clients, 124 (60.5% of the 205 clients) completed a subsequent BDI-II and their mean BDI-II scores improved significantly $F(1, 123) = 195.56, p < .001$, $ES(d) = 1.21$ (95% CI 1.04 to 1.39). Participant flow is shown, as are demographic characteristics in Appendix 3; mean start and subsequent BDI-II scores, effect sizes and significance values for change in repeated measures analysis of variance are shown (Table 8).

The aim of the outcomes part of this research was to report on clients starting their first session of therapy with a clinical level of depression (BDI-II ≥ 14) and a subsequent measure of their depression. Of the 205 clients with a depression measurement at their first session, 162 (79.0% of 205 clients) had a clinical level of depression at their first session (BDI-II score ≥ 14). For these 162 clients, on an LOCF basis, mean BDI-II scores improved significantly $F(1, 161) = 127.74, p < .001$, $ES(d) = 1.02$ (95% CI .85 to 1.20). Of these clients 111 (68.5% of 162 clients) completed a subsequent BDI-II and their mean BDI-II scores improved significantly $F(1, 110) = 212.60, p < .001$, $ES(d) = 1.48$ (95% CI 1.28 to 1.68). It was this group that were the target of the depression outcomes hypothesis (A.1), the evidence suggested mean client BDI-II scores improved significantly and the null hypothesis, mean BDI-II scores did not improve, was highly unlikely. These effect sizes were compared with other studies (Table 9), see Discussion. Using transformed scores ($n = 111$) the mean start score of 5.49 (SD .83) became mean subsequent score 3.94 (SD 1.07) with a slightly higher effect size $ES(d) = 1.8713$ which would be difficult to compare with other studies using raw scores.

Subtracting clients with ‘severe depression’ from the clients with ‘clinical depression’ left a sample of ‘non-severely depressed clients’ with what might be termed ‘low severity depression’ (Dimidjian, et al., 2006). For ‘low severity depression’ clients (start BDI-II scores ≥ 14 and < 29) on an LOCF basis ($n = 110$) start BDI-II mean score 20.8, SD 4.08, LOCF BDI-II mean score 14.1, SD 7.21 a statistically significant improvement $F(1, 109) = 92.66, p < .001$, $ES(d) = 1.62$ (95% CI 1.28 to 1.95) and with subsequent measure ($n = 70$)

start BDI-II mean 21.0, SD 4.23, subsequent BDI-II mean score 10.3, SD 6.4, a statistically significant improvement $F(1, 69) = 179.23$, $p < .001$, $ES(d) = 2.50$ (95% CI 2.13 to 2.88).

Table 8: Depression (BDI-II) outcomes.

Group	n	First		Subsequent		ES(d) ^a	p
		Mean	SD	Mean	SD		
All clients with any BDI-II score at first session:							
All _{LOCF} ^b	205	22.21	10.948	14.54	9.800	.7005	<.001
UCS _{LOCF}	87	21.26	10.144	15.93	9.633	.5254	<.001
PP _{LOCF}	118	22.91	11.497	13.52	9.837	.8167	<.001
All _{subseq} ^c	124	24.96	10.703	11.92	9.017	1.2183	<.001
UCS _{subseq}	48	21.40	11.058	11.29	8.032	.9142	<.001
PP _{subseq}	76	27.21	9.899	12.32	9.618	1.5041	<.001
All clients with clinical BDI-II score at first session (BDI-II ≥ 14):							
All _{LOCF}	162	25.83	9.163	16.40	9.989	1.0291	<.001
UCS _{LOCF}	68	24.91	8.109	18.60	8.999	.7781	<.001
PP _{LOCF}	94	26.50	9.845	14.80	10.404	1.1884	<.001
All_{subseq}	111	26.86	9.550	12.68	9.116	1.4848	<.001
UCS _{subseq}	36	25.69	9.017	13.19	8.031	1.3862	<.001
PP _{subseq}	75	27.41	9.805	12.43	9.634	1.5277	<.001
All clients with severe BDI-II score at first session (BDI-II ≥ 29):							
All _{LOCF}	52	36.73	7.151	21.33	12.935	2.1535	<.001
UCS _{LOCF}	17	36.24	7.111	23.76	12.862	1.7550	.005
PP _{LOCF}	35	36.97	7.262	20.14	12.989	2.3175	<.001
All _{subseq}	41	36.93	7.414	16.68	11.499	2.7313	<.001
UCS _{subseq}	10	37.40	7.989	16.20	10.830	2.6536	.002
PP _{subseq}	31	36.77	7.352	16.84	11.875	2.7108	<.001

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at start. ^b Last Observation Carried Forward. ^c Clients with a BDI-II measurement subsequent to their first session score. UCS = University Counselling Service. PP = Private Practice. **Bold** indicates clients in the hypotheses testing samples.

Table 9: Depression (BDI-II) outcomes: Comparison with other studies.

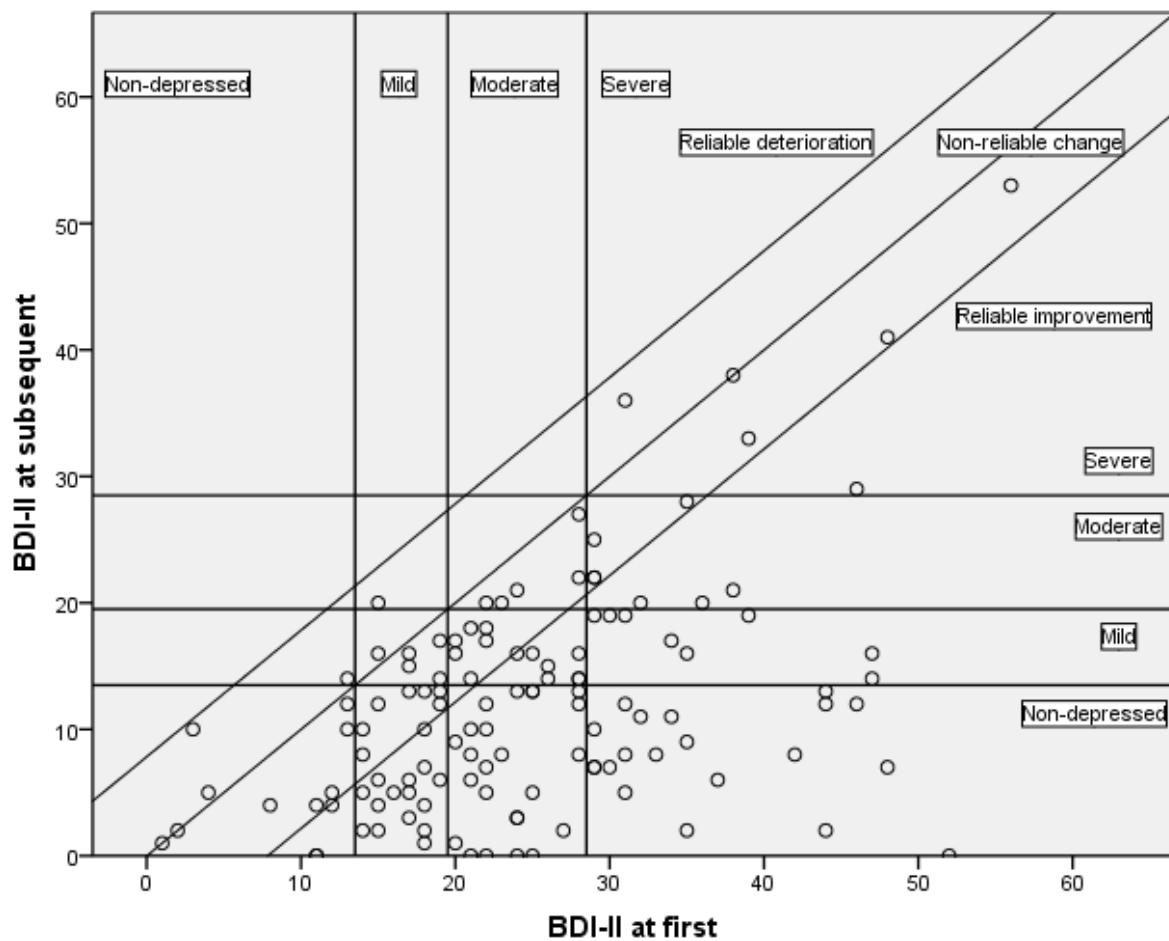
Study	n	Start		Subsequent		ES(d) ^a	Group	Comment
		Mean	SD	Mean	SD			
This study	162	25.83	9.163	16.40	9.989	1.0291	LOCF ^b	Clinical depression (Start with BDI-II >= 14).
	111	26.86	9.550	12.68	9.116	1.4848	Subseq^c	
	52	36.73	7.151	21.33	12.935	2.1535	LOCF ^b	Severe Depression (Start with BDI-II >= 29).
	41	36.93	7.414	16.68	11.499	2.7313	Subseq^c	
Ward, et al., (2000)	56	27.6	8.4	12.7	9.5	1.7738	CBT	Using BDI. Outcomes at 4 months for randomised completer clients. Data taken from table 2 of reference.
	62	25.4	8.6	11.5	7.7	1.6162	PCT	
	62	26.5	8.9	17.2	11.9	1.0449	Usual GP care	
Watson, et al., (2003)	33	26.00	9.03	10.27	9.62	1.7419	CBT	Using BDI. Outcomes for completer clients. Data taken from table 2 of reference.
	33	23.24	7.81	9.03	8.63	1.8194	PE	
Elliott, et al., (2004)	Meta-analysis of 23 studies (9 classical PCT, 6 PE, etc.)					1.18		Depression outcomes. Data taken from page 514 of reference.
Missirlian, et al., (2005)	32	24.59	6.08	9.16	5.13	2.5378	PE	Using BDI. Pre and post completer data from table 1 of reference.
Dimidjian, et al., (2006)	17	27.30	6.89	9.76	8.15	2.5457	Cognitive Therapy	Using BDI. Low severity completer clients at 16 weeks, apart from placebo, data at 8 weeks (treatment change at 8 weeks for this group). Data from table 2 of reference. Note 48.8% of Paroxetine group had left trial by 16 weeks.
	13	28.72	4.59	11.00	10.08	3.8605	Behavioural Activation	
	22	27.79	5.67	7.91	6.29	3.5061	Paroxetine	
	19	26.59	5.43	14.68	7.81	2.1933	Pill Placebo	
Minami, et al., (2007)	Meta-analysis of 29 studies					1.859	Completers	Aggregated benchmarks using BDI. Typical duration 15-16 weeks. Data from table 1 of reference. Control is natural history benchmark.
	Meta-analysis of 11 studies					1.706	LOCF	
	Meta-analysis of 11 studies					.371	Control	

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at start. ^b Last Observation Carried Forward. ^c Clients with a BDI-II measurement subsequent to their first session score. **Bold** indicates clients in the hypotheses testing samples for this research.

b) Reliable change

In addition to reporting mean change in outcome measures it is good practice to analyse reliable change (Jacobson & Truax, 1991). A Jacobson plot is provided for the 124 clients with any BDI-II score at First Session and a subsequent BDI-II score, of which 111 clients started therapy with a clinical depression score (Figure 5).

Figure 5: Jacobson Plot for Depression (BDI-II) Outcomes.



Note: Severity levels as per author's definitions (Beck et al 1996) and reliable change index as per this research. Changes greater than 7.8 BDI-II units were considered 'reliable'.

Client numbers in each severity group at First and Subsequent Session are shown (Table 10).

Table 10: Depression (BDI-II) severity at first and subsequent measurement.

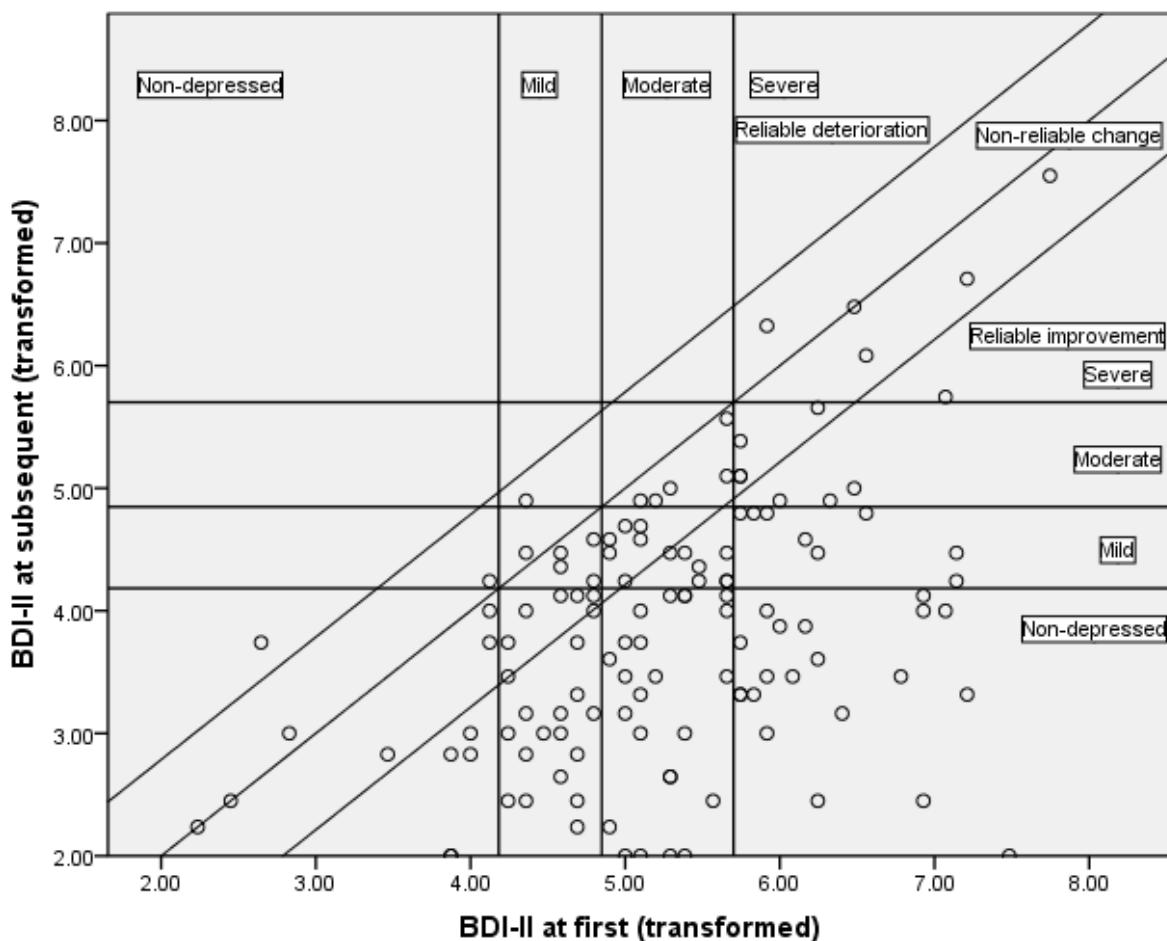
		BDI-II Severity at First Session				Total	
		Non	Mild	Moderate	Severe		
BDI-II Severity at Subsequent Session	Non	12	22	24	20	78	
	Mild	1	5	13	8	27	
	Moderate	-	1	5	7	13	
	Severe	-	-	-	6	6	
		Total	13	28	42	41	124

For the 124 clients with a BDI-II score at First Session and a subsequent session there were no clients who reliably deteriorated. For the 111 clients who started therapy with a BDI-II score in the clinical range, 29.7% had no reliable change and 70.3% reliably improved (Table 11); the 70.3% with reliable improvement further breaks down into 53.2% of clients who had 'recovered' (reliable change and non-clinical BDI-II score at subsequent measurement) and 17.1% of clients who had reliable improvement only. Transformed data gave a slightly higher estimate of 71.2% (cf. 70.3%) of clients with reliable improvement, see Figure 6.

Table 11: Percentages of clients with reliable change and recovered from depression (BDI-II).

Group	n	Reliable Deterioration ^a	No Reliable Change ^b	Reliable Change ^c		
				I or R ^d	Improved ^e	Recovered ^f
All clients with clinical BDI-II score at first session (BDI >= 14)						
All _{subseq}	111	0.0%	29.7%	70.3%	17.1%	53.2%
UCS _{subseq}	36	0.0%	41.7%	58.3%	13.9%	44.4%
PP _{subseq}	75	0.0%	24.0%	76.0%	18.7%	57.3%
All clients with severe BDI-II score at first session (BDI >= 29)						
All _{subseq}	41	0.0%	22.0%	78.0%	29.3%	48.8%
UCS _{subseq}	10	0.0%	20.0%	80.0%	40.0%	40.0%
PP _{subseq}	31	0.0%	22.6%	77.4%	25.8%	51.6%

Note: ^a 'Reliable deterioration' was a deterioration of more than 7.8 BDI-II units. ^b 'No reliable change' was a change of less than 7.8 BDI-II units. ^c 'Reliable change' was an improvement of more than 7.8 BDI-II units. ^d 'I or R' was 'improved' or 'recovered' at subsequent measurement, i.e. 'improved' (improvement of more than 7.8 BDI-II units) or 'recovered' (an improvement of more than 7.8 BDI-II units and a subsequent score of 13 or less). ^e 'Improved' was reliable change only, i.e. improvement of more than 7.8 BDI-II units and subsequent score was >=14. ^f 'Recovered' was reliable change and non-clinical score at subsequent, i.e. improvement of more than 7.8 BDI-II units and a subsequent score of 13 or less.

Figure 6: Jacobson Plot for Depression (BDI-II) Outcomes using transformed data.

Note: Severity levels (transformed) as per author's definitions (Beck et al 1996) and reliable change index (transformed) as per this research. Changes greater than 0.786 transformed BDI-II units were considered 'reliable'.

c) Severe depression

NICE are interested in outcomes for clients with severe depression (NICE, 2004b) and outcomes are shown (Table 8) for the subgroup of clients with depression scores where the BDI-II scores at first session were in the range the authors of this questionnaire (Beck et al 1996) defined as 'severe' ($BDI-II >= 29$). It was intended that outcomes from this study would be compared with the studies of Missirlian, et al., (2005) and Dimidjian, et al., (2006) which each evaluated clients with relatively severe depression. Of the 205 clients who completed a BDI-II at their first session, 52 clients (25.4%) met the criteria for 'severe depression'. On an LOCF basis there was an overall significant improvement $F(1, 51) = 66.14$, $p < .001$, $ES(d) = 2.15$ (95% CI 1.62 to 2.68). Of the 52 clients, 41 (78.8%) had a

subsequent measurement of their depression and there was on average a significant improvement $F(1, 40) = 109.94$, $p < .001$, $ES(d) = 2.73$ (95% CI 2.20 to 3.25).

Clients with severe depression appear on the Jacobson plot (Figure 6) and in the severity table (Table 10). No clients reliably deteriorated during treatment (Table 11), 22.0% had no reliable change, 78.0% had reliable change; this broke down into 29.3% with reliable change only and 48.8% 'recovered' (reliable change and non-clinical BDI-II).

6.1.2 Anxiety (BAI) outcomes

a) Effect sizes

There were 204 clients who completed a BAI at their first therapy session, and for these clients, starting with any level of BAI score, on an LOCF basis mean client BAI scores improved significantly $F(1, 203) = 63.68, p < .001, ES(d) = .46$ (95% CI .35 to .58). Of these clients, 102 (50.0% of the 204 clients) completed a subsequent BAI and their mean BAI scores improved significantly $F(1, 101) = 98.75, p < .001, ES(d) = .94$ (95% CI .75 to 1.13). Participant flow is shown as are demographic characteristics in Appendix 3; mean start and subsequent BAI scores, effect sizes and significance values for change in repeated measures analysis of variance are shown (Table 12).

The aim of the outcomes part of this research was to report on clients starting their first session of therapy with a clinical level of anxiety (BAI ≥ 8) and a subsequent measure of their anxiety. Of the 204 clients with an anxiety measurement at their first session, 156 (76.5% of 204 clients) had a clinical level of anxiety at their first session (BAI score ≥ 8). For these 156 clients, on an LOCF basis, mean BAI scores improved significantly $F(1, 155) = 77.20, p < .001, ES(d) = .69$ (95% CI .54 to .85). Of these clients, 91 (58.3% of 156 clients) completed a subsequent BAI and their mean BAI scores improved significantly $F(1, 90) = 127.88, p < .001, ES(d) = 1.15$ (95% CI .95 to 1.35). It was this group that were the target of the anxiety outcomes hypothesis (A.3), the evidence suggested mean client BAI scores improved significantly and the null hypothesis, mean BAI scores did not improve, was highly unlikely. These effect sizes were compared with other studies (Table 13), see Discussion. Using transformed scores ($n = 91$) the mean start score of 4.44 (SD .99) became mean subsequent score 2.89 (SD 1.13) with a slightly higher effect size $ES(d) = 1.5687$ which would be difficult to compare with other studies using raw scores.

It was intended that outcomes from this study would be compared with the study of Barrowclough, et al., (2001) which had a mean start BAI score of 27.26 (SD 9.44), for comparison a sample from this study was selected with BAI ≥ 19 which gave a mean start BAI score of 28.61 (SD 8.199) and mean end 10.39 (SD 8.192), this was a significant improvement $F(1, 37) = 165.09, p < .001, ES(d) = 2.2222$ (95% CI 1.87 to 2.57), see discussion.

Table 12: Anxiety (BAI) outcomes.

Group	n	First		Subsequent		ES(d) ^a	p
		Mean	SD	Mean	SD		
All clients with any BAI score at first session:							
All _{LOCF} ^b	204	15.38	10.058	10.66	8.468	.4692	<.001
UCS _{LOCF}	87	15.20	10.526	12.59	8.958	.2479	<.001
PP _{LOCF}	117	15.52	9.739	9.22	7.819	.6468	<.001
All _{subseq} ^c	102	17.99	10.242	8.32	7.043	.9441	<.001
UCS _{subseq}	47	15.53	11.300	10.66	7.976	.4309	<.001
PP _{subseq}	55	20.09	8.813	6.33	5.457	1.5613	<.001
All clients with clinical BAI score at first session (BAI >= 8):							
All _{LOCF}	156	18.76	9.060	12.44	8.743	.6975	<.001
UCS _{LOCF}	68	18.40	9.632	14.69	8.689	.3851	<.001
PP _{LOCF}	88	19.05	8.638	10.70	8.428	.9666	<.001
All_{subseq}	91	19.65	9.560	8.59	7.124	1.1569	<.001
UCS _{subseq}	37	18.59	10.792	11.73	8.143	.6356	<.001
PP _{subseq}	54	20.37	8.647	6.44	5.438	1.6109	<.001

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at start. ^b Last Observation Carried Forward. ^c Clients with a BAI measurement subsequent to their first session score. **Bold** indicates clients in the hypotheses testing samples.

Table 13: Anxiety (BAI) outcomes: Comparison with other studies.

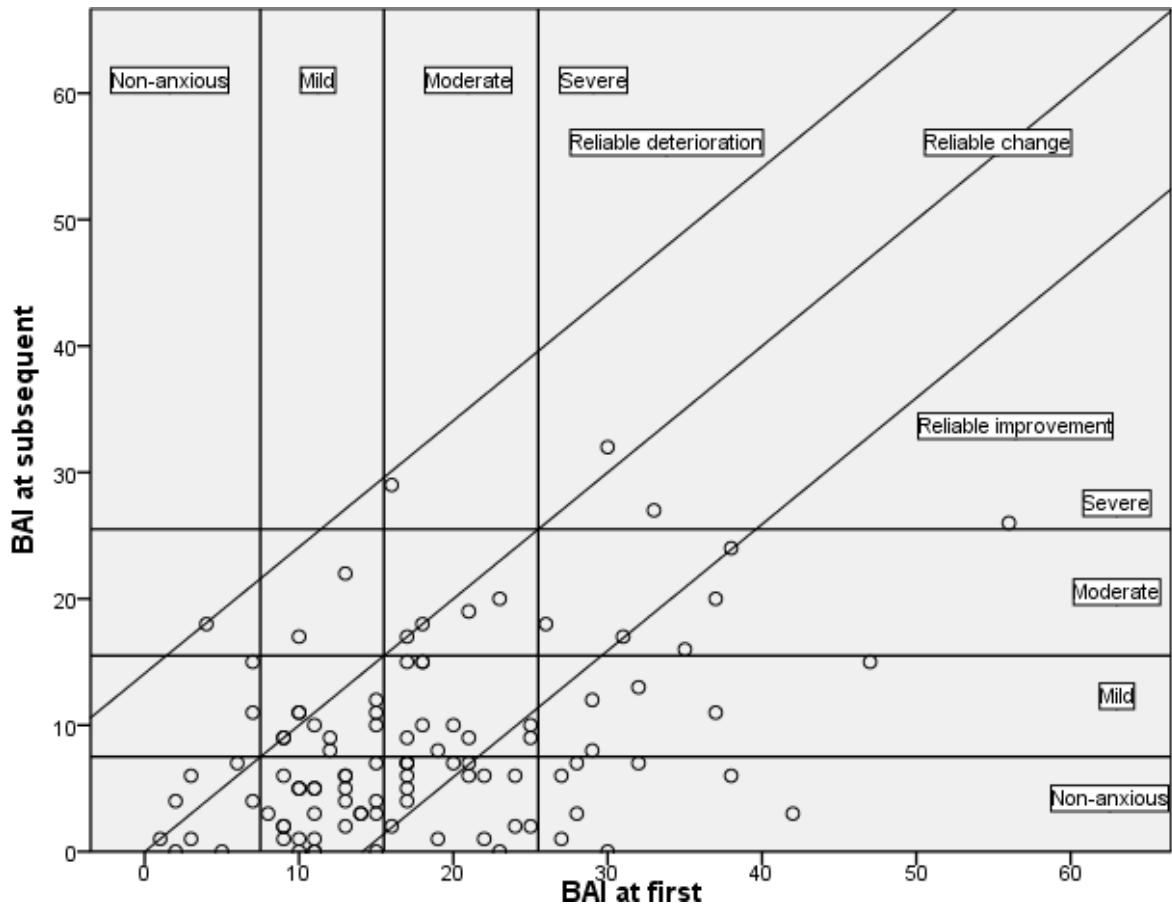
Study	n	Start		Subsequent		ES(d) ^a	Group	Comment	
		Mean	SD	Mean	SD				
This study	156	18.76	9.060	12.44	8.743	.6975	LOCF ^b	Clinical anxiety (Start with BAI >= 8)	
	91	19.65	9.560	8.59	7.124	1.1569	Subseq^c		
Borkovec and Whisman, (1996) ⁴	Meta-analysis of 2 studies				.24	CT	Cognitive therapy		
	Meta-analysis of 7 studies				.90	BT	Behaviour therapy		
	Meta-analysis of 7 studies				1.01	CBT			
Gould, et al., (1997) ⁴	Meta-analysis of 22 studies				.34	Relax +	Relaxation training with bio-feedback (2 studies)		
					.51	BT	Behaviour therapy (3 studies)		
					.59	CT	Cognitive therapy (3 studies)		
					.64	Relax	Relaxation training (3 studies)		
					.91	CBT	(8 studies)		
Bryant, et al., (1998)	12	STAI State STAI Trait			1.55	CBT	Five 1.5 hour sessions of CBT for 'acute stress disorder' within 2 weeks of 'civilian trauma'.		
					.86				
Bryant, et al., (1999)	15	STAI Trait			1.45	PE+AM	Prolonged exposure and anxiety management.		
	14				1.39	PE only	Prolonged exposure only.		
	16				.85	SC	Supportive counselling (psychological placebo).		
Barrowclough, et al., (2001)	19	27.26	9.44	11.58	9.17	1.68	CBT	Test of CBT (2 therapists) for anxiety in older adults versus SC (supportive counselling) provided by one counsellor.	
	24	26.46	12.84	17.46	12.17	.71	SC		
Elliott, et al., (2004)	Meta-analysis of 8 studies				1.30	PE	Process Experiential therapies. Page 511.		
Westen and Morrison, (2001) ^d	Meta-analysis of 5 studies				2.09	CBT	Conservative inclusion criteria.		

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at start. ^b Last Observation Carried Forward. ^c Clients with a BAI measurement subsequent to their first session score. ^d Also included in subsequent meta-analysis (Deacon & Abramowitz, 2004). **Bold** indicates clients in the hypotheses testing samples for this research.

b) Reliable change

A Jacobson plot is provided for the 102 clients with any BAI score at First Session and a subsequent BAI score, of which 91 clients started therapy with a clinical anxiety score (Figure 7).

Figure 7: Jacobson Plot for Anxiety (BAI) Outcomes.



Note: Severity levels as per author's definitions (Beck & Steer, 1993) and reliable change index as per this research; Changes greater than 14.1 BAI units were considered 'reliable'.

Client numbers in each severity group at First and Subsequent Session are shown (Table 14).

Table 14: Anxiety (BAI) severity at first and subsequent measurement.

		BAI Severity at First Session				Total	
		Non	Mild	Moderate	Severe		
BAI Severity at Subsequent Session	Non	8	26	17	8	59	
	Mild	2	10	10	5	27	
	Moderate	1	2	4	5	12	
	Severe	-	-	1	3	4	
		Total	11	38	32	21	102

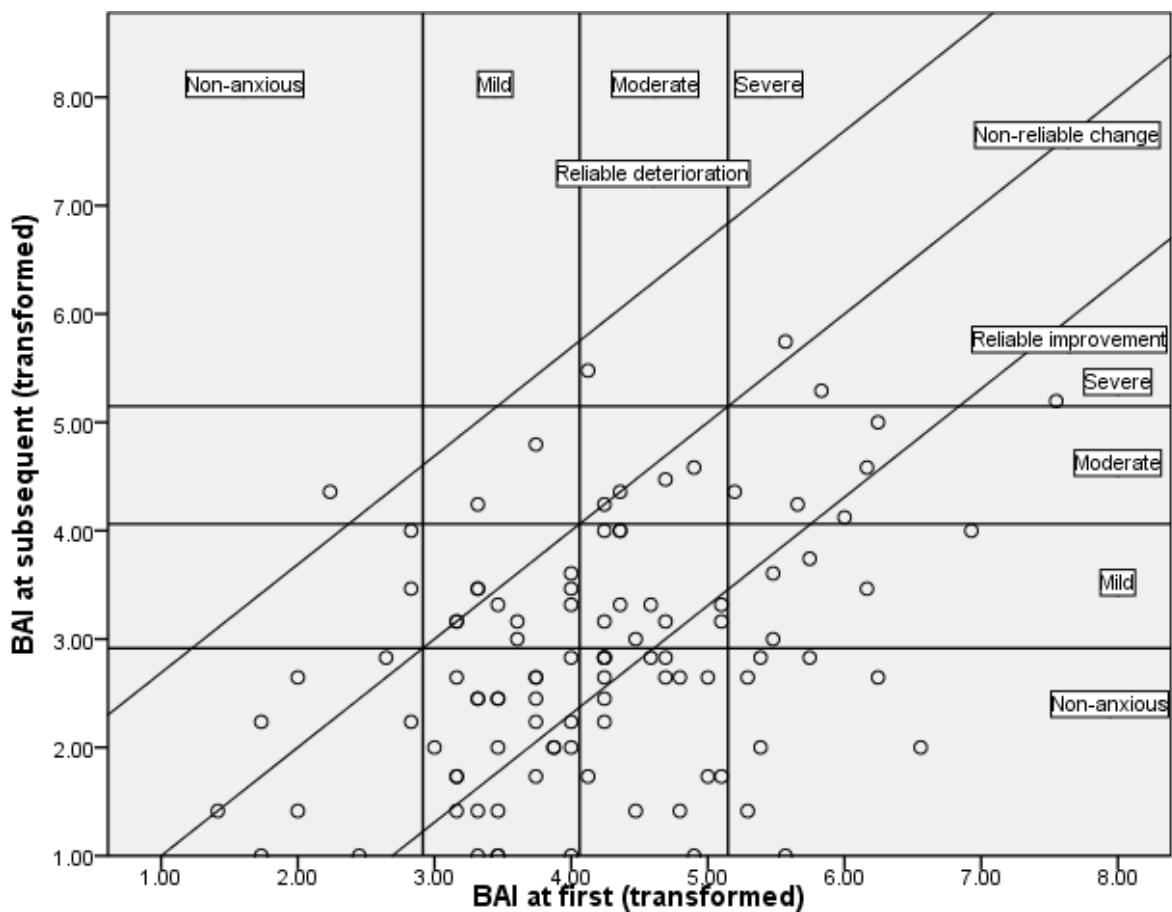
For the 102 clients with a BAI score at First Session and a subsequent session there were no clients who reliably deteriorated (Figure 7). For the 91 clients who started therapy with a BAI score in the clinical range, no clients reliably deteriorated, 70.3% had no reliable change and 29.7% reliably improved (Table 15); the 29.7% with reliable improvement further breaks down into 18.7% of clients who had ‘recovered’ (reliable change and non-clinical BAI score at subsequent measurement) and 11.0% of clients who had reliable improvement only. With the transformed data of the 91 starting with clinical anxiety, 46.2% reliably improved, see Figure 8.

Table 15: Percentages of clients with reliable change and recovered from anxiety (BAI).

Group	n	Reliable Deterioration ^a	No Reliable Change ^b	Reliable Change ^c		
				I or R ^d	Improved ^e	Recovered ^f
All clients with clinical BAI score at first session (BAI ≥ 8)						
All _{subseq}	91	0.0%	70.3%	29.7%	11.0%	18.7%
UCS _{subseq}	37	0.0%	83.8%	16.2%	16.2%	0.0%
PP _{subseq}	54	0.0%	61.1%	38.9%	7.4%	31.5%

Note: ^a ‘Reliable deterioration’ was a deterioration of more than 14.1 BAI units. ^b ‘No reliable change’ was a change of less than 14.1 BAI units. ^c ‘Reliable change’ was an improvement of more than 14.1 BAI units. ^d ‘I or R’ was ‘improved’ or ‘recovered’ at subsequent measurement, i.e. ‘improved’ (improvement of more than 14.1 BAI units) or ‘recovered’ (an improvement of more than 14.1 BAI units and a subsequent score of BAI score of 7 or less). ^e ‘Improved’ was reliable change only, i.e. improvement of more than 14.1 BAI units and subsequent BAI score was 8 or more. ^f ‘Recovered’ was reliable change and non-clinical score at subsequent measurement, i.e. improvement of more than 14.1 BAI units and a subsequent BAI score of 7 or less.

Figure 8: Jacobson Plot for Anxiety (BAI) Outcomes using transformed data.



6.1.3 Distress (CORE-OM) outcomes

a) Effect sizes

There were 155 clients who completed a CORE-OM at their first therapy session, and for these clients, starting with any level of CORE-OM score, on an LOCF basis mean CORE-OM scores improved significantly $F(1, 154) = 84.01, p < .001, ES(d) = .74$ (95% CI .58 to .90). Of these clients, 89 (57.4% of the 155 clients) completed a subsequent CORE-OM and their mean CORE-OM scores improved significantly $F(1, 88) = 141.41, p < .001, ES(d) = 1.28$ (95% CI 1.07 to 1.50). Participant flow is shown as are demographic characteristics in Appendix 3; mean start and subsequent CORE-OM scores, standard deviations, effect sizes and significance values for change in repeated measures analysis of variance are shown (Table 16).

The aim of the outcomes part of this research was to report on clients starting their first session of therapy with a clinical level of distress (CORE-OM ≥ 10) and a subsequent measure of their distress. Of the 155 clients with a distress measurement at their first session, 130 (83.9% of 155 clients) had a clinical level of distress at their first session (CORE-OM score ≥ 10). For these 130 clients, on an LOCF basis, mean CORE-OM scores improved significantly $F(1, 129) = 95.83, p < .001, ES(d) = 1.11$ (95% CI .89 to 1.34). Of these clients, 79 (60.8% of 130 clients) completed a subsequent CORE-OM and their mean CORE-OM scores improved significantly $F(1, 78) = 182.71, p < .001, ES(d) = 1.80$ (95% CI 1.53 to 2.06). It was this group that were the target of the distress outcomes hypothesis (A.3), the evidence suggested mean CORE-OM scores improved significantly and the null hypothesis, mean CORE-OM scores did not improve, was highly unlikely. These effect sizes were compared with other studies (Table 17), see discussion.

Table 16: Distress (CORE-OM) outcomes.

Group	n	First		Subsequent		ES(d) ^a	p
		Mean	SD	Mean	SD		
All clients with any CORE score at first session:							
All _{LOCF} ^b	155	16.163	6.1266	11.600	6.2261	.7447	<.001
UCS _{LOCF}	87	15.568	6.1172	12.191	6.0648	.5520	<.001
PP _{LOCF}	68	16.925	6.0988	10.843	6.3917	.9972	<.001
All _{subseq} ^c	89	16.980	6.1955	9.015	5.1618	1.2856	<.001
UCS _{subseq}	47	15.357	6.6514	9.068	4.7036	.9455	<.001
PP _{subseq}	42	18.796	5.1334	8.957	5.6884	1.9166	<.001
All clients with clinical CORE-OM score at first session (CORE-OM >= 10):							
All _{LOCF}	130	17.962	4.8705	12.527	6.2784	1.1159	<.001
UCS _{LOCF}	72	17.52	4.6882	13.460	5.7537	.8660	<.001
PP _{LOCF}	58	18.509	5.0751	11.369	6.7461	1.4068	<.001
All_{subseq}	79	18.392	4.9641	9.453	5.1722	1.8007	<.001
UCS _{subseq}	38	17.663	5.0138	9.969	4.4751	1.5345	<.001
PP _{subseq}	41	19.067	4.8813	8.974	5.7579	2.0676	<.001

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at start. ^b Last Observation Carried Forward. ^c Clients with a CORE-OM measurement subsequent to their first session score. **Bold** indicates clients in the hypotheses testing samples.

Table 17: Distress (CORE-OM) outcomes: Comparison with other studies.

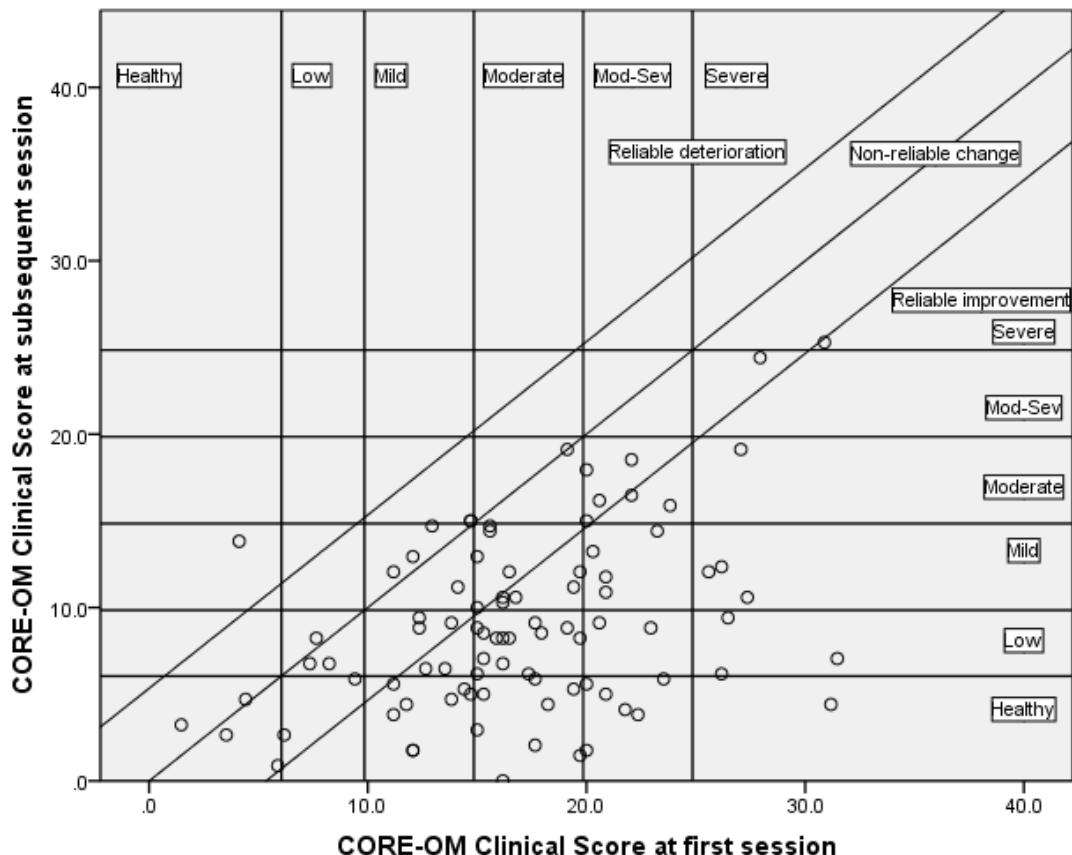
Study	n	Start		Subsequent		ES(d) ^a	Group	Comment
		Mean	SD	Mean	SD			
This study	130	17.962	4.8705	12.527	6.2784	1.1159	LOCF ^b	Clinical distress (CORE-OM >= 10)
	79	18.392	4.9641	9.453	5.1722	1.8007	Subseq^c	
Elliott, et al., (2004)	127 treatment groups					.99	PE	Group mean (Process experiential therapies).
						.82	PE	Weighted by sample size
	42 controlled studies					.89	PE	Group mean
						.78	PE	Weighted by sample size
						.11		Untreated conditions
Stiles, et al., (2006)	1,309	17.41	6.52	8.50	6.27	1.36		58 NHS primary care settings (CBT, PDT, PCT)
Mullin, et al., (2006)	11,953	17.5	6.3	8.5	6.3	1.42		32 NHS primary care counselling services
Stiles, et al., (2007)	5,613	17.60	6.33	8.77	6.43	1.39		32 NHS primary care settings (CBT, PDT, PCT)

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at start. ^b Last Observation Carried Forward. ^c Clients with a CORE-OM measurement subsequent to their first session score. **Bold** indicates clients in the hypotheses testing samples for this research.

b) Reliable change

A Jacobson plot is provided for the 89 clients with any CORE-OM score at First Session and a subsequent CORE-OM score, of which 79 clients started therapy with a clinical anxiety score (Figure 9).

Figure 9: Jacobson Plot for Distress Outcomes.



Note: Severity levels as per author's definitions (Barkham, et al, 2006) and reliable change index as per this research; Changes greater than 5.3 CORE-OM units were considered 'reliable'.

Client numbers in each severity group at First and Subsequent Session are shown (Table 18).

Table 18: Distress (CORE-OM) severity at first and subsequent measurement.

		CORE-OM Severity at First Session						Total
		Healthy	Low	Mild	Moderate	Mod/Sev	Severe	
CORE-OM Severity at subseq. Session	Healthy	4	2	8	8	6	1	29
	Low	-	3	5	13	2	3	26
	Mild	1	-	4	10	4	3	22
	Moderate	-	-	2	1	6	1	10
	Mod/Sev	-	-	-	-	-	1	1
	Severe	-	-	-	-	-	1	1
Total		5	5	19	32	18	10	89

For the 89 clients with a CORE-OM score at First Session and a subsequent session there was one client who reliably deteriorated (Figure 9). For the 79 clients who started therapy with a CORE-OM score in the clinical range, no clients reliably deteriorated, 25.3% had no reliable change and 74.7% reliably improved (Table 19); the 78.5% with reliable improvement further breaks down into 54.4% of clients who had ‘recovered’ (reliable change and non-clinical CORE-OM score at subsequent measurement) and 20.3% of clients who had reliable improvement only.

Table 19: Percentages of clients with reliable change or recovered from symptoms of distress (CORE-OM).

Group	n	Reliable Deterioration ^a	No Reliable Change ^b	Reliable Change ^c		
				I or R ^d	Improved ^e	Recovered ^f
All clients with clinical CORE-OM score at first session ≥ 10 :						
All _{subseq}	79	0.0%	25.3%	74.7%	20.3%	54.4%
UCS _{subseq}	38	0.0%	36.8%	63.2%	15.8%	47.4%
PP _{subseq}	41	0.0%	14.6%	85.4%	24.4%	61.0%
Clients with mild/moderate/moderate-to-severe CORE-OM score at first session:						
All _{subseq}	69	0.0%	27.5%	72.5%	15.9%	56.5%
UCS _{subseq}	34	0.0%	41.2%	58.8%	11.8%	47.1%
PP _{subseq}	35	0.0%	14.3%	85.7%	20.0%	65.7%
Clients with severe CORE-OM score at first session:						
All _{subseq}	10	0.0%	10.0%	90.0%	50.0%	40.0%
UCS _{subseq}	4	0.0%	0.0%	100.0%	50.0%	50.0%
PP _{subseq}	6	0.0%	16.7%	83.3%	50.0%	33.3%

Note: ^a ‘Reliable deterioration’ was a deterioration of more than 5.3 CORE-OM units. ^b ‘No reliable change’ was a change of less than 5.3 CORE-OM units. ^c ‘Reliable change’ was an improvement of more than 5.3 CORE-OM units. ^d ‘I or R’ was ‘improved’ or ‘recovered’ at subsequent measurement, i.e. ‘improved’ (improvement of more than 5.3 CORE-OM units) or ‘recovered’ (an improvement of more than 5.3 CORE-OM units and a subsequent CORE-OM score of less than 10). ^e ‘Improved’ was reliable change only, i.e. improvement of more than 5.3 CORE-OM units and subsequent CORE-OM score was 10 or more. ^f ‘Recovered’ was reliable change and non-clinical score at subsequent measure, i.e. improvement of more than 5.3 CORE-OM units and a subsequent CORE-OM score of less than 10.

Naturalistic observational studies can suffer from a lack of comparison between treatment and control cells. This drawback can be partly addressed by comparing this observational data with another study. Recovery, improvement and deterioration benchmarks have been published for 11,953 clients from 32 primary care NHS counselling and psychological therapy services (Mullin, et al., 2006). Results from this research ($n = 79$) were compared with the published benchmarks (Table 20). For the purposes of amalgamation ‘desirable’ qualities – high recovery rate, high improvement rate, low rate of ‘no reliable change’ and low rate of ‘reliable deterioration’ – were labelled ‘lower quartile’, ‘average’ or upper quartile’ as per the benchmark data, such that on this basis an all round well performing service would appear in the upper quartile for reliable deterioration, no reliable deterioration, reliable change, improved and recovered. Note: This is different to how the authors presented the benchmarks (Mullin, et al., 2006).

Table 20: Benchmarking percentages of clients with reliable change or recovered from symptoms of distress (CORE-OM): Compared with published benchmarks by Mullin et al., 2006.

Group	n	Reliable Deterioration ^a	No Reliable Change ^b	Reliable Change ^c		
				I or R ^d	Improved ^e	Recovered ^f
All clients with clinical CORE-OM score at first session $\geq 10^7$.						
All _{subseq}	79	Upper Q	Average	UQ/Av	Upper Q	Average
UCS _{subseq}	38	Upper Q	Lower Q	Lower Q	Lower Q	Lower Q
PP _{subseq}	41	Upper Q	Upper Q	Upper Q	Upper Q	Upper Q
Clients with mild/moderate/moderate-to-severe CORE-OM score at first session ⁷ .						
All _{subseq}	69	n/a	Average	Average	Average	Average
UCS _{subseq}	34	n/a	Lower Q	Lower Q	Average	Lower Q
PP _{subseq}	35	n/a	Upper Q	Upper Q	Upper Q	Upper Q
Clients with severe CORE-OM score at first session ⁷ .						
All _{subseq}	10	n/a	Average	Upper Q	Upper Q	Average
UCS _{subseq}	4	n/a	Upper Q	Upper Q	Upper Q	Upper Q
PP _{subseq}	6	n/a	Average	Average	Upper Q	Average

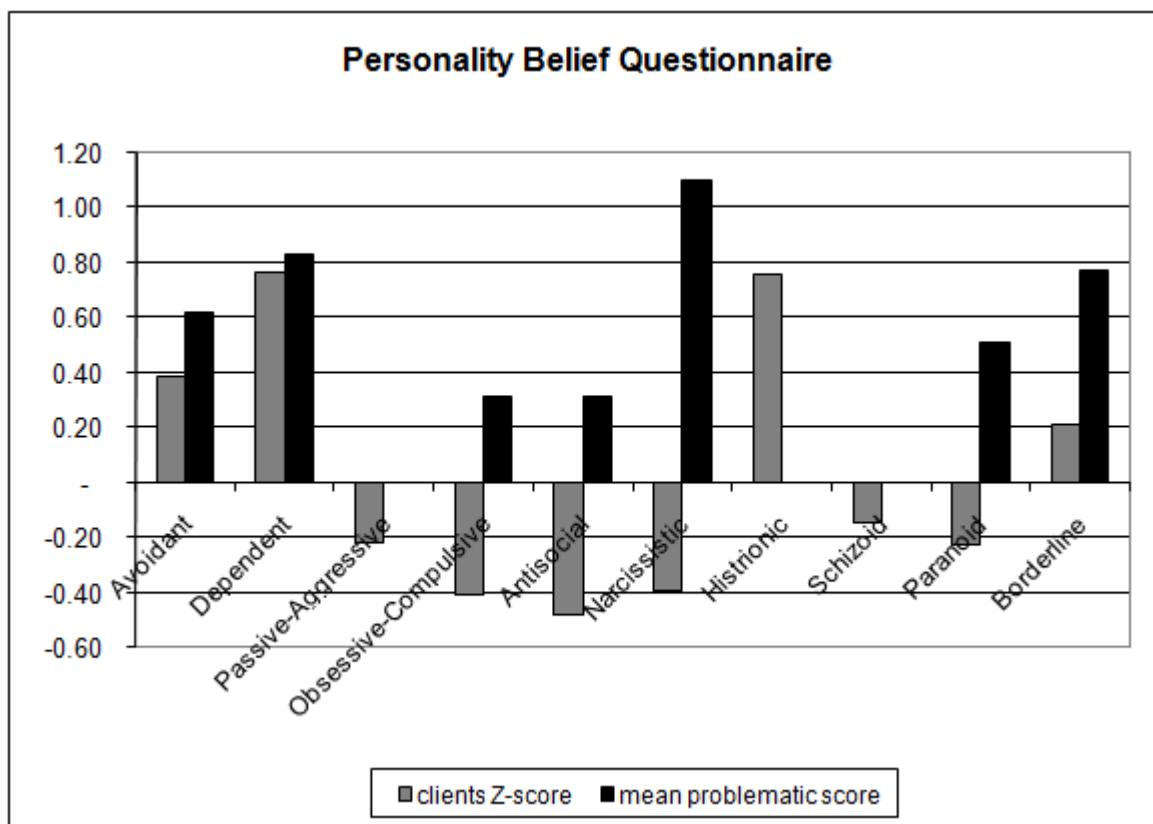
Note: ^a 'Reliable deterioration' was a deterioration of more than 5.3 CORE-OM units. ^b 'No reliable change' was a change of less than 5.3 CORE-OM units. ^c 'Reliable change' was an improvement of more than 5.3 CORE-OM units. ^d 'I or R' was 'improved' or 'recovered' at subsequent measurement, i.e. 'improved' (improvement of more than 5.3 CORE-OM units) or 'recovered' (an improvement of more than 5.3 CORE-OM units and a subsequent CORE-OM score of less than 10). ^e 'Improved' was reliable change only, i.e. improvement of more than 5.3 CORE-OM units and subsequent CORE-OM score was 10 or more. ^f 'Recovered' was reliable change and non-clinical score at subsequent measure, i.e. improvement of more than 5.3 CORE-OM units and a subsequent CORE-OM score of less than 10. ^g Compared with appropriate published data and labelled 'lower quartile', 'average' or 'upper quartile' on the basis that desirable qualities were high recovery rate, high improvement rate, low rate of 'no reliable change' and low rate of 'reliable deterioration'. n/a = Cell sizes for comparing reliable deterioration too small to provide adequate statistical comparison.

The cell sizes for comparing reliable deterioration when clients were split into 'less severe' and 'more severe' groups were too small to provide an adequate statistical comparison and are therefore 'not applicable' (n/a) as per Mullin et al., (2006). Thirteen comparisons were thus possible between this research, for the overall sample, and the published benchmarks; overall 4.5/13 cells were 'upper quartile' and 8.5/13 cells were 'average'. Amalgamating the benchmark comparisons in this way suggested that the reliable change results from the 79 clients in this research were about average or perhaps slightly above average when compared with the experience of 11,953 clients who attended 32 primary care NHS counselling and psychological therapy services.

6.1.4 Personality Disorder Prevalence (PBQ)

From 11th November 2007 the aim was that all new PP clients would complete a PBQ on intake. Several clients had scores consistent with more than one personality disorder. The questionnaire's authors instructed that for clients with more than one subscale score exceeding a clinical cut-off score for a personality disorder subscale the client should be recorded as having a primary personality disorder of the type shown by the subscale with the highest z-score. An example of a client PBQ score is shown (Figure 10).

Figure 10: Example of client PBQ score: Black bars indicate mean z-scores for patients with diagnosis of each personality disorder whilst this questionnaire was being tested (Beck, et al., 2001) and grey bars indicate this client's score for each personality disorder subscale.



Note: During testing there were insufficient numbers of clients to provide a mean z-score for the personality disorder subscales Passive-Aggressive, Histrionic and Schizoid (Beck, et al., 2001)

The client in this example (Figure 10) was recorded as having a primary personality disorder of 'dependent' in line with the authors' instructions (Beck, et al., 2001) plus 'avoidant' personality disorder because this client's score was closer to the clinical than the non-clinical score obtained during the questionnaire's development. In addition this client was also considered as having a 'histrionic' personality disorder because of how the high z-score compared with the other subscale clinical scores e.g. without this 'dependent' z-score this client would have been classified as having a primary personality disorder of 'histrionic' in line with the authors' instructions.

Four existing clients were invited to complete a PBQ and the scores for each of these were consistent with the presence of one or more personality disorders. There were 59 new clients to 20th August 2009 and 49 completed a PBQ, for various reasons ten did not, mainly because they had only one session. Of the 49 new clients tested, 41 had a score consistent with one or more personality disorders, implying a personality disorder prevalence rate of 83.7% for clients seen in the PP. In total 53 clients were tested (four existing and 49 new clients) and 45 had scores consistent with one or more personality disorders (84.9%). The most frequent primary personality disorder was avoidant with 9 clients each obtaining scores consistent with an avoidant personality disorder (Table 21).

Table 21: Primary Personality Disorder within Private Practice sample: Highest PBQ subscale z-scores for each client (n = 53).

Primary Personality Disorder	Frequency	Percent %	Cumulative Percent %
Avoidant	9	17.0	17.0
Dependent	6	11.3	28.3
Passive-Aggressive ^a	2	3.8	32.1
Obsessive-Compulsive	5	9.4	41.5
Antisocial	1	1.9	43.4
Narcissistic	2	3.8	47.2
Histrionic ^a	5	9.4	56.6
Schizoid ^a	5	9.4	66.0
Paranoid	8	15.1	81.1
Borderline	2	3.8	84.9
None	8	15.1	100.0
Total	53	100.0	100.0

^a These PBQ subscales had the highest z-scores for these clients and according to the instructions from the questionnaire's authors (Beck, et al., 2001) these clients were marked as having these primary personality disorders. All but two of these clients would have satisfied criteria for another personality disorder. Given the scores for other clients the two clients that did not satisfy criteria for another personality disorder were considered best represented as having scores consistent with histrionic and schizoid personality disorder.

PBQ subscales, passive-aggressive, histrionic and schizoid, did not have enough clients in the original sample to define a clinical mean score (Beck, et al., 2001). As shown in Table 21 clients were recorded with a primary personality disorder of passive-aggressive, histrionic or schizoid where they would otherwise meet the criteria for another personality disorder and the z-score was higher for passive-aggressive, histrionic or schizoid, with two exceptions; for these two clients (one histrionic and one schizoid) their score was considered to be sufficiently high, relative to the other subscale z-scores for clinical cut-offs to merit recording with a primary personality disorder of histrionic and schizoid respectively; these scores also appeared to fit with clinical observations and client's self-report.

From these results it appeared a relatively high proportion of the PP sample may have had co-morbid personality disorders e.g. circa 80-85%. These initial results suggested the PBQ can be used to check prevalence of personality disorders.

.

6.2 Predictor analysis

6.2.1 Depression (BDI-II) outcome prediction by therapeutic relationship (BLRI)

There were 92 clients with measured depression at first and a subsequent session (BDI-II) plus a measure of the client's perception of the therapeutic relationship (BLRI). The results of multiple regression analysis are shown and these suggested the therapeutic relationship as defined by BLRI scores did not significantly predict the subsequent depression score when controlling for depression score at the first counselling session (BDI-II scores). Whilst there was a statistically significant effect of start depression on subsequent depression ($B_1 = .373$, $p < .001$), the therapeutic relationship did not significantly predict subsequent depression whilst controlling for start depression ($B_2 = -.025$, $p = .539$), see Table 22.

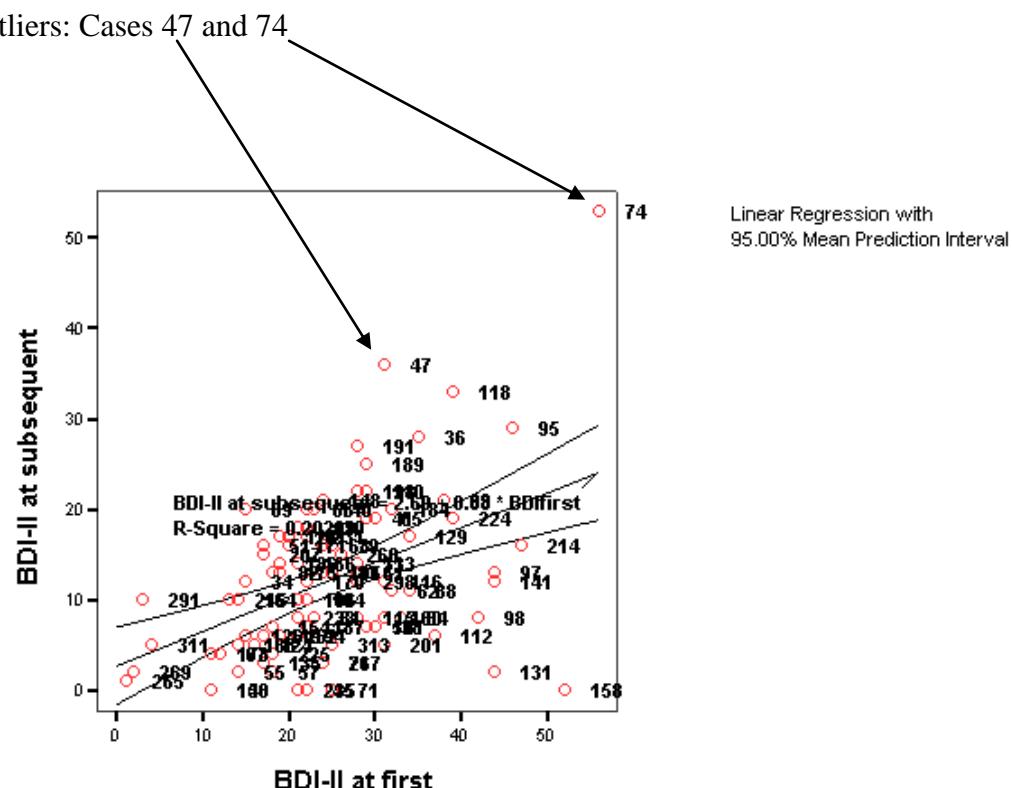
Table 22: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Depression (BDI-II) at First Session in Predicting Depression at Last Session for 92 clients with measurement of depression at first and subsequent session plus BLRI measurement.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
1 (1)	Dep Subseq	Dep First	.373	.081	.439			<.001
(2)		BLRI	-.025	.040	-.059			.539
						.20	11.44	<.001

Note: n = 92. ^a df = 2, 89. DV = dependent variable. Dep Subseq = Depression at Subsequent Session (BDI-II). Dep First = Depression at First Session (BDI-II). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory.

Examination of the regression model revealed that two cases had standardised residuals greater than a value of three, cases 47 and 74 (Figure 11) and the analysis was re-run excluding these cases.

Figure 11: Regression of depression (BDI-II) at first and subsequent session for full sample (n=92): Two outliers with standardised residuals ≥ 3 .



The analysis was re-run omitting these two outlying cases ($n = 90$) and the results shown (Table 23). BLRI scores did not significantly predict subsequent BDI-II scores. Start depression significantly predicted subsequent depression ($B_1 = .250$, $p = .001$) and the therapeutic relationship did not significantly predict subsequent depression whilst controlling for start depression ($B_2 = -.036$, $p = .311$).

Table 23: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Depression (BDI-II) at First Session in Predicting Depression at Subsequent Session for 90 clients with measurement of depression at first and subsequent session plus BLRI measurement.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
2 (1)	Dep Subseq	Dep First	.250	.075	.337			.001
(2)		BLRI	-.036	.035	-.103			.311
						.14	6.81	.002

Note: n = 90. ^a df = 2, 87. DV = dependent variable. Dep Subseq = Depression at Subsequent Session (BDI-II). Dep First = Depression at First Session (BDI-II). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory.

Further results – using the amended model ($n = 90$) and controlling for the number of sessions elapsed until BLRI completed and or the number of sessions elapsed until the subsequent BDI-II measure was completed showed the effect of BLRI scores in predicting the relationship between start and subsequent BDI-II scores was not big enough to be anything other than a chance finding ($p > .05$), see section 7.

6.2.2 Anxiety (BAI) outcome prediction by therapeutic relationship (BLRI)

There were 75 clients with measured anxiety at first and subsequent sessions (BAI) plus a measure of the client's perception of the therapeutic relationship (BLRI). The results of multiple regression analysis are shown and these suggested the therapeutic relationship as defined by BLRI scores did not significantly predict the subsequent anxiety score when controlling for anxiety at the first counselling session. Whilst there was a statistically significant effect of start anxiety on subsequent anxiety ($B_1 = .338$, $p < .001$), the therapeutic relationship did not significantly predict subsequent anxiety whilst controlling for start anxiety ($B_2 = -.004$, $p = .908$), see Table 24.

Table 24: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Anxiety (BAI) at First Session in Predicting Anxiety at Subsequent Session for 75 clients with measurement of anxiety at first and subsequent session plus BLRI measurement.

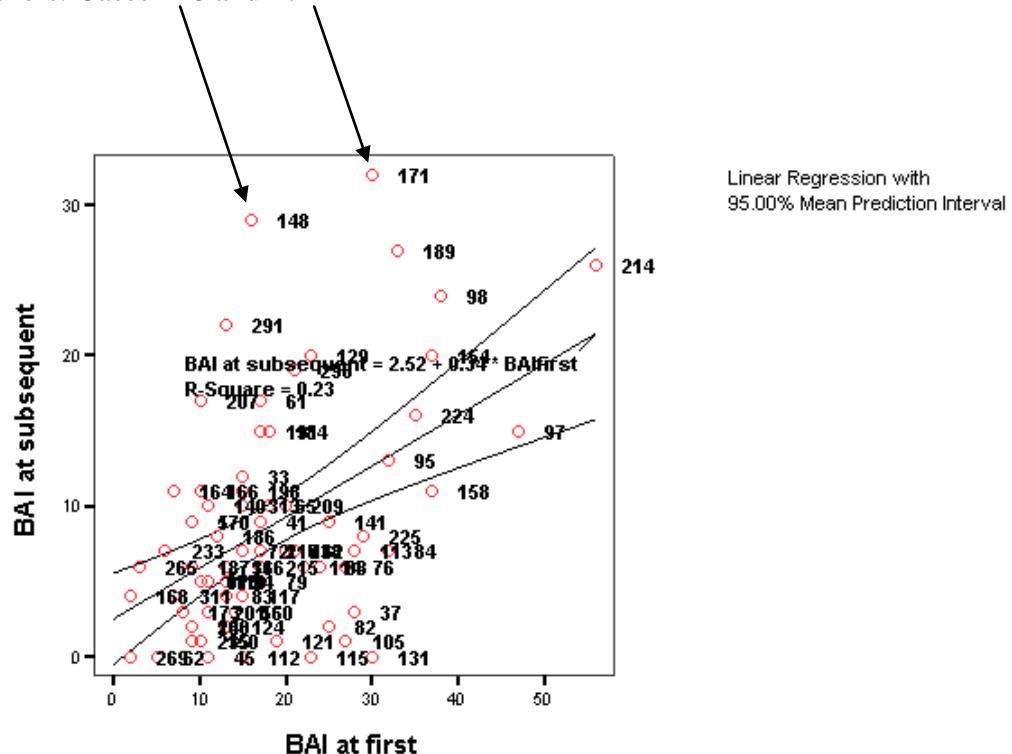
Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
1 (1)	Anx Subseq	BLRI	-.004	.034	-.012			.908
(2)		Anx First	.338	.073	.478			<.001
						.23	10.62	<.001

Note: $n = 75$. ^a For model 1 df = 2, 72. DV = dependent variable. Anx Subseq = Anxiety at Subsequent Session (BAI). Anx First = Anxiety at First Session (BAI). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory.

Examination of the regression models revealed that two cases had standardised residuals greater than a value of three (Figure 12) and the analysis was re-run excluding these two cases.

Figure 12: Regression of anxiety (BAI) at first and subsequent session for full sample (n=75): Two outliers with standardised residuals ≥ 3 .

Outliers: Cases 148 and 171



The analysis was re-run omitting these two outlying cases ($n = 73$) and the results shown (Table 25). BLRI scores did not significantly predict subsequent BAI scores. There was a statistically significant effect of start anxiety on subsequent anxiety ($B_1 = .313, p < .001$) and the therapeutic relationship did not significantly predict subsequent anxiety whilst controlling for start anxiety ($B_2 = -.008, p = .786$).

Table 25: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Anxiety (BAI) at First Session in Predicting Anxiety (BAI) at Subsequent Session for 73 clients with measurement of anxiety at first and subsequent session plus BLRI measurement.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
2 (1)	Anx Subseq	BLRI	.008	.029	.028			.786
(2)		Anx First	.314	.064	.506			<.001

Note: $n = 73$. ^a For model df = 2, 70. DV = dependent variable. Anx Subseq = Anxiety at Subsequent Session (BAI). Anx First = Anxiety at First Session (BAI). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory.

Further results – using the amended model ($n = 73$) and controlling for the number of sessions elapsed until BLRI completed and or the number of sessions elapsed until the subsequent BAI measure was completed showed the effect of BLRI scores in predicting subsequent BAI scores was not big enough to be anything other than a chance finding ($p > .05$).

6.2.3 Distress (CORE-OM) prediction by therapeutic relationship (BLRI)

There were 54 clients with measured symptoms of distress at first and subsequent sessions (CORE-OM) plus a measure of the client's perception of the therapeutic relationship (BLRI). The results of multiple regression analysis are shown in Table 26 and these suggested the therapeutic relationship as defined by BLRI scores (Rogers, 1957) did not significantly predict the subsequent distress (CORE-OM scores). Whilst there was a significant total effect of start distress on subsequent distress ($B_1 = .406$, $p < .001$) the therapeutic relationship did not significantly predict subsequent distress whilst controlling for start distress ($B_2 = -.004$, $p = .897$), see Table 26

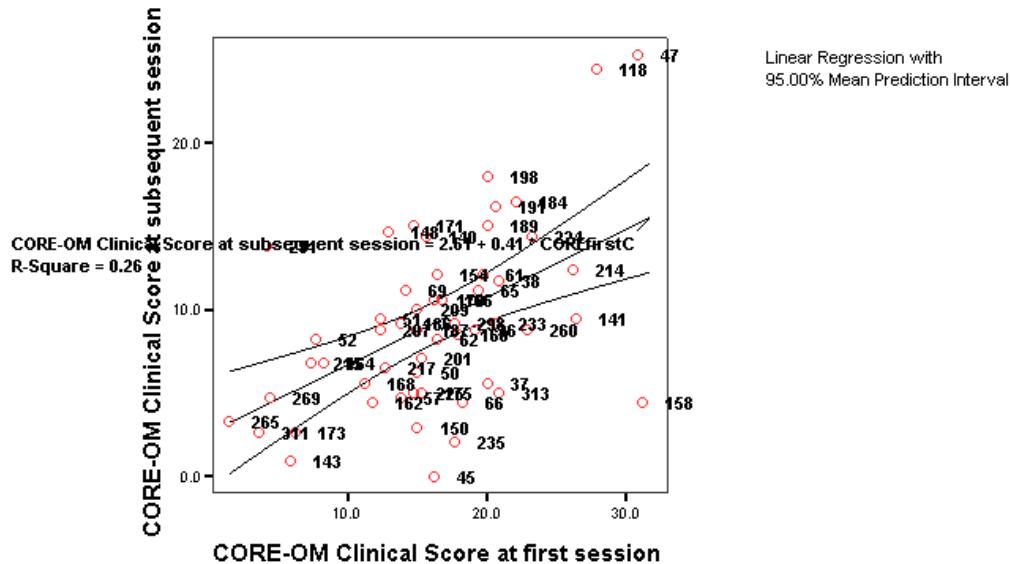
Table 26: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Distress (CORE-OM) at First Session in Predicting Distress (CORE-OM) at Subsequent Session for 54 clients with measurement of distress at first and subsequent session plus BLRI measurement.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
1 (1)	CORE-OM Subseq	BLRI	-.004	.034	-.016			.897
(2)		CORE-OM First	.402	.102	.501			<.001
						.26	8.79	.001

Note: $n = 54$. ^a df = 2, 51. DV = dependent variable. CORE-OM Subseq = Distress at Subsequent Session (CORE-OM). CORE-OM First = Distress at First Session (CORE-OM). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory.

No cases had standardised residuals greater than a value of three, see Figure 13.

Figure 13: Regression of anxiety at first and subsequent session for full sample (n= 54): No outliers with standardised residuals ≥ 3 .



Further results – using this model ($n = 54$) and controlling for the number of sessions elapsed until BLRI completed and or the number of sessions elapsed until the subsequent CORE-OM measure was completed showed the effect of BLRI scores in mediating the relationship between start and subsequent CORE-OM scores was not big enough to be anything other than a chance finding ($p > .05$).

7. Further results

The results of the hypotheses tests were presented together with some additional results (section 6) to fill out the wider picture of the outcome and predictor hypotheses. In this section some further results are examined in particular to address some of the criticisms of uncontrolled naturalistic research and to question to what extent the outcome results can be relied upon. Furthermore some questions about the non-significant findings for the predictor hypotheses are examined, in particular what, if any, were the shortcomings of this part of the research and what might have been their impact. Given the 'exploratory' nature of these results, sample characteristics are not provided for all the groups analysed, although they are for the wait-controlled outcomes part of the research (Appendix 3).

7.1 Outcomes research: Addressing some of the criticisms of uncontrolled naturalistic research

Non-randomised naturalistic outcome research has appeared in the literature e.g. Stiles, et al., (2006), Stiles W, et al., (2007). Criticisms of this approach (Clark, et al., 2007) and their rebuttal (Stiles, et al., 2008) appeared after much of the data collection for this research was completed. Nevertheless it was possible to address some of the areas of concern for this type of research with the data that had been collected. This section examines expressed 'areas of concern' (Clark, et al., 2007) to the extent that it was possible with the data collected, specifically:

- 'Missing cases' – how come some clients were 'lost to follow-up' and what impact did this have on the effect size and 'reliable and clinically significant improvement' data reported?
- 'No control for other causes of recovery' – was any observed change simply 'regression to the mean', or 'natural recovery in recent onset cases' or perhaps 'attributable to concurrently administered medication'?
- 'Lack of randomisation to different treatments'
- 'No evidence that treatments appropriately delivered'

7.1.1 Missing cases: Depression

What follows is a qualitative analysis of quantitative data.

On an LOCF basis there were 162 clients whose mean BDI-II scores improved significantly $F(1, 161) = 127.74, p < .001$, $ES(d) = 1.02$ (95% CI .85 to 1.20). Of the 162 clients who started therapy with a clinical depression score 111 (68.5% of 162) had a subsequent BDI-II measurement; 51 clients were ‘missing cases’. In this section the ‘missing cases’ at PP and UCS are looked at separately as there were differences between the two sites and then the overall impact of these missing cases is assessed.

a) Private Practice (PP) clients

There were 19 ‘missing cases’ at the PP, 20.2% of the 94 with clinical level depression who started therapy. In their rejoinder to Clark, et al. (2007), Stiles, et al. (2008) made the point that in contrast to randomised controlled trials where clients have been assessed, selected and assigned to groups, ‘non-completion in routine practice is more often attributable to personal, institutional, social and economic conditions than to the theoretical approach the therapist uses’ (p. 2). There appeared to be some validity in this comment with regard to the PP where apparent reasons for non-completion were individually assessed and grouped as follows:

- 6 clients stopped attending because they wanted to address a relationship issue and it was clear from discussions with their partner, that change was not possible in the way they envisaged, e.g. ‘I’m no longer coming as I’ve decided to leave the relationship’.
- 4 clients stopped attending because they decided they didn’t want to address the issue they initially said they wanted to address e.g. child sexual abuse.
- 2 clients were prevented from attending because it turned out they were closely related to clients previously seen by the author.
- 2 clients stopped attending because they decided it was too far to travel (1 was sent by their employer).
- 2 clients were sent by their employer and attended only one session. It is not known why they stopped although it is possible that they felt some coercion to at least attend one session.
- 1 client phoned to say one session had been sufficient and he no longer needed to attend.

- 1 client decided to work with a different therapist. Given the research findings about the importance of the therapeutic relationship (Norcross, 2002) the author is keen to encourage clients to ‘try me out’ and be open minded to clients choosing to work with someone else if they so choose.
- 1 client completed the outcome questionnaires at the start of therapy and then made it clear that they were attending for a one-off session paid for by a relative to address a particular issue that had occurred.

Clearly this kind of analysis risks post-rationalisation error. The numbers of sessions attended by those PP clients who were ‘missing cases’ are shown in Table 27.

Table 27: Numbers of sessions attended by 19 PP ‘missing cases’.

Number of sessions	Number of clients	Cumulative %
1	15	78.9%
2	1	84.2%
3	3	100.0%
Total	19	

b) University Counselling Service (UCS) clients

At the UCS there were 32 ‘missing cases’, 47.1% of 68, who began therapy with a clinical level of depression. Less is known about the UCS clients, because they were seen by therapists in addition to the author. However, there were differences between the two settings that likely go someway to explaining the difference between the PP where 20.2% are ‘missing’ and UCS where 47.1% are missing. Clearly the author was motivated to encourage questionnaire completion! At the UCS there were some therapists who were ‘for’ the research, some ‘against’ and some who changed their allegiance e.g. one of the ‘missing cases’ had 20 sessions to what was recorded by the therapist, who had agreed to participate in the research, as a ‘mutually agreed and satisfactory ending’ although no last session questionnaires were completed. As the research progressed the author was using an ‘outcomes management’ approach by asking clients to complete outcome questionnaires on a regular basis (none of the ‘missing cases’ at the PP had more than three sessions) whereas at the UCS the protocol was for clients to complete their outcome questionnaire only at the last

session. The numbers of sessions attended by those UCS clients who were ‘missing cases’ are shown in Table 28.

Table 28: Numbers of sessions attended by 32 UCS ‘missing cases’.

Number of sessions	Number of clients	Cumulative %
1	6	18.8%
2	1	21.9%
3	5	37.5%
4	6	56.3%
5	2	62.5%
6	2	68.8%
7	1	71.9%
8	1	75.0%
11	1	78.1%
15	1	81.3%
18	1	84.4%
20	1	87.5%
22	1	90.6%
23	1	93.8%
26	1	96.9%
31	1	100.0%
Total	32	

It seemed plausible that some of the cases that were ‘missing’ from the UCS probably had perfectly good outcomes that simply went unrecorded, for whatever reason, e.g. therapist did not ask for questionnaires to be completed, client stopped therapy earlier than therapist expected, etc. Equally there could be some ‘treatment failures’ amongst this group. Clients at the UCS were not immune from ‘political’ reasons to attend for therapy e.g. amongst this group of 32 clients were two clients who attended because they had been referred by the Dean of Students Office, two who had been referred by advisers/lecturers/academic staff, two who had been sent by the University Health Centre and one who had been told to go to the UCS by ‘friends’. In terms of recorded endings for this group of 32 clients; 9 were marked as ‘premature ending (client or counsellor leaving), 5 were marked as ‘mutually agreed and satisfactory ending’ and 18 simply as ‘Other (including DNA)’.

After assessing each individual case of the 32 UCS 'missing cases' with the evidence available the author decided that there were 14 clients that might be considered 'treatment failures'; six clients who attended only one session, who had decided they did not intend to be treated; eleven clients who had an ending marked by the therapist as either a 'mutually agreed and satisfactory' or 'premature – client or counsellor leaving'; plus one client who attended for 31 sessions and then did not attend, on the basis that it seemed unlikely someone would attend for 31 sessions and then conclude this was wholly a 'treatment failure'.

Clearly this kind of analysis risks post-rationalisation error. The numbers of sessions attended by those UCS clients who might be considered 'treatment failures' is shown in Table 29.

Table 29: Number of sessions attended by UCS clients who might be considered 'treatment failures'.

Number of sessions ^a	Number of clients	Cumulative %
3	4	28.6%
4	4	57.1%
5	2	71.4%
6	2	85.7%
7	1	92.9%
8	1	100.0%
Total	14	

c) Private Practice (PP) and University Counselling Service (UCS) clients

Scientifically the reason for examining 'missing cases' is to determine to what extent claims for 'treatment successes' should be tempered with an allowance for 'treatment failures'. In this case the extent to which the effect size for clients with a subsequent measure of their depression, $n = 111$, $ES(d) = 1.48$ should be reduced to allow for any 'treatment failures' e.g. $ES(d) = 1.02$ for $n = 162$ on an LOCF basis. This so called 'intent to treat' basis in a naturalistic setting does not allow for those clients who 'do not intend to be treated' e.g. 'I am only attending because my wife/ husband/ employer/ lecturer made me'.

One approach to estimating the impact of 'missing cases' is to group cases together and then to estimate the impact on the overall effect size from that group. For example the following steps could be taken to analyse this, see Table 30 and Figure 14:

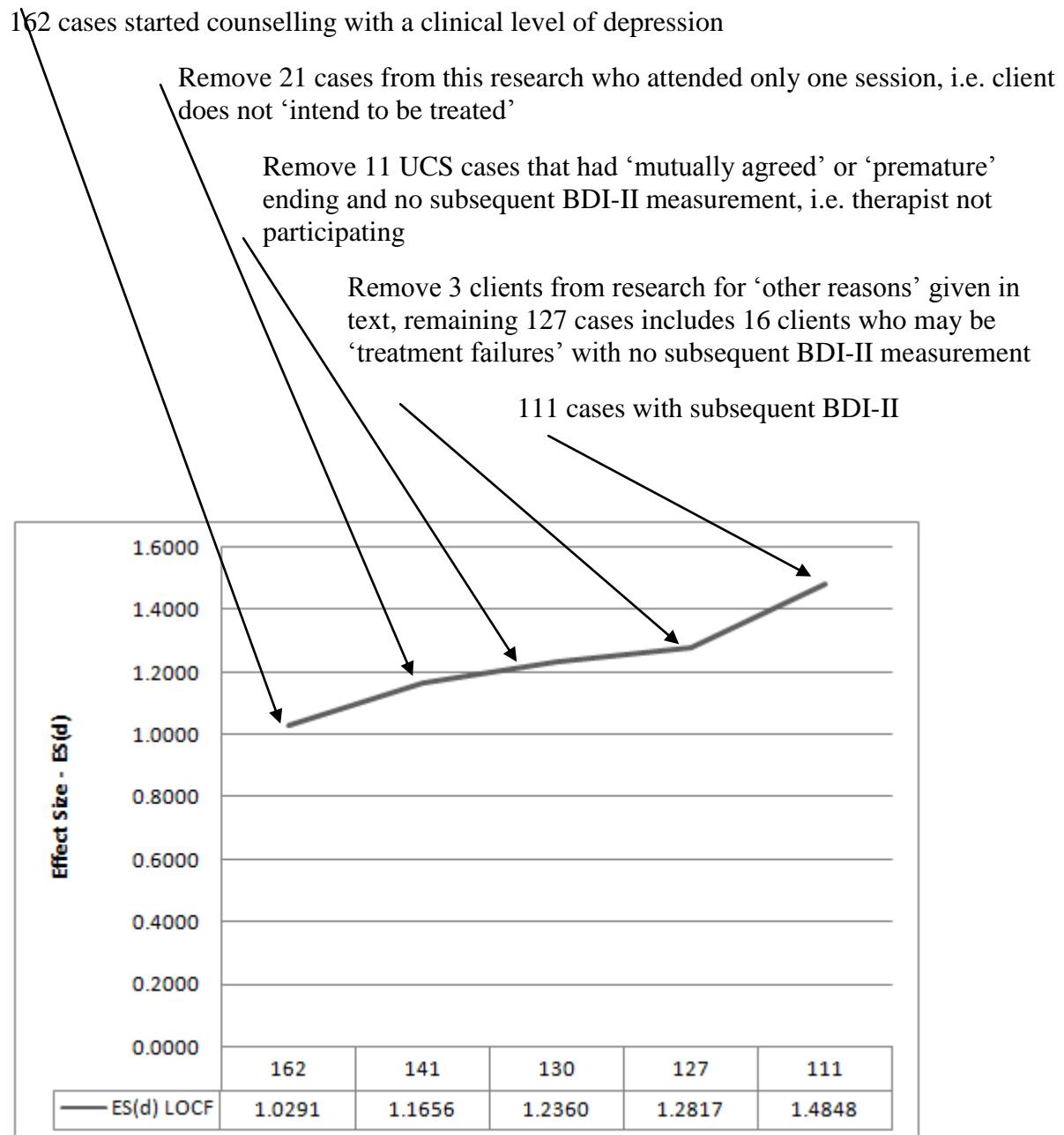
1. 162 clients start therapy with clinical depression and on an LOCF basis, mean BDI-II scores improved significantly $F(1, 161) = 127.74$, $p < .001$, $ES(d) = 1.0291$ (95% CI .85 to 1.20).
2. There were 21 'missing clients' that attended only one session ($162 - 21 = 141$ clients); arguably these clients hadn't really 'started'. Subtracting these 21 cases from the 162 clients at first session with a clinical BDI-II score left 141 cases that 'really start'. On an LOCF basis these 141 clients change significantly $F(1, 140) = 147.30$, $p < .001$, $ES(d) = 1.1656$ (95% CI .9752 to 1.3547).
3. Once the clients who had only one session were removed there were 11 UCS clients who were recorded as having a 'mutually agreed and satisfactory' or 'premature – client or counsellor leaving' ending. There was some likelihood that subsequent BDI-II questionnaires were not completed for these 11 clients for reasons other than 'treatment failure', e.g. therapist didn't ask for questionnaire to be completed, client or therapist left university. On an LOCF basis the remaining 130 clients change significantly $F(1, 129) = 161.60$, $p < .001$, $ES(d) = 1.2360$ (95% CI 1.0438 to 1.4286).
4. Beyond this a further three clients could be removed from the analysis: one client had 31 sessions at the UCS and it seems unlikely that a person would attend 31 sessions and consider this wholly a 'treatment failure', one client had 3 sessions at the PP and decided to stop attending because they no longer wanted to be in their couple relationship (communication received after session 3), one client had 2 sessions at the PP before it became clear that they were in a close relationship with a former client. This left 127 clients who, on an LOCF basis, had a statistically significant improvement $F(1, 126) = 164.80$, $p < .001$, $ES(d) = 1.2817$ (95% CI 1.0849 to 1.4803). Compared with the 111 clients who had a subsequent BDI-II measure this sample of 127 LOCF clients thus included 16 clients who might be termed 'treatment failures': one client at the PP who after 3 sessions decided they would rather not at this stage examine their childhood sexual abuse, one client at the PP who after 3 sessions decided they would ignore the advice of a former client about attending counselling, and 14 clients at the UCS who attended between 3 and 8 sessions with no subsequent BDI-II measurement and an ending recorded as 'Other (including DNA)', see Table 30; the extent to which these clients would consider themselves 'treatment failures' is not known.

5. There were 111 clients who completed a subsequent BDI-II (68.5% of 162 clients) and their mean BDI-II scores improved significantly $F(1, 110) = 212.60, p < .001$, $ES(d) = 1.4848$ (95% CI 1.28 to 1.68). It was this group that were the target of the depression outcomes hypothesis (A.1).

Table 30: Groups of clients who could be included or excluded from outcome measurement and estimated impact on effect size and reliable change percentage.

Reason for inclusion/exclusion	Δn	$\Delta ES(d)$	$\Delta \% \text{ reliable change}$
Attend only one session	21	.1365	7.2%
Probably did improve according to another measure	11	.0704	4.7%
Reason for stopping probably not related to lack of progress	3	.0457	1.4%
Possible 'treatment failures' although no direct evidence of this	16	.2031	8.9%
Total	51	.4557	22.2%

Figure 14: Hypothetical impact of ‘missing cases’ on effect size: removal of cases from the research for reasons given in text (x-axis is descending number of cases left ‘in the research’ once these cases are removed and y-axis is mean effect size on an LOCF basis).



Of the 111 clients who started therapy with a clinical level of depression and had a subsequent measure of their depression, 73 had 'reliable change', of which 17 were 'improved' and 56 'recovered'. These 73 clients can be set against different denominators as described above (see also Table 30) and illustrated in Figure 14 to give differing percentages with 'reliable change' as shown in Table 31, varying from 34.6% 'reliable and clinically significant change' to 50.5%, depending upon the viewpoint taken.

Table 31: For different denominators: percentages of clients with reliable change and recovered from depression.

Group	n	Reliable Deterioration ^a	No Reliable Change ^b	Reliable Change ^c		
				I or R ^d	Improved ^e	Recovered ^f
All _{LOCF}	162	0.0%	51.9%	48.1%	11.7%	36.4%
Less sessions ≥ 2	141	0.0%	44.7%	55.3%	13.5%	41.8%
Less non-DNA endings	130	0.0%	40.0%	60.0%	14.6%	45.4%
Less 3 other clients (text)	127	0.0%	38.6%	61.4%	15.0%	46.4%
All _{Subseq}	111	0.0%	29.7%	70.3%	17.1%	53.2%

Note: ^a 'Reliable deterioration' was a deterioration of more than 9.4 BDI-II units. ^b 'No reliable change' was a change of less than 7.8 BDI-II units. ^c 'Reliable change' was an improvement of more than 7.8 BDI-II units. ^d 'I or R' was 'improved' or 'recovered' at subsequent measurement, i.e. 'improved' (improvement of more than 7.8 BDI-II units) or 'recovered' (an improvement of more than 7.8 BDI-II units and a subsequent score of 13 or less). ^e 'Improved' was reliable change only, i.e. improvement of more than 7.8 BDI-II units and subsequent score was ≥ 14 . ^f 'Recovered' was reliable change and non-clinical score at subsequent, i.e. improvement of more than 7.8 BDI-II units and a subsequent score of 13 or less.

For depression outcomes this section has considered 'missing cases' and examined how some clients were 'lost to follow-up' and assessed what impact this may have had on the effect size and 'reliable and clinically significant improvement' data reported. This exercise was not repeated for anxiety outcomes as these clients were mainly a subset of the clients in the depression outcomes grouping, BDI-II was used in preference to BAI as a first line measure of progress (Newman, et al., 2006, p. 193). The exercise was not repeated for distress outcomes as these were a smaller and temporally earlier group in this research and they were mainly UCS clients, already described in this section, who had also completed BDI-II.

7.1.2 Controlling for other causes of recovery

a) Depression

Since this was naturalistic research, rather than a randomised controlled trial, it was not possible to know what average changes might have occurred without the therapeutic intervention. For the UCS part of the research it was possible to structure a 'wait' condition, with which to compare changes made during treatment. The question of whether being in therapy was responsible for the observed improvements, was addressed by seven analyses:

1. By estimating the effect of regression to the mean during treatment.
2. By looking at a subset of clients who acted as their own wait list control to see if they changed during wait or treatment.
3. Because on average clients were in treatment longer than waiting, the effect of time on change in depression was assessed.
4. By comparing the subset of clients who acted as their 'own wait list control' with the rest of clients who were in a 'wait condition only' to see if there was any significant difference between these two groups, in terms of depression scores.
5. By comparing the subset of clients who acted as 'own wait list control' with the rest of clients who were in a 'treatment condition only' to see if there was any significant difference between these two groups, in terms of depression scores.
6. By comparing the clients who started therapy with a clinical depression score and didn't do a subsequent measurement of their depression with those who did, to see if there was any significant difference between these two groups, in terms of depression scores. The rationale for this was to see if there was any evidence that it was the 'easier' (lower depression) cases that were completing and the 'harder' (higher depression) cases that were terminating early.
7. By comparing clients in therapy who said they were taking no prescription medication with those who said they were, to see if there was any significant difference between these two groups in terms of depression scores.

1. Were any observed changes simply regression to the mean?

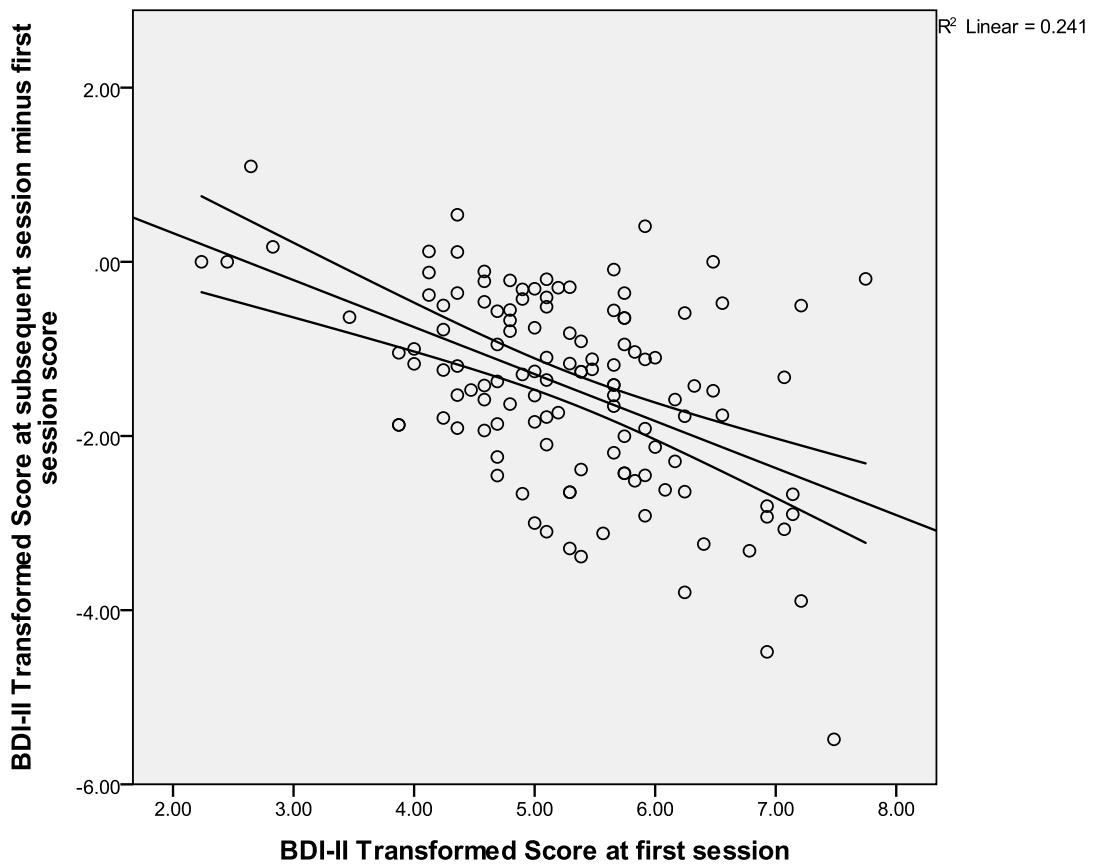
Correction for regression to the mean effects requires data to be distributed not significantly different from normal. BDI-II scores at the first time of completion were distributed significantly different from normal and data was transformed to a distribution not significantly different from normal by adding four to each score and taking the square root of the resultant sum for this part of the analysis. References to BDI-II scores in this section are to the transformed scores.

During treatment, 124 clients had their BDI-II scores recorded at their first and subsequent counselling session. The mean score for this group was 5.29, with a standard deviation of 1.02, the population was distributed not significantly different from normal K-S D(124) = .05, $p > .2$.

Hypothesis A1 was about clients with 'clinical depression', raw scores of 14 and above, and so this cut-off was applied to derive the 'clinical sample' of 111 clients (those with a raw BDI-II score of 14 or above at their first session), from the sample of 124 clients. These 111 clients with 'clinical depression' had a mean score of 5.49 at the first session, SD .83, and a mean score of 3.94 at their subsequent measurement, SD 1.07. To what extent this observed change of 1.55 in mean BDI-II scores, $ES = 1.8713$, for the subsample starting with 'clinical distress', was simply a regression to the mean effect is considered below using the method described by Barnett, van der Pols, and Dobson (2005).

BDI-II score, at first session versus subsequent counselling session minus score at first session is plotted below, Figure 15.

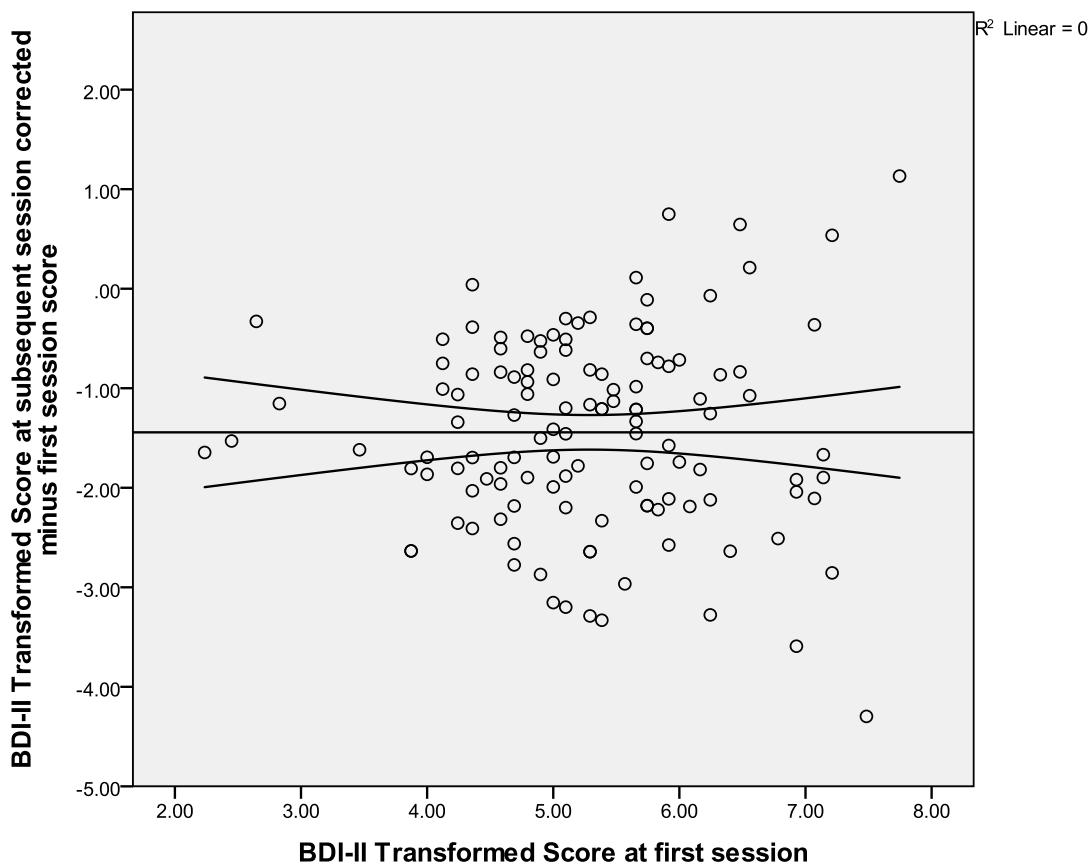
Figure 15: Scatter plot of transformed BDI-II at first session versus subsequent counselling session minus score at first session.



There was some evidence of a regression to the mean effect, in that some clients whose BDI-II scores were low (below the population mean) at the first session had increased at the subsequent session and some of those with higher scores (above the population mean) had reduced at the subsequent session.

Regression to the mean effects were corrected for using the method described above with $B = .460$, $p < .001$, see Figure 16.

Figure 16: Scatter plot of regression to mean corrected transformed BDI-II at subsequent session minus first session score versus first session score.



For the 111 clients with clinical scores (raw BDI-II ≥ 14) at first session, the mean score at subsequent session corrected for regression to the mean effects was 4.05, SD 1.30. Using repeated measures analysis of variance there was a significant improvement from first to subsequent for both the original data $F(1, 110) = 221.93, p < .001$ and the corrected data $F(1, 110) = 228.97, p < .001$. Correcting for regression to the mean reduced the size of the change between first and subsequent session to 1.44, $ES(d) 1.7359$; suggesting .11 transformed BDI-II units, $ES(d) .1354$ or 7.2% of the effect was due to regression to the mean.

Comparing the mean change without correction for regression to the mean, 1.55, SD 1.10, with the mean change with correction for regression to the mean, 1.44, SD 1.00, with a paired samples t-test was statistically significant $t(110) = 2.64, p = .009, r = .24$. This was a small-medium effect, mean estimate 0.11 transformed BDI-II units, 95% confidence interval 0.028 to 0.196. Converting back to BDI-II units was not straightforward in the sense that the transformed figure represented the difference between two square root expressions. One

solution that fits this expression is that the mean estimate for the regression to the mean effect during treatment was approximately equal to the difference between BDI-II scores of 15 and 14, i.e. $\sqrt{15+4} - \sqrt{14+4} = .116$. For sake of simplicity on average it could be considered that the size of the regression to the mean effect during treatment was approximately one BDI-II unit. Whilst change from first to subsequent session was much greater than could be explained by regression to the mean effects, regression to the mean was a statistically significant small-medium effect.

2. *Did clients change during wait or treatment condition?*

At the UCS there were 36 clients who had BDI-II scores at exploratory, first session and a subsequent session and had a clinical depression score (BDI-II ≥ 14) at first session (Sample characteristics, Appendix 3). The results of a repeated measures ANOVA showed (Table 32) mean BDI-II scores did not improve significantly during 'wait', from exploratory session to first session, $F(1, 33) = 1.28$, $p = .266$, $ES(d) = .05$ (95% CI -.18 to .28) and did improve significantly during 'treatment', from first session to a subsequent session, $F(1, 33) = 9.10$, $p = .005$, $ES(d) = 1.26$ (95% CI .82 to 1.69). Given the non-significant change during wait this suggested the wait-controlled effect size for these 36 UCS clients was $ES(d) = 1.26$. This suggested mean client BDI-II scores did not improve during 'wait', did improve during 'treatment' and therefore improved more during treatment than during wait. The benchmarks for depression (Table 9, Minami, et al., 2007) suggested a small/medium improvement of $ES(d) = .371$ over a period of 15-16 weeks and another meta-analysis (Elliott, et al., 2004) found $ES(d) = .11$ for untreated control clients. These findings were similar to those found by Svartberg et al. (2001) who found 'patients with major depression did not change after the pretherapy interviews, whereas anxiety patients changed substantially and significantly faster than the depressed patients' (p. 201. The analysis from this research suggested it was the treatment condition that was responsible for the observed improvements. A number of other explanations are possible e.g. only clients who were still depressed returned for treatment or very depressed clients sought help elsewhere (see analyses 4 and 6 below) or there was a difference in the length of the wait and treatment periods. In fact, clients did wait a significantly shorter number of days (Median 21 days) than they were in treatment for (Median 49 days) $z = 3.56$, $p < .001$, a medium sized effect, $r = .34$. The effect of time on change in depression was therefore examined in the next section (below).

Table 32: Depression (BDI-II) outcomes: Comparing ‘wait’ with ‘treatment’.

Group	n	Start		Subsequent		ES(d) ^a	p
		Mean	SD	Mean	SD		
All clients with clinical BDI-II score at first session (BDI-II ≥ 14):							
Exp to First	36	26.19	9.908	25.69	9.017	.0504	.266
First to Subsequent	36	25.69	9.017	13.19	8.031	1.2616	.005

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at exploratory session.

3. What effect did time have on change in depression?

Including both wait period in days and treatment period in days as covariates (repeated measures ANCOVA) suggested there was no significant interaction between wait period (days) and change in depression score during wait, from exploratory session to first session, $F(1, 33) = .45$, $p = .508$, however there was a significant interaction between treatment period (days) and change in depression score during treatment, from first therapy session to a subsequent measure, $F(1, 33) = 5.00$, $p = .032$. This suggested time in treatment interacted with change in depression score (longer treatment, lower depression) and there was insufficient evidence to suggest time in wait interacted with change in depression.

Whilst there was insufficient evidence of a change during wait, examination of the interaction between wait (days) and change in depression (BDI-II) during the wait period (Table 33) suggested there was no statistically significant improvement following the exploratory session and there was no significant change predicted by the wait period.

Table 33: Summary of Regression Analysis for Wait period (days) in predicting change in Depression (BDI-II) during wait: Model 1 has no outlying cases (standardised residuals ≥ 3).

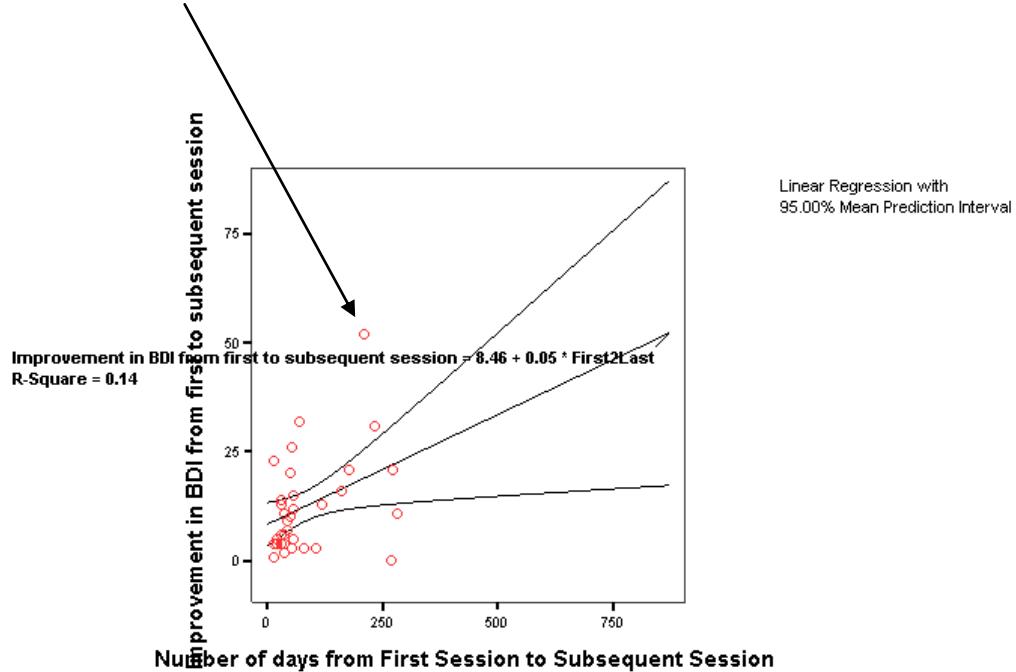
Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	P
1	Imp Dep Exp to First	Constant	1.185	1.453				.420
		Wait period (days)	-.025	.042	-.103			.549
						.01	.37	.549

Note: n = 36. ^a For model 1 df = 1, 34. DV = dependent variable. Imp Dep Exp to First = Improvement in Depression (BDI-II) from Exploratory Session to First Session.

Further examination of the interaction between treatment period (days) and change in depression during treatment with regression analysis suggested treatment period (days) significantly predicted change in depression ($B = .050$, $p = .026$), see Model 1 in Table 34. Examination of residuals showed one case was an outlier (standardised residual ≥ 3), see Figure 17.

Figure 17: Scatter plot of treatment period (number of days from first session to subsequent session) versus improvement in depression (BDI-II) during treatment period: One outlier with standardised residual ≥ 3 .

Outlier: Case 158



Note: A positive value for an improvement in depression (BDI-II) was a lessening of depression.

Once the outlying case was deleted from the model, treatment period (days) no longer significantly predicted change in depression during treatment ($B = .029$, $p = .128$), see Model 2 in Table 34. Note the constant in both models was statistically significant, suggesting that on average immediately following the first session (Analysis 1) clients improved by 8.5 BDI-II units (95% CI 3.5 to 13.4) or (Analysis 2) clients improved by 9.2 BDI-II units (95% 5.1 to 13.2). The mean changes were greater than the reliable change criteria calculated for BDI-II. There is a reasonably extensive literature on 'sudden gains', especially for CBT (e.g. Tang &

DeRubeis, 1999, Tang, Luborsky, & Andrusyna, 2002, Tang, DeRubeis, Beberman & Pham, 2005, Tang, DeRubeis, Hollon, Amsterdam & Shelton, 2007).

In summary, neither wait (days) nor treatment period (days) predicted change in depression score (BDI-II). This analysis suggested on average the passage of time did not have a predictable effect on depression levels. The extent to which these 36 clients were representative, in terms of their depression scores, of the other clients in this research was addressed in the next section (below).

Table 34: Summary of Regression Analysis for Treatment period (days) in predicting change in Depression during treatment: Model 1 with outlying case included (standardised residual ≥ 3) and Model 2 with outlying case deleted from the model.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
1	Imp Dep	Constant	8.457	2.428				.001
		Treatment period (days)	.050	.022	.370			.026
	First to Subseq					.37	5.40	.026
2	Imp Dep	Constant	9.157	1.999				<.001
		Treatment period (days)	.029	.018	.262			.128
	First to Subseq					.26	2.44	.128

Note: n = 36 for model 1 and n = 35 for model 2. ^a For model 1 df = 1, 34, for model 2 df = 1, 33. DV = dependent variable. Imp Dep First to Subseq = Improvement in Depression (BDI-II) from First Session to Subsequent Session.

4. How representative were the clients in the 'wait time control of treatment outcomes' of those other clients who waited for therapy? And (5.) how representative were they of those other clients who had a subsequent measure of their depression?

'In psychotherapy research, the best predictor by far of post-treatment scores is usually pre-treatment scores on the same measure' (Stiles, et al., 2008, p. 4) and on that basis the extent to which the 36 clients in the wait time control of treatment outcomes were representative of the larger samples, was assessed in terms of depression scores at the different stages and was addressed in two ways as shown below:

- Firstly, by comparing the 36 UCS clients with both wait and treatment outcomes with the 32 UCS clients who had data only for the wait condition. A repeated measures analysis of variance with group membership as a between-subjects factors showed scores for both groups combined did not change significantly during wait $F(1, 66) = .10$, $p = .752$ and there was no statistically significant difference between the two group means at exploratory and first session combined, $F(1, 66) = .97$, $p = .329$, see Figure 18.
- Secondly, by comparing the 36 UCS clients with both wait and treatment outcomes with the 75 PP clients who had data only for the treatment condition. A repeated measures analysis of variance with group membership as a between-subjects factor showed scores for both groups combined changed significantly during treatment $F(1, 109) = 175.75$, $p < .001$, ES (d) = 1.48 (95% CI 1.22 to 1.65). (Note: This was the depression hypothesis sample, A.1, the F ratio and 95% CI for effect size differ slightly from those shown above as this is a different statistical test). There was no statistically significant difference between the two group means at first and last session combined, $F (1, 109) = .09$, $p = .765$, see Figure 19.

Figure 18: Estimated marginal means at exploratory session (stage 1) and first session (stage 2) for 36 UCS clients in the ‘wait and treatment’ condition and 32 UCS clients in the ‘wait only’ condition: stage (1 = exploratory session, 2 = first session) is x-axis and depression (BDI-II) score is y-axis.

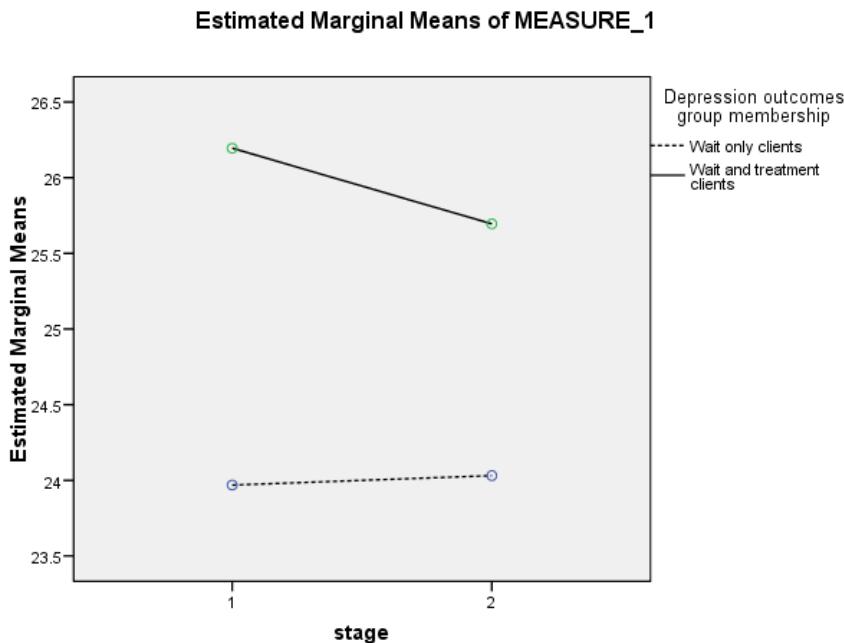
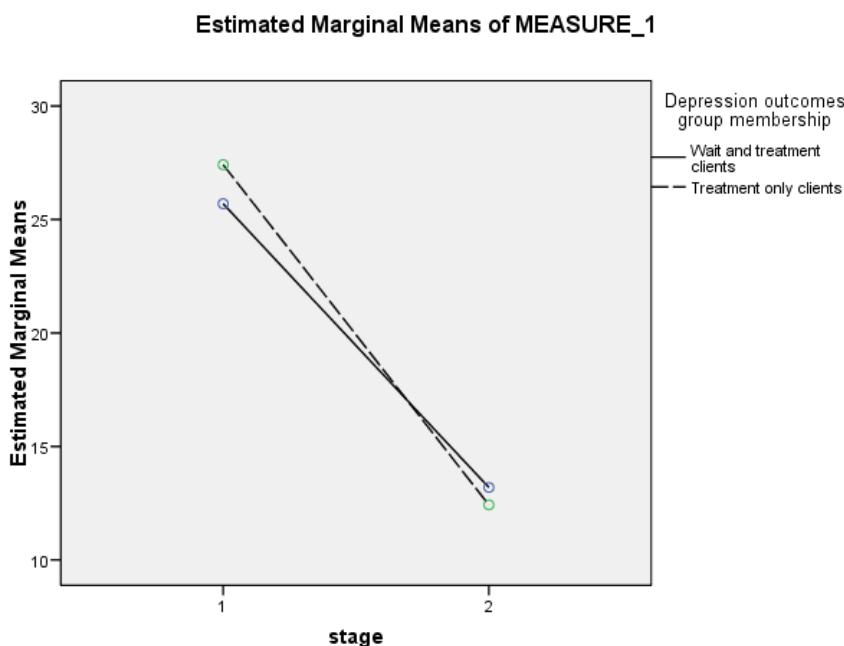


Figure 19: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 36 UCS clients in the ‘wait and treatment’ condition and 75 PP clients in the ‘treatment only’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and depression (BDI-II) score is y-axis.

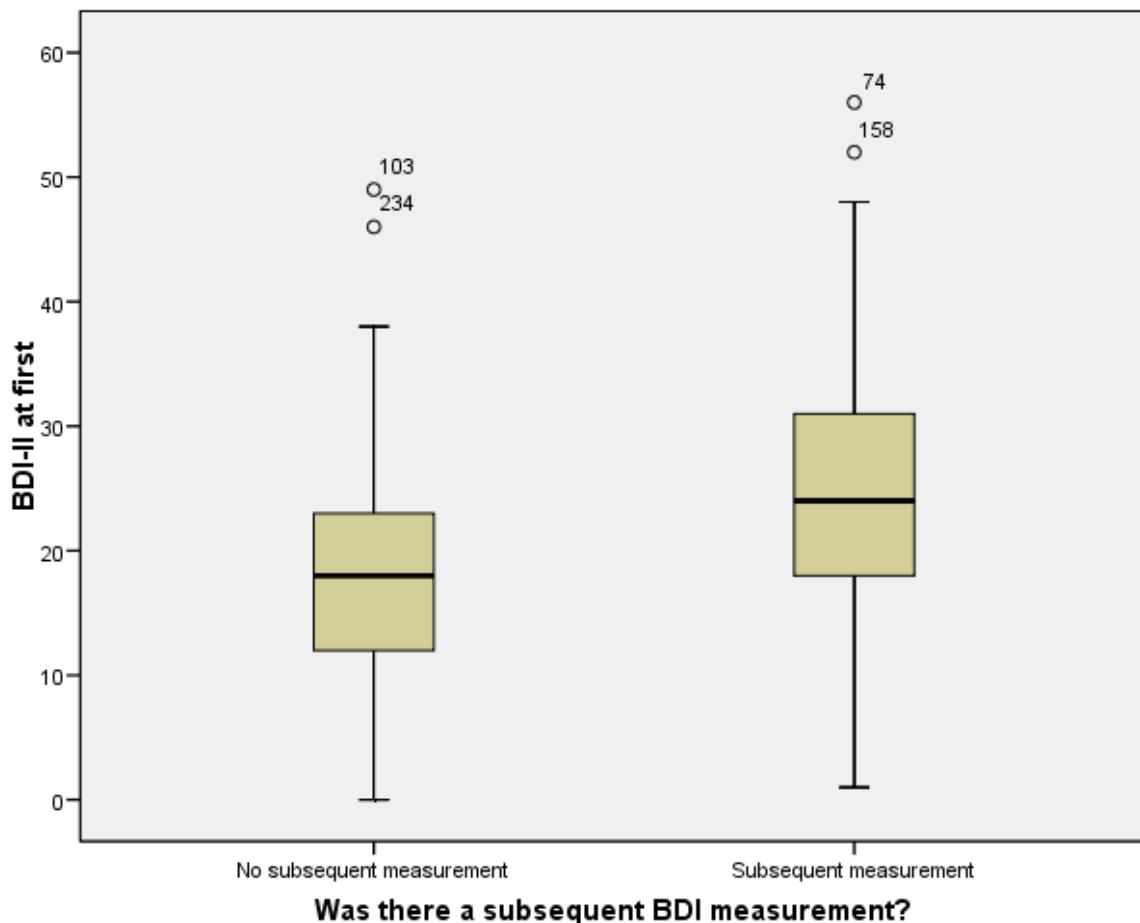


The 36 UCS clients with both wait and treatment outcomes were representative of the 32 UCS clients with data for wait only and the 75 PP clients with data for treatment only. To what extent clients who started in therapy with clinical depression and had a subsequent measure of their depression were representative of those who didn't do a subsequent measure of their depression was assessed, in the next section.

6. How representative, in terms of depression scores, were those clients who started therapy with a clinical depression score and didn't do a subsequent measurement of their depression, with those who did?

The 205 clients with a BDI-II score at first session were examined for this part of the analysis. The 124 clients with a subsequent BDI-II score had a mean score at first session (mean = 24.96, SD = 10.70) higher than the 81 clients with no subsequent BDI-II score (mean = 18.00, SD = 9.991) and this difference was significant $t(203) = 4.67$, $p < .001$, a medium effect $r = .31$, see Figure 20. On average it was the clients who started off more depressed who were more likely to have their progress measured. There was no evidence that it was the 'easier' (lower depression) cases that were continuing and the 'harder' (higher depression) cases that were terminating early. In fact the evidence pointed in the opposite direction; clients who started off more depressed were more likely to continue to a subsequent measurement.

Figure 20: Box plot of depression (BDI-II) scores at first session (y-axis) for clients with no subsequent BDI-II measurement (81 clients) and a subsequent BDI-II measurement (124 clients): Clients with a subsequent measurement started therapy with a significantly higher level of depression $p < .001$.

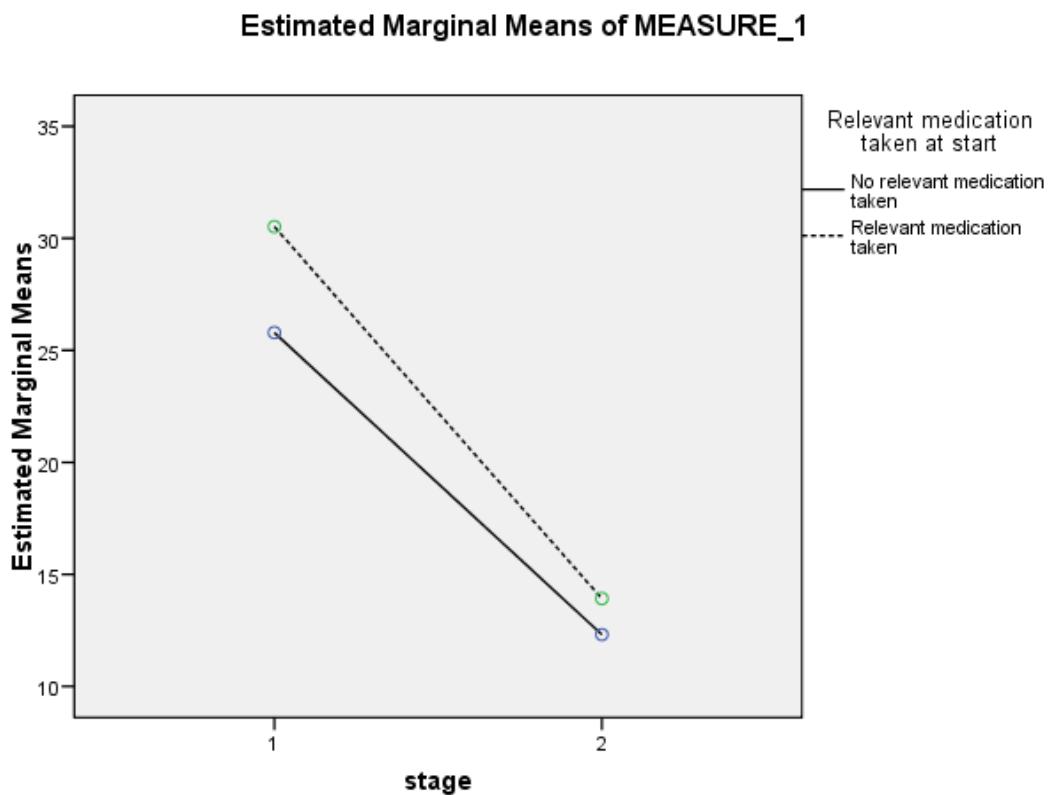


7. What effect did prescription medication have on change in depression?

The 111 clients who started therapy with a clinical level of depression and had a subsequent measure of their depression included 25 clients who were taking some form of prescription medication that might have had some psychological impact: 20 were taking anti-depressants, one was taking anxiolytics, one was taking anti-depressants plus anti-IBS drugs, one was taking a sedative, one was taking anti-psychotics and one was taking a psycho-active drug of unknown type. A repeated measures analysis of variance with group membership (relevant prescription medication taken or not) as a between-subjects factors showed scores for both groups combined changed significantly during treatment $F(1, 109) = 168.13, p < .001$, $ES(d) = 1.48$ (95% CI 1.33 to 1.81) and there was no statistically significant difference between the two group means at first and last session combined, $F(1, 109) = 3.26, p = .074$, see Figure 21.

There was no statistically significant interaction between stage of therapy and medication taking, $F(1, 109) = 1.81$, $p = .181$, suggesting no significant difference in how much the medication taken or not groups changed during therapy.

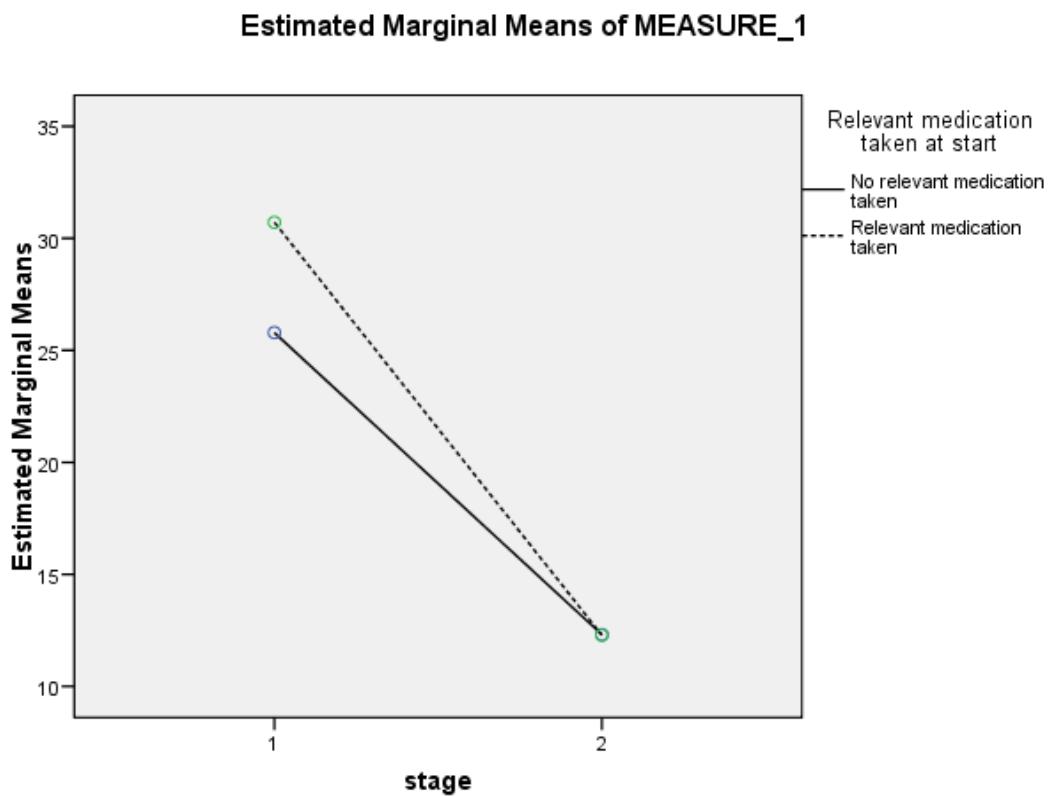
Figure 21: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 25 clients in the ‘relevant medication taken’ condition and 86 clients in the ‘no relevant medication taken’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and depression (BDI-II) score is y-axis.



Since this was an analysis of change in depression, the analysis was re-run to compare the 86 clients taking no relevant medication with only the 21 clients taking anti-depressants. A repeated measures analysis of variance with group membership (no relevant prescription medication taken or anti-depressants taken) as between-subjects factors showed scores for both groups combined changed significantly during treatment $F(1, 105) = 167.37$, $p < .001$, $ES(d) = 1.51$ (95% CI 1.41 to 1.92) and there was no statistically significant difference between the two group means at first and last session combined, $F(1, 105) = 1.71$, $p = .194$, see Figure 22. There was a statistically significant interaction between the stage of therapy and medication taken status, $F(1, 105) = 4.032$, $p = .047$. This suggested there was a

significant difference in how much the medication taken or not groups changed during therapy; the medication taken group changed on average by 18.43 BDI-II units and the no medication group by 13.48 BDI-II units, a mean difference of 4.95 BDI-II units. An independent samples t test confirmed the significance of this difference in change scores $t(105) = 2.00$, $p = .047$, $r = .19$, a small effect.

Figure 22: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 21 clients in the ‘anti-depressant medication taken’ condition and 86 clients in the ‘no relevant medication taken’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and depression (BDI-II) score is y-axis.



These findings were similar to those reported by Stiles, et al., (2007) in their further analysis (Stiles, et al., 2008) in that those prescribed medication tended to start with higher scores. In this research any effect of ‘relevant medication taken’ was not big enough to be anything other than a chance finding although there was evidence of a statistically significant small effect for ‘anti-depressant medication taken’ status, although it was unclear whether the anti-

depressant medication itself was causative. There was insufficient evidence that the clients in this research got better solely because of concurrent medication taken cf. criticisms of this type of uncontrolled naturalistic research (Clark, et al., 2007).

As this was naturalistic research, rather than a randomised controlled trial, it was not possible to compare two identical conditions, a wait condition with a treatment condition, and deduce (deductive logic) that any observed difference was solely because of the difference between wait and treatment. However, a series of analyses was conducted to see what could be inferred (inductive logic) about the effect of being in therapy on the observed improvements. In summary, a group of 36 clients was examined to compare what happened to depression scores whilst these clients waited for therapy with what happened during treatment. On average these clients changed significantly during treatment and not during wait. The treatment period was significantly longer than the wait period, so the effect of time in predicting change in depression scores was analysed and it was found there was insufficient evidence of time (days) in either condition predicting change in depression scores, suggesting it wasn't simply the passage of time that resulted in change in depression for this group of clients. This group of clients might have been different to other clients in the research so the depression scores of these clients were compared with those other clients in the research who waited for therapy and also those other clients who had treatment. There was no significant difference in depression scores or their changes during wait or treatment for the 36 clients in the 'wait time control of treatment outcomes' part of the research and other clients, as defined. This suggested that, in terms of depression scores, the 36 clients were representative of other clients in the research and the finding, clients probably changed more in treatment than in waiting, could likely be extended to other clients in the research. Finally the clients who received treatment were compared with those who started treatment and had no subsequent measure of the effect of treatment. It was found that on average it was the more depressed clients who were measured for progress in therapy. Furthermore, it was found that there was no significant difference in outcomes for those taking medication compared with those who were not (although there was a significant and small effect amongst the small proportion taking anti-depressants which may or may not have been due to the anti-depressants).

Taken together these findings suggested it was unlikely cf. some criticisms that have been made of naturalistic research (Clark, et al., 2007), that:

- There were lots of ‘missing cases’ who were really treatment failures
- These findings were restricted to the ‘easiest’ clients in the research
- These findings were restricted to the ‘best’ outcomes
- These findings were simply ‘regression to the mean’
- These findings were simply ‘natural recovery’
- These findings were ‘attributable to concurrently administered medications’

It seemed plausible that person-centred therapy in this research was at least partly responsible for the observed improvements.

b) Anxiety

As described above, seven analyses were performed to assess to what extent therapy might be responsible for the observed improvements.

1. Were any observed changes simply regression to the mean?

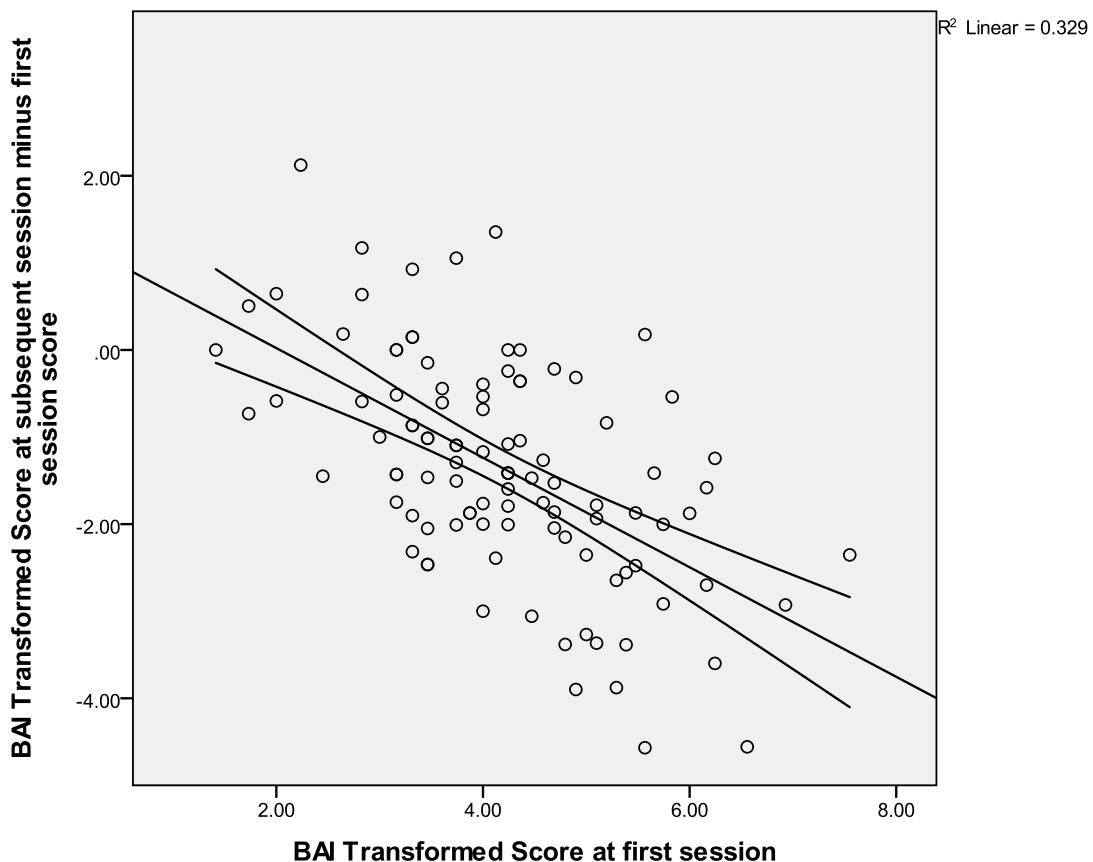
Correction for regression to the mean effects requires data to be distributed not significantly different from normal. As described above, BAI scores at the first time of completion were distributed significantly different from normal and the data was transformed to a distribution not significantly different from normal by adding one to each score and taking the square root of the resultant sum for this part of the analysis. References to BAI scores in this section are to the transformed scores.

During treatment, 102 clients had their BAI scores recorded at their first and subsequent counselling session. The mean score for this group was 4.20, with a standard deviation of 1.17; the population was distributed not significantly different from a normal distribution $K-S D(102) = .07, p > .2$.

Hypothesis A2 was about clients with 'clinical anxiety', raw scores of 8 and above, and so this cut-off was applied to derive the 'clinical sample' of 91 clients (those with a raw BAI score of 8 or above at their first session), from the sample of 102 clients. These 91 clients with 'clinical anxiety' had a mean score of 4.44 at the first session, SD .99, and a mean score of 2.89 at their subsequent measurement, SD 1.13. To what extent this observed change of 1.55 in mean BAI scores, $ES = 1.5687$, for the subsample starting with 'clinical anxiety', was simply a regression to the mean effect is considered below, using the method described by Barnett, et al, (2005).

BAI score, at first session versus subsequent counselling session minus score at first session is plotted below, Figure 23.

Figure 23: Scatter plot of transformed BAI at first session versus subsequent counselling session minus score at first session.

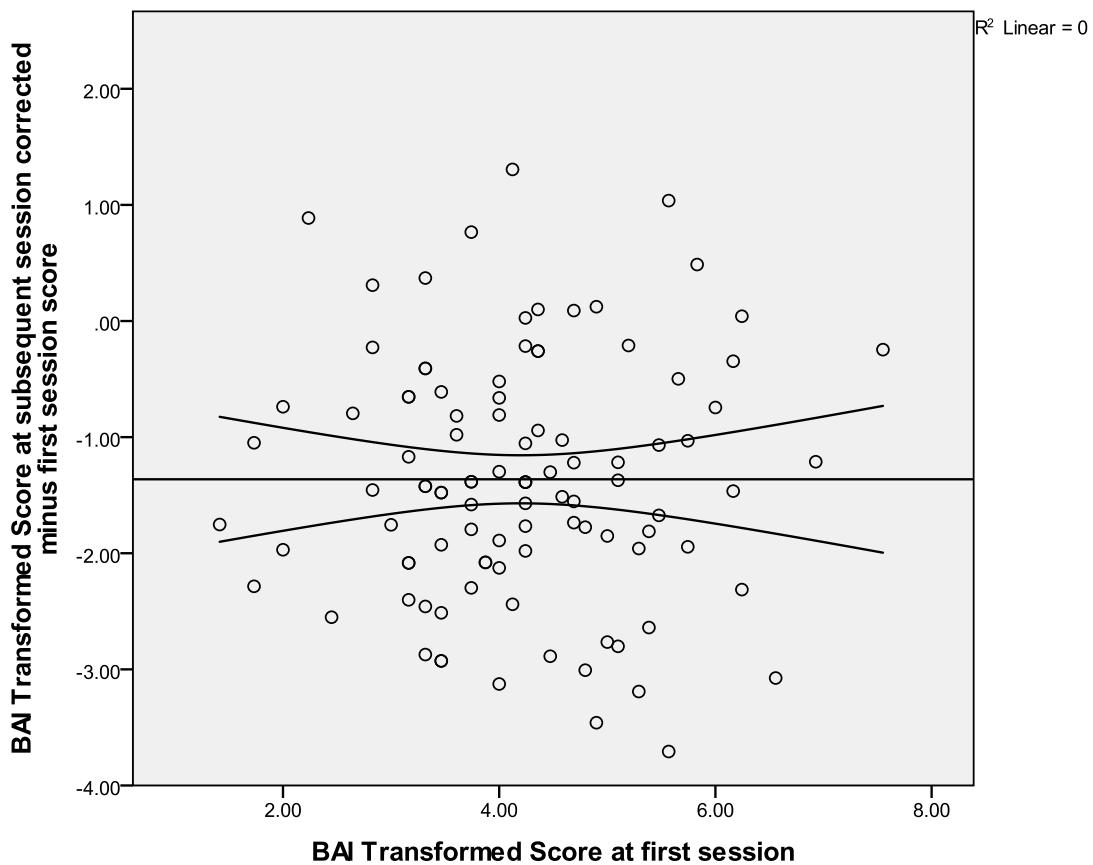


There was some evidence of a regression to the mean effect, in that some clients whose BAI scores were low (below the population mean) at the first session had increased at the

subsequent session and some of those with higher scores (above the population mean) had reduced at the subsequent session.

Regression to the mean effects were corrected for using the method described above with $B = .371$, $p < .001$, see Figure 24.

Figure 24: Scatter plot of regression to mean corrected transformed BAI at subsequent session minus first session score versus first session score.



For the 91 clients with clinical scores (raw BAI ≥ 8) at first session, the mean score at subsequent session corrected for regression to the mean effects was 3.04, SD 1.48. Using repeated measures analysis of variance there was a significant improvement from first to subsequent for both the original data $F(1, 90) = 155.53$, $p < .001$ and the corrected data $F(1, 90) = 163.59$, $p < .001$. Correcting for regression to the mean reduced the size of the change between first and subsequent session to 1.40, $ES(d) = 1.4187$; suggesting .15 transformed BAI units, $ES(d) = .15$ or 9.6% of the effect was due to regression to the mean.

Comparing the mean change without correction for regression to the mean, 1.55, SD 1.19, with the mean change with correction for regression to the mean, 1.40, SD 1.04, with a paired samples t-test was statistically significant $t(90) = 2.28$, $p = .025$, $r = .23$. This was a small-medium effect; mean estimate 0.149 transformed BAI units, 95% confidence interval 0.019 to 0.278. Converting back to BAI units was not straightforward in the sense that the transformed figure represented the difference between two square root expressions. One solution that fit this expression was that the mean estimate for the regression to the mean effect during treatment was approximately equal to the difference between raw BAI scores of 11 and 10, i.e. $\sqrt{11+1} - \sqrt{10+1} = .147$. For sake of simplicity on average it could be considered that the size of the regression to the mean effect during treatment was approximately one BAI unit. Whilst change from first to subsequent session was much greater than could be explained by regression to the mean effects, regression to the mean was a statistically significant small-medium effect.

2. Did clients change during wait or treatment condition?

At the UCS there were 36 clients who had anxiety (BAI) scores at exploratory, first session and a subsequent session, had a clinical anxiety score (BAI ≥ 8) at first session and had data about the length (days) of their wait and treatment. The results of a repeated measures ANOVA showed (Table 35) mean client BAI scores improved significantly during 'wait', from exploratory session to first session $F(1, 33) = 6.69$, $p = .014$, $ES(d) = .18$ (95% CI -.08 to .45) and also improved significantly during 'treatment', from first session to a subsequent session $F(1, 33) = 4.86$, $p = .035$, $ES(d) = .57$ (95% CI .25 to .89). Mean client BAI scores improved significantly during both wait and treatment; on average there was a greater improvement during treatment than wait, incremental $ES(d) = .38$ (.5737 minus .1878). The mean change during wait was 2.22 BAI units (SD 7.834) and during treatment 6.78 BAI units (SD 9.037). A paired samples t-test comparing the improvement during wait with that during treatment suggested this difference was not significant $t(35) = 1.96$, $p = .058$, $r = .31$, a medium effect. It was plausible that the sample size was underpowered for a significant finding for this medium-sized effect suggesting this effect should not be ignored. In general clients waited a significantly shorter number of days (Median 20.5 days) than they were in treatment for (Median 49 days) $z = 3.39$, $p = .001$, $r = .32$, so the effect of time on change in anxiety was assessed, in the next section.

These findings were similar to those found by Svartberg et al. (2001) who found 'patients with major depression did not change after the pretherapy interviews, whereas anxiety patients changed substantially and significantly faster than the depressed patients' (p. 201).

Table 35: Anxiety (BAI) outcomes comparing wait and treatment.

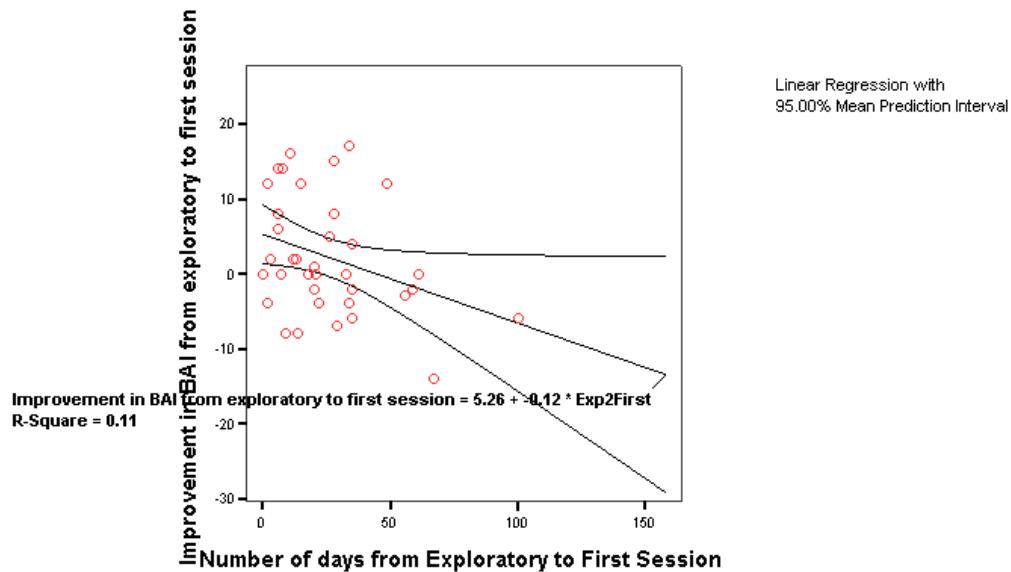
Group	n	Start		Subsequent		ES(d) ^a	p
		Mean	SD	Mean	SD		
All clients with clinical BAI score at first session (BAI ≥ 8):							
Exp to First	36	20.86	11.82	18.64	10.94	.1878	.014
First to Subsequent	36	18.64	10.94	11.86	8.22	.5737	.035

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at exploratory session.

3. What effect did time have on change in anxiety?

Including the wait period (days) and the treatment period (days) as covariates (repeated measures ANCOVA) suggested there was a significant interaction between wait period (days) and change in anxiety score from exploratory session to first session $F(1, 33) = 4.36$, $p = .045$, i.e. longer wait was associated with worsening anxiety (Figure 25). A 'negative improvement' was a worsening in anxiety.

Figure 25: Scatter plot of wait period (number of days from exploratory session to first session) versus improvement in anxiety (BAI) during wait period: A negative improvement was a worsening in anxiety score.



There was no significant interaction between treatment period (days) and change in anxiety score from first therapy session to a subsequent measure $F(1, 33) = 1.73, p = .197$. There was insufficient evidence to suggest time in treatment interacted with change in anxiety score and evidence to suggest time waiting interacted with change in anxiety (longer wait, worsening anxiety).

Further examination of the interaction between wait period (days) and change in anxiety during wait with regression analysis suggested wait period (days) significantly predicted change in anxiety ($B = -.118, p = .048$), see analysis 1 in Table 36. Examination of residuals showed there were no outliers (standardised residuals ≥ 3), see also Figure 25. On average, immediately following the exploratory session it appeared that clients had a statistically significant improvement in anxiety of 5.3 BAI units, $p = .010$ (95% CI 1.3 to 9.2).

Table 36: Summary of Regression Analysis for Wait period (days) in predicting change in Anxiety (BAI) during wait: Model 1 had no outlying cases (standardised residuals ≥ 3).

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	P
1	Imp Anx	Constant	5.261	1.937				.010
		Wait period (days)	-.118	.058	-.332			.048
						.33	4.21	.048

Note: n = 36. ^a For model 1df = 1, 34. DV = dependent variable. Imp Anx Exp to First = Improvement in Anxiety (BAI) from Exploratory Session to First Session.

Examination of the interaction between treatment period (days) and change in anxiety during treatment with regression analysis suggested treatment period (days) did not significantly predict change in anxiety; see analysis 1 in Table 37. There were no outliers (standardised residuals ≥ 3). The statistically significant constant term suggested on average immediately following the first session clients improved by 4.7 BAI units (95% CI .4 to 9.0).

Table 37: Summary of Regression Analysis for Treatment period (days) in predicting change in Anxiety (BAI) during treatment: Model 1 had no outlying cases (standardised residuals ≥ 3).

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	P
1	Imp Anx	Constant	4.704	2.098				.032
		Treatment period (days)	.026	.019	.233			.171
						.05	1.96	.171

Note: n = 36. ^a For model 1df = 1, 34. DV = dependent variable. Imp Anx First to Subseq = Improvement in Anxiety (BAI) from First Session to Subsequent Session.

Analysis of the subset for which it was possible to compare changes in anxiety during wait and treatment suggested a complicated mixture of forces at play. It appeared clients improved during wait and treatment, and treatment offered only an incremental benefit. From the regression analysis (Table 36) it appeared immediately following the exploratory session that clients improved on average by 5.26 BAI points, p = .010 (95% CI for improvement 1.32 to

9.20 BAI points) immediately following the exploratory session (Figure 25); and the average improvement was eroded by passage of time to the first session, so average BAI score worsened by .118 BAI units per day, $p = .048$ (95% CI for rate of deterioration .001 to .236 BAI units per day). On this basis an average initial improvement, or 'early treatment gain', of 5.26 BAI units eroding at .118 BAI units per day had on average returned to a zero improvement (i.e. no change) after 44.6 days (5.26/.118). It appeared treatment offered only an incremental benefit to waiting; however, it was likely this was because clients waited for a relatively short time (median 20.5 days) compared with treatment (median 49 days) and built an initial, but perhaps unstable, early 'treatment' gain after the exploratory session until the time of their first session. The regression analysis showed it was predictable clients deteriorated with the passage of time after their exploratory session. It appeared on average following the first session clients improved by 4.7 BAI units and there was then no significant relationship between time and change in anxiety score. The extent to which these 36 clients were representative, in terms of their anxiety scores, of the other clients in this research was assessed in the next section.

4. How representative were the clients in the 'wait time control of treatment outcomes' of those other clients who waited for therapy? And (5.) how representative were they of those other clients who had a subsequent measurement of their anxiety?

To what extent the 36 clients in the wait time control of treatment outcomes were representative of the larger samples, in terms of anxiety scores at the different stages was addressed in two ways:

- Firstly, by comparing the 36 clients with both wait and treatment outcomes with the 30 UCS clients who had data only for the wait condition. A repeated measures analysis of variance with group membership as a between-subjects factor showed scores for both groups combined changed significantly during wait $F(1, 64) = 5.33, p = .024$ and there was no statistically significant difference between the two group means at exploratory and first session combined, $F(1, 64) = .06, p = .814$, see Figure 26.
- Secondly, by comparing the 36 clients with both wait and treatment outcomes with the 54 clients (PP) who had data only for the treatment condition. A repeated measures analysis of variance with group membership as a between-subjects factor

showed scores for both groups combined changed significantly during treatment $F(1, 88) = 121.21, p < .001$, ES (d) = 1.01 (95% CI .88 to 1.27). (Note: This is not the anxiety hypothesis sample, since case 303 was omitted from the 'wait and treatment condition' as no data on wait and treatment period was available for this case). There was no statistically significant difference between the two group means at first and last session combined, $F (1, 88) = 1.48, p = .227$, see Figure 27.

Figure 26: Estimated marginal means at exploratory session (stage 1) and first session (stage 2) for 36 UCS clients in the 'wait versus treatment' condition and 30 UCS clients in the 'wait only' condition: stage (1 = exploratory session, 2 = first session) is x-axis and anxiety (BAI) score is y-axis.

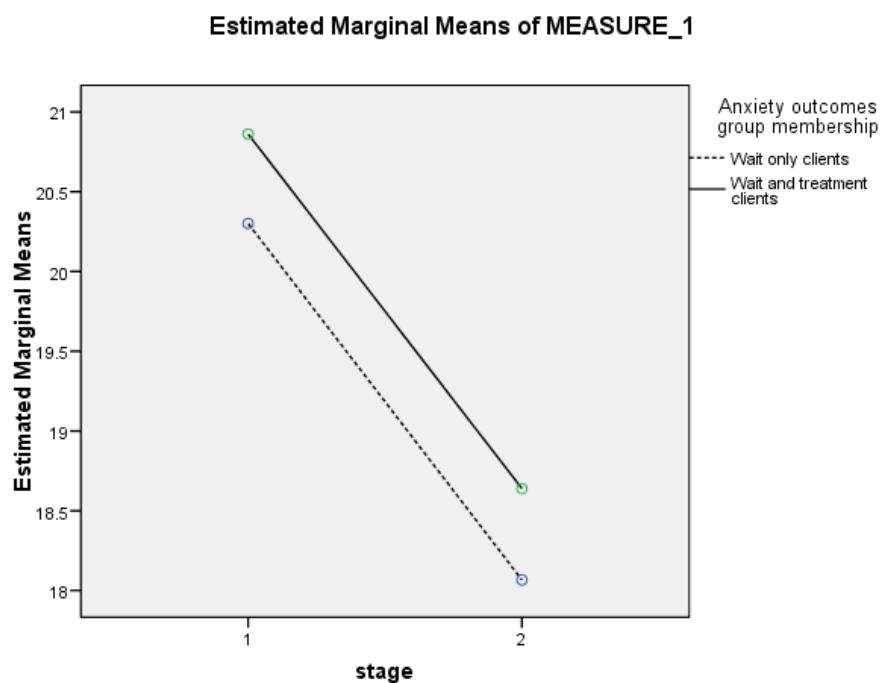
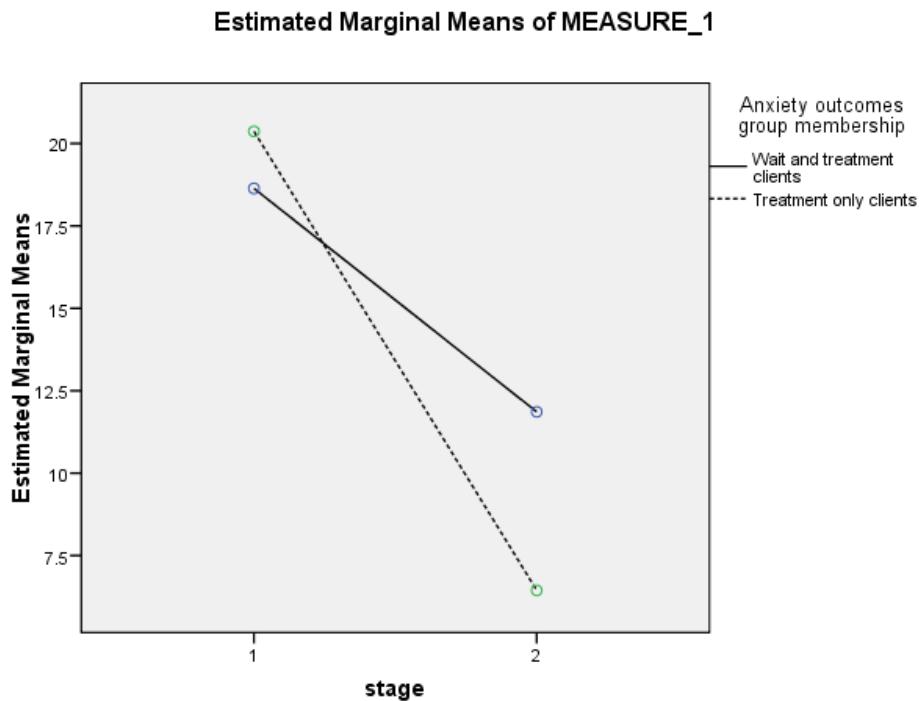


Figure 27: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 36 UCS clients in the ‘wait versus treatment’ condition and 54 PP clients in the ‘treatment only’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and anxiety (BAI) score is y-axis.



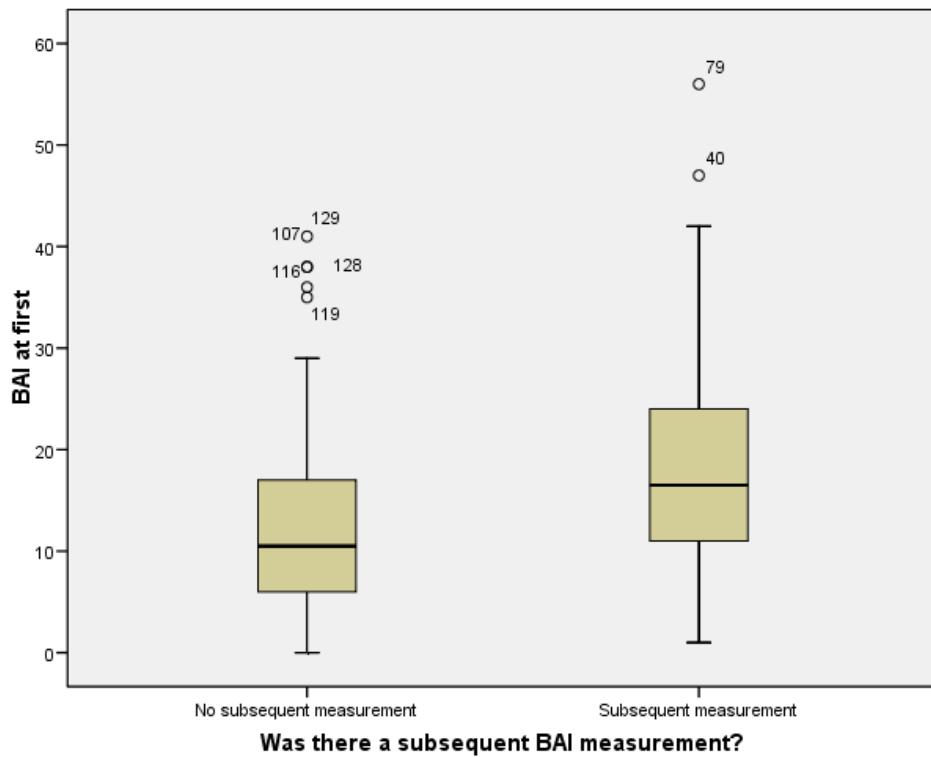
The 36 UCS clients with both wait and treatment outcomes were representative of the 30 UCS clients with data for wait only and the 54 PP clients with data for treatment only. To what extent clients who started in therapy with clinical anxiety and had a subsequent measurement of their anxiety were representative of those who didn't do a subsequent measure of their anxiety was assessed, in the next section.

6. How representative, in terms of anxiety scores, were those clients who started in therapy with a clinical anxiety score and didn't do a subsequent measurement of their anxiety, with those who did?

The 204 clients with a BAI score at first session were examined for this part of the analysis. The 102 clients with a subsequent BAI score had a mean score at first session (mean = 17.99, SD = 10.24) higher than the 102 clients with no subsequent BAI score (mean = 12.77, SD = 9.203) and this difference was significant $t(202) = 5.22$, $p < .001$, a medium effect $r = .34$, see

Figure 28. On average it was the clients who started off more anxious who were more likely to have their progress measured.

Figure 28: Box plot of anxiety (BAI) scores at first session (y-axis) for clients with no subsequent BAI measurement (102 clients) and a subsequent BAI measurement (102 clients): Clients with a subsequent measurement started therapy with a significantly higher level of anxiety $p < .001$.

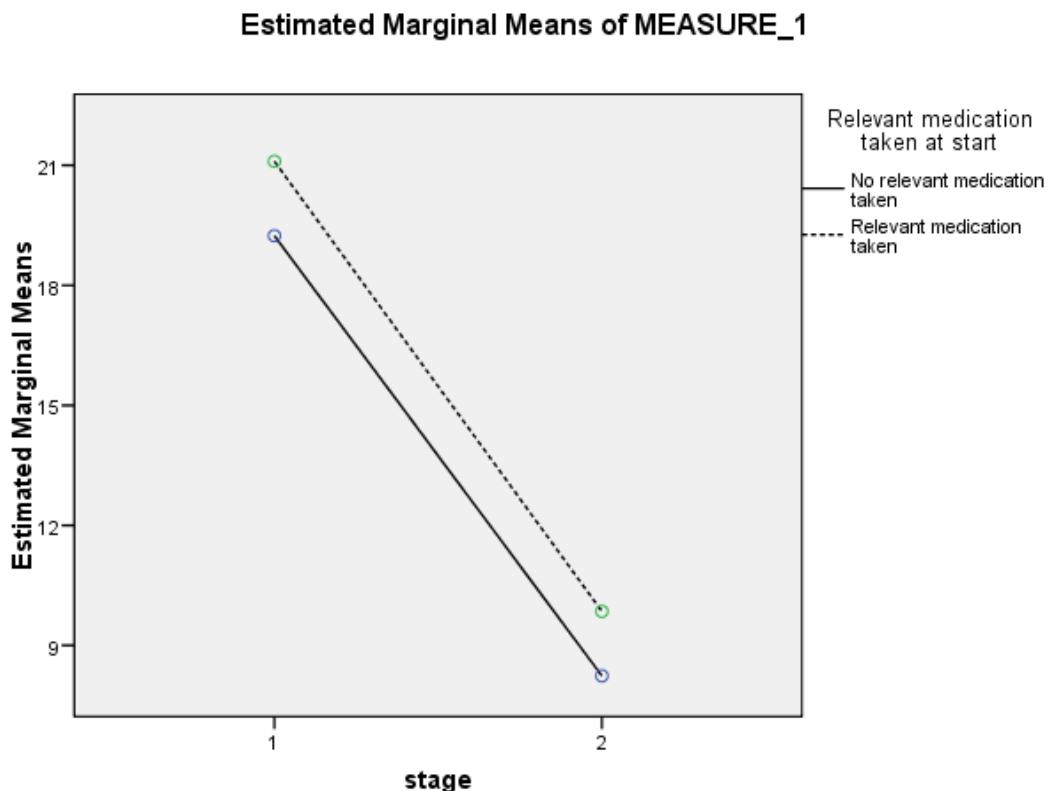


7. *What effect did prescription medication have on change in anxiety?*

The 91 clients who started therapy with a clinical level of anxiety and had a subsequent measure of their anxiety included 20 clients who were taking some form of prescription medication that might have had some psychological impact: 16 were taking anti-depressants, two were taking anxiolytics, one was taking anti-depressants plus anti-IBS drugs and one was taking a psycho-active drug of unknown type. A repeated measures analysis of variance with group membership (relevant prescription medication taken or not) as a between-subjects factors showed scores for both groups combined changed significantly during treatment $F(1, 89) = 87.86, p < .001$, $ES(d) = 1.15$ (95% CI .91 to 1.41) and there was no statistically significant difference between the two group means at first and last session combined, $F(1,$

89) = .95, $p = .332$, see Figure 29. The interaction term was not significant $F(1, 89) = .01, p = .916$.

Figure 29: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 20 clients in the ‘relevant medication taken’ condition and 71 clients in the ‘no relevant medication taken’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and anxiety (BAI) score is y-axis.



In this research any effect of ‘relevant medication taken’ was not big enough to be anything other than a chance finding. There was insufficient evidence that the clients in this research got better because of concurrent medication taken.

Taken together these findings suggested it was unlikely cf. some criticisms that have been made of naturalistic research (Clark, et al., 2007), that:

- There were lots of ‘missing cases’ who were really treatment failures
- These findings were restricted to the ‘easiest’ clients or ‘best’ outcomes
- These findings were simply ‘regression to the mean’ or ‘natural recovery’

- These findings were ‘attributable to concurrently administered medications’

It seemed plausible person-centred therapy in this research was at least partly responsible for the observed improvements.

c) Distress

As described above, seven analyses were performed to assess to what extent therapy might be responsible for the observed improvements.

1. Were observed treatment changes simply regression to the mean?

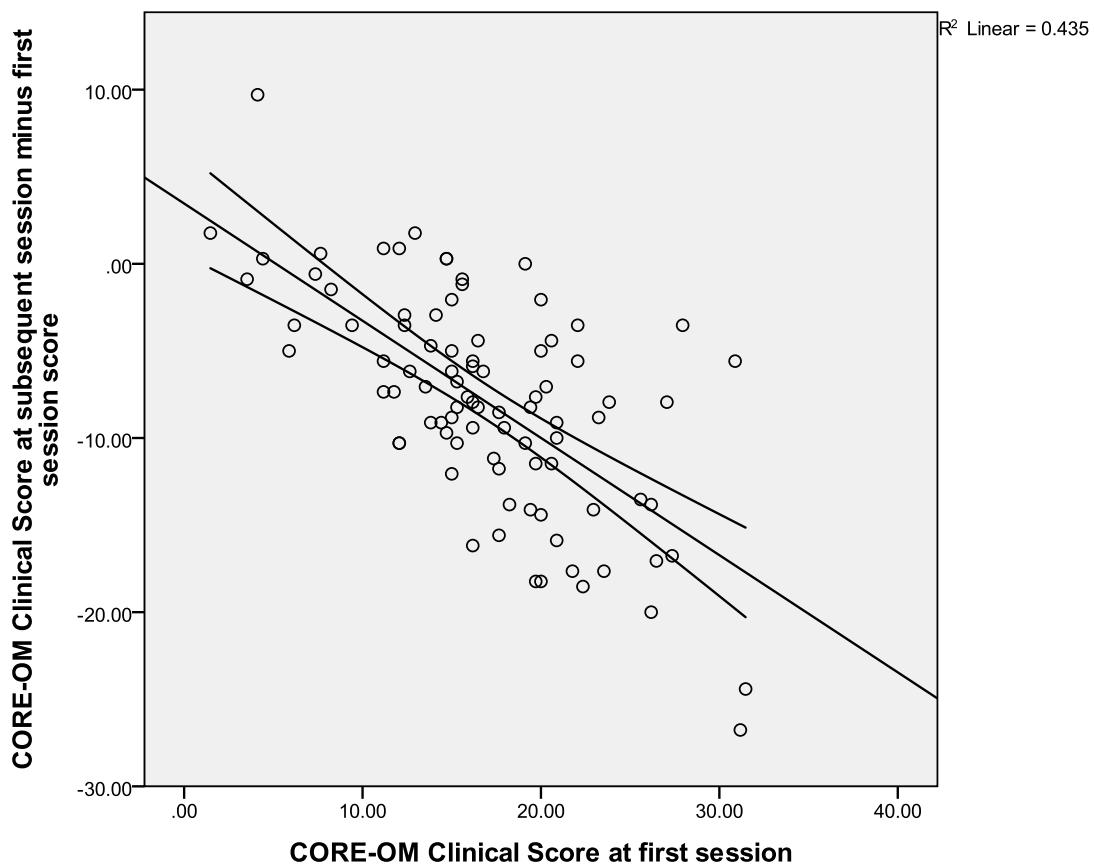
The effects of regression to the mean were assessed for the treatment phase, CORE-OM scores at the first time of completion were distributed not significantly different from normal.

During treatment, eighty nine clients had their CORE-OM scores recorded at their first and subsequent counselling session. The mean score for this group was 16.98, with a standard deviation of 6.20, the population was distributed not significantly different from a normal distribution K-S D(89) = .07, $p > .2$.

Hypothesis A3 was about clients with ‘clinical distress’, scores of 10.0 and above, and so this cut-off was applied to derive the ‘clinical sample’ of 79 clients (those with a CORE-OM score of 10.0 or above at their first session), from the sample of 89 clients. These 79 clients with ‘clinical distress’ had a mean score of 18.39 at the first session, SD 4.96, and a mean score of 9.45 at their subsequent measurement, SD 5.17. To what extent this observed change of 8.939 in mean CORE-OM scores, $ES = 1.8007$, for the subsample starting with ‘clinical distress’, was simply a regression to the mean effect is considered below.

CORE-OM score, at first session versus subsequent counselling session minus score at first session is plotted below, Figure 30.

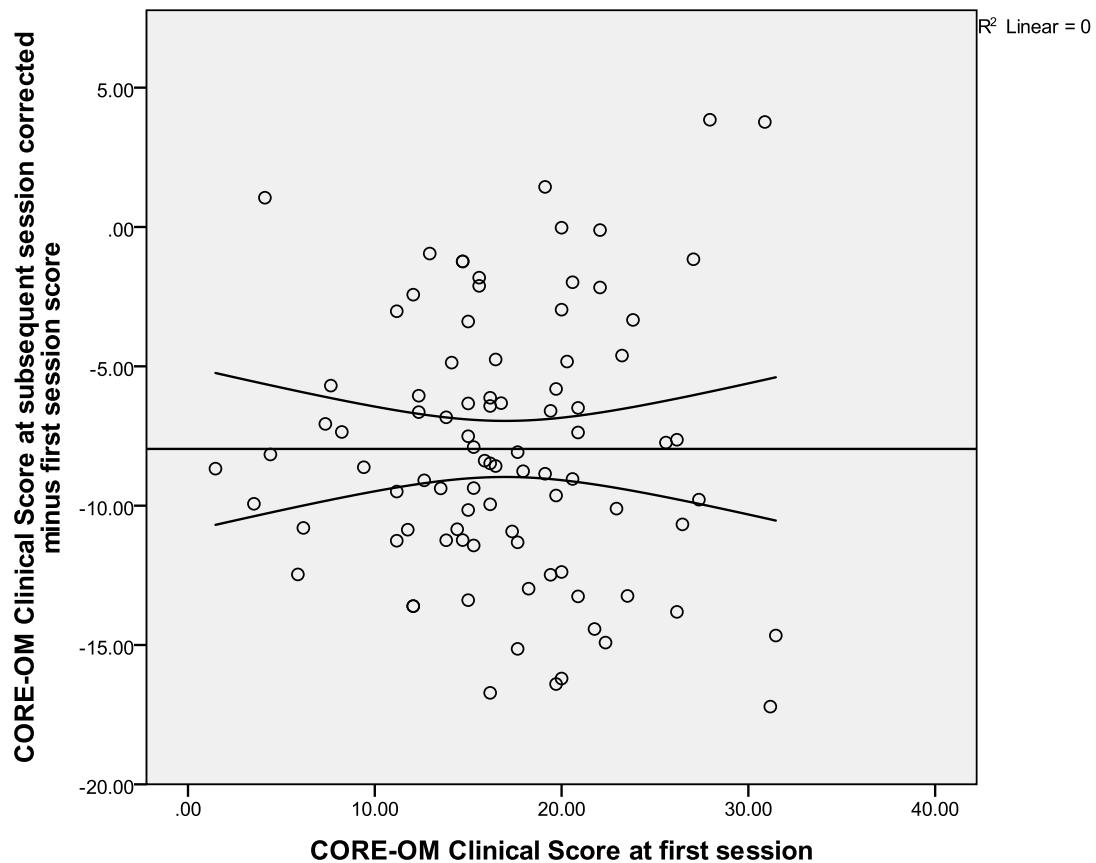
Figure 30: Scatter plot of CORE-OM at first session versus subsequent counselling session minus score at first session.



There was some evidence of a regression to the mean effect, in that some clients whose CORE-OM scores were low (below the population mean) at the first session had increased at the subsequent session and some of those with higher scores (above the population mean) had reduced at the subsequent session.

Regression to the mean effects were corrected for using the method described above with $B = .327$, $p < .001$, see Figure 31.

Figure 31: Scatter plot of regression to mean corrected CORE-OM at subsequent session minus first session score versus first session score.



For the 79 clients with clinical scores ($\text{CORE-OM} \geq 10.00$) at first session, the mean score at subsequent session corrected for regression to the mean effects was 10.40, SD 7.02. Using repeated measures analysis of variance there was a significant improvement from first to subsequent for both the original data $F(1, 78) = 182.71, p < .001$ and the corrected data $F(1, 78) = 211.14, p < .001$. Correcting for regression to the mean reduced the size of the change between first and subsequent session to 7.989, $\text{ES}(d) 1.6092$; suggesting 0.95 CORE-OM units, $\text{ES}(d) .1913$ or 10.6% of the effect was due to regression to the mean.

Comparing the mean change without correction for regression to the mean, 8.94, SD 5.88, with the mean change with correction for regression to the mean, 7.99, SD 4.89, with a paired samples t-test was statistically significant $t(78) = 2.53, p = .013, r = .28$. This was a medium effect, mean estimate 0.95 CORE-OM units, 95% confidence interval .202 to 1.699 CORE-OM units. Whilst change from first to subsequent session was much greater than could be explained by regression to the mean effects, regression to the mean was a statistically significant medium effect.

2. Did clients change during wait or treatment condition?

At the UCS there were 37 clients who had CORE-OM scores at exploratory, first session and a subsequent session, had a clinical distress score (CORE-OM ≥ 10) at first session and had data about wait and treatment periods (days). The results showed (Table 38) mean client CORE-OM scores improved significantly during 'wait', from exploratory session to first session $F(1, 34) = 10.02$, $p = .003$, $ES(d) = .23$ (95% CI minus .01 to plus .46; minus figure due to Bonferroni correction adjusted estimate for marginal means) and also improved significantly during 'treatment', from first session to a subsequent session $F(1, 34) = 12.47$, $p = .001$, $ES(d) = 1.45$ (95% CI .98 to 1.92). Mean client CORE-OM scores improved significantly during both wait and treatment; on average there was a greater improvement during treatment than wait, incremental $ES(d) = 1.2243$ (1.4520 minus .2277). The mean change during wait was 1.21 CORE-OM units (SD 3.232) and during treatment 7.70 CORE-OM units (SD 6.39), with a paired samples t-test this was statistically significant $t(36) = 4.79$, $p < .001$, $r = .62$, a large effect. On this basis the mean improvement during treatment was significantly greater than during wait and treatment appeared to offer a significant advantage to wait. However, in general clients waited a significantly shorter number of days (Median 20 days) than they were in treatment for (Median 49 days) $z = 3.92$, $p < .001$, $r = .37$, a medium effect.

Table 38: Distress (CORE-OM) outcomes comparing wait and treatment

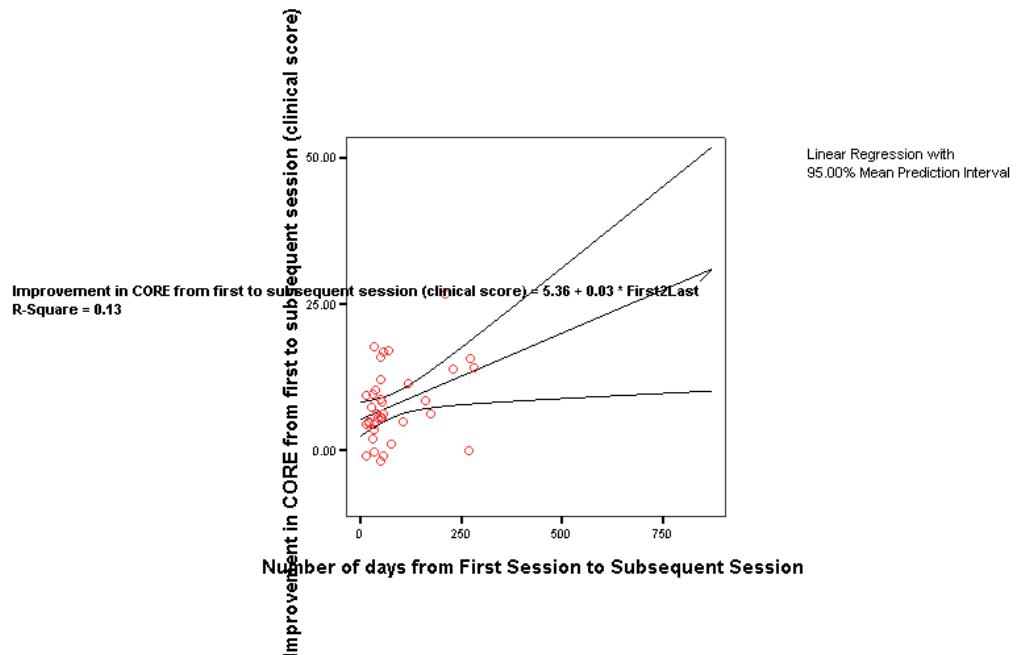
Group	N	Start		Subsequent		ES(d) ^a	p
		Mean	SD	Mean	SD		
All clients with clinical CORE-OM score at first session (CORE-OM ≥ 10):							
Exp to First	37	19.05	5.305	17.84	4.963	.2277	.003
First to Subsequent	37	17.84	4.963	10.14	4.416	1.4520	.001

Note: ^a Effect size was calculated by outcome measure at start minus subsequent outcome measure divided by standard deviation of outcome measure at exploratory session.

3. What effect did time have on change in distress?

Including the wait period (days) and the treatment period (days) as covariates (repeated measures ANCOVA) suggested there was no significant interaction between wait period (days) and change in distress score during 'wait', from exploratory session to first session, $F(1, 34) = 1.62$, $p = .212$, however there was a significant interaction between 'treatment' period (days) and change in distress score during 'treatment', from first therapy session to a subsequent measure, $F(1, 34) = 5.01$, $p = .032$, on average a longer treatment period resulted in a greater improvement in distress, see Figure 32.

Figure 32: Scatter plot of treatment period (number of days from first session to a subsequent session) versus improvement in distress (CORE-OM) during treatment period: A positive improvement in CORE-OM score was a lessening in distress.



There was evidence to suggest time in treatment interacted with change in distress score (longer treatment, greater improvement in distress) and insufficient evidence to suggest time waiting interacted with change in distress.

Whilst there was insufficient evidence of an interaction between time and change in distress during wait, examination of the interaction between wait (days) and change in distress during the wait period (Table 39) suggested that on average immediately following the exploratory session there was a clinically significant improvement of 1.929 CORE-OM units (95% CI .251 to 3.606).

Table 39: Summary of Regression Analysis for Wait period (days) in predicting change in Distress (CORE-OM) during wait: Model 1 has no outlying cases (standardised residuals ≥ 3).

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	P
1	Imp Distress Exp to First	Constant	1.929	.826				.025
		Wait period (days)	-.028	.025	-.188			.264
						.04	1.29	.264

Note: n = 36. ^a For model 1 df = 1, 35. DV = dependent variable. Imp Distress Exp to First = Improvement in Distress (CORE-OM) from Exploratory Session to First Session.

Further examination of the interaction between treatment period (days) and change in distress during treatment period, with regression analysis, suggested treatment period (days) significantly predicted change in distress ($B = .029$, $p = .029$), see Table 40. Examination of residuals showed there were no outliers (standardised residual ≥ 3).

Table 40: Summary of Regression Analysis for Treatment period (days) in predicting change in Distress (CORE-OM) during treatment: No outlying cases (standardised residual ≥ 3).

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
1	Imp CORE First to Subsequent	Constant	5.359	1.430				.001
		Treatment period (days)	.029	.013	.360			.029
						.13	5.20	.029

Note: n = 37. ^a For model 1 df = 1, 35. DV = dependent variable. Imp CORE First to Subsequent = Improvement in Distress (CORE-OM) from First Session to Subsequent Session.

Whilst the regression model predicting change in distress from exploratory to first session from number of days spent waiting was not significant ($p = .264$) the constant in the equation, 1.9 CORE-OM units was significant, $p = .025$. The exploratory session was with a qualified therapist, 27.0% of clients spent approximately quarter of an hour with their exploratory session therapist, 43.2% spent about half an hour and 29.7% spent about an hour with the therapist doing the exploratory. This suggested that following the exploratory session clients immediately saw on average an 'early treatment gain' of 1.9 CORE-OM units. Unlike the comparable anxiety regression model there was insufficient evidence of a statistically significant rate of subsequent deterioration. The regression model predicting change in distress during treatment (Table 40) was statistically significant ($p = .029$). The constant in the equation, 5.36 (95% CI 2.456 to 8.261) was clinically and statistically significant ($p = .001$) and this suggested that following the first session clients immediately saw on average an 'early treatment gain' of 5.4 CORE-OM units and this improved by .029 CORE-OM units per day (95% CI .003 to .055), $p = .029$. The extent to which these 37 clients were representative, in terms of their distress scores, of the other clients in this research was assessed below.

4. How representative were the clients in the 'wait time control of treatment outcomes' of those other clients who waited for therapy? And (5.) how representative were they of those other clients who had a subsequent measure of their distress?

To what extent the 37 clients in the wait time control of treatment outcomes were representative of the larger samples, in terms of distress scores at the different stages was assessed in two ways:

- Firstly, by comparing the 37 UCS clients with both wait and treatment outcomes with the 34 UCS clients who had data only for the wait condition (Figure 33). A repeated measures analysis of variance with group membership as a between-subjects factors showed scores for both groups combined changed significantly during wait $F(1, 69) = 8.32$, $p = .005$ ES(d) = .2492 (95% CI .0769 to .4223) and there was no statistically significant difference between the two group means at exploratory and first session combined, $F(1, 69) = .17$, $p = .678$, see Figure 33. There was no significant interaction

between group membership and stage of counselling $F(1, 69) = .01$, $p = .941$, there was no significant difference in the amount of change from exploratory session to first session between groups.

- Secondly, by comparing the 37 UCS clients with both wait and treatment outcomes with the 41 PP clients who had data only for the treatment condition. A repeated measures analysis of variance with group membership as a between-subjects factor showed scores for both groups combined changed significantly during treatment $F(1, 76) = 181.35$, $p < .001$, $ES (d) = 1.8182$ (95% CI 1.5387 to 2.0729). (Note: This is not the distress hypothesis sample, since case 303 was omitted from the 'wait and treatment condition' as no data on wait and treatment period was available for this case). There was no statistically significant difference between the two group means at first and last session combined, $F (1, 76) = .00$, $p = .971$, see Figure 34. There was no significant interaction between group membership and stage of counselling $F(1, 76) = 3.27$, $p = .074$, there was no significant difference in the amount of change from first session to subsequent session between groups.

Figure 33: Estimated marginal means at exploratory session (stage 1) and first session (stage 2) for 37 UCS clients in the ‘wait versus treatment’ condition and 34 UCS clients in the ‘wait only’ condition: stage (1 = exploratory session, 2 = first session) is x-axis and distress (CORE-OM) score is y-axis.

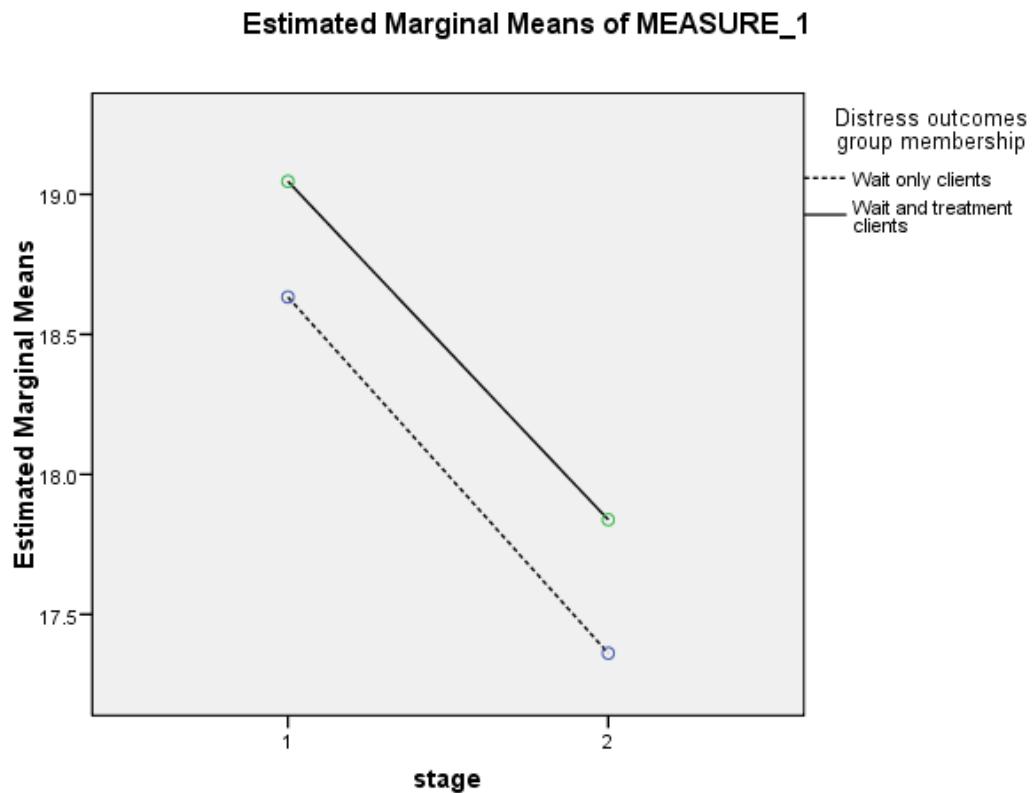
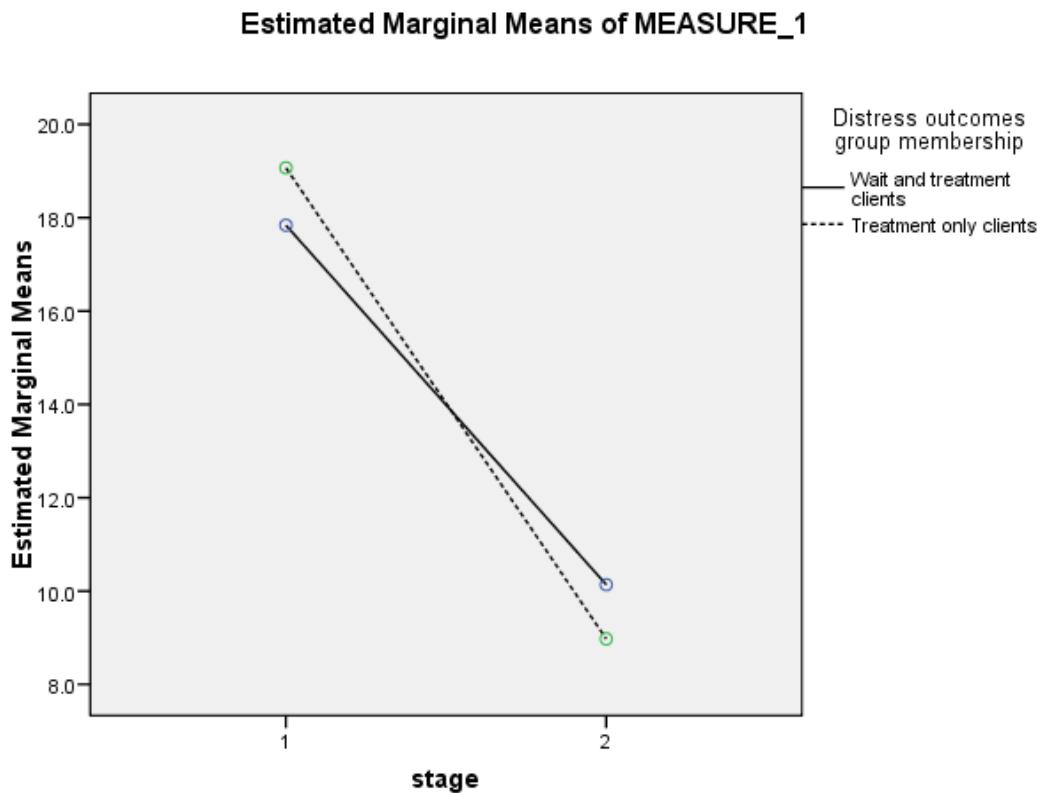


Figure 34: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 37 UCS clients in the ‘wait versus treatment’ condition and 41 PP clients in the ‘treatment only’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and distress (CORE-OM) score is y-axis.



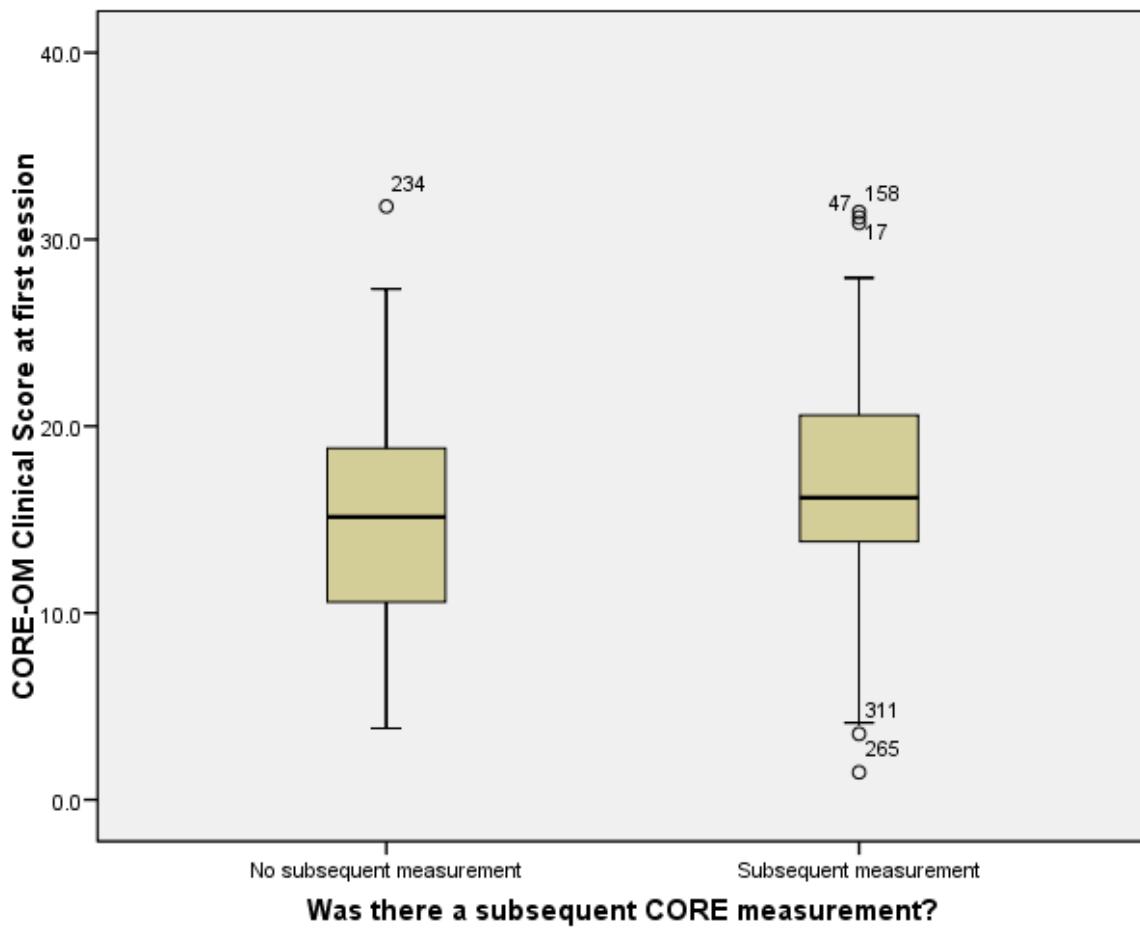
The 37 UCS clients with both wait and treatment outcomes were representative of the 34 UCS clients with data for wait only and the 41 PP clients with data for treatment only. To what extent clients who started in therapy with clinical levels of distress and had a subsequent measurement of their distress were representative of those who didn't do a subsequent measure of their distress was assessed, in the next section.

6. How representative, in terms of distress scores, were those clients who started therapy with a clinical distress score and didn't do a subsequent measurement of their distress with those who did?

The 155 clients with a CORE-OM score at first session were examined for this part of the analysis. The 89 clients with a subsequent CORE-OM score had a mean score at first session

(mean = 16.89, SD = 6.20) higher than the 66 clients with no subsequent CORE-OM score (mean = 15.06, SD = 5.90) and this difference was not significant $t(153) = 1.94$, $p = .054$, see Figure 35.

Figure 35: Box plot of distress (CORE-OM) scores at first session (y-axis) for clients with no subsequent CORE-OM measurement (66 clients) and a subsequent CORE-OM measurement (89 clients): Clients with a subsequent measurement started therapy with a non-significant higher level of distress ($p=.054$).

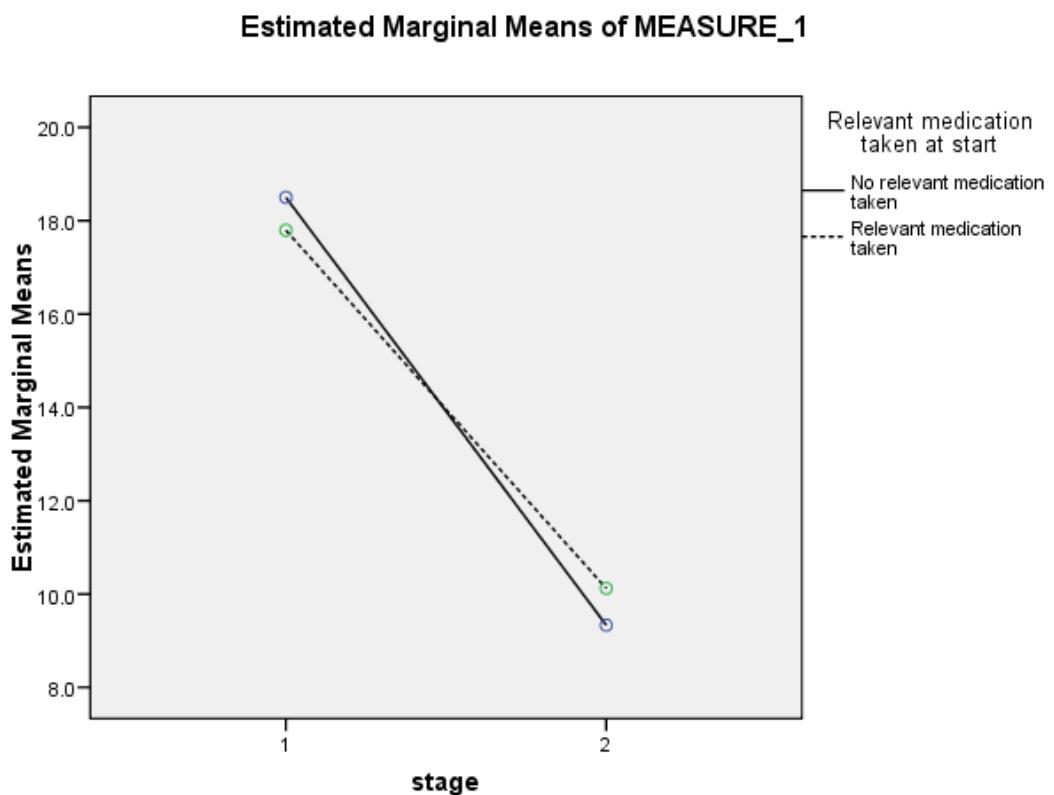


7. *What effect did prescription medication have on change in distress?*

The 79 clients who started therapy with a clinical level of distress and had a subsequent measurement of their distress included 12 clients who were taking some form of prescription medication that might have had some psychological impact: seven were taking anti-depressants, two were taking anxiolytics, one was taking an anti-psychotic, one was taking

anti-depressants plus anti-IBS drugs and one was taking a psycho-active drug of unknown type. A repeated measures analysis of variance with group membership (prescription medication taken or not) as a between-subjects factors showed scores for both groups combined changed significantly during treatment $F(1, 77) = 83.14$, $p < .001$, $ES(d) = 1.80$ (95% CI 1.32 to 2.06) and there was no statistically significant difference between the two group means at first and subsequent session combined, $F(1, 77) = .00$, $p = .974$, see Figure 36. There was no statistically significant interaction between stage of therapy (first session, subsequent session) and medication status (relevant medication taken or not) $F(1, 77) = .66$, $p = .421$.

Figure 36: Estimated marginal means at first (stage 1) and subsequent session (stage 2) for 12 clients in the ‘relevant medication taken’ condition and 67 clients in the ‘no relevant medication taken’ condition: stage (1 = first session, 2 = subsequent session) is x-axis and distress (CORE-OM) score is y-axis.



In this research any effect of ‘relevant medication taken’ was not big enough to be anything other than a chance finding. There was insufficient evidence that the clients in this research get better because of concurrent medication taken.

Taken together these findings suggested it was unlikely cf. some criticisms that have been made of naturalistic research (Clark, et al., 2007), that:

- There were lots of ‘missing cases’ who were really treatment failures
- These findings were restricted to the ‘easiest’ clients in the research
- These findings were restricted to the ‘best’ outcomes
- These findings were simply ‘regression to the mean’
- These findings were simply ‘natural recovery’
- These findings were ‘attributable to concurrently administered medications’

It seemed plausible person-centred therapy in this research was at least partly responsible for the observed improvements.

Thus far the ‘further results’ section has looked at some criticisms of uncontrolled naturalistic research and considered: firstly, the impact of ‘missing clients’ for depression only for the reasons given; and secondly, other possible causes of recovery: (1) clients simply got better through the passage of time, or (2) clients got better because of concurrently administered medication, for each of depression, anxiety and distress. There appears to be little evidence to support these criticisms and we now turn to a third area of criticism, ‘lack of randomisation to different treatments’ in the next section, followed by the fourth and final area ‘no evidence that treatments appropriately delivered’ before looking at further results for the mediation part of this research.

7.1.3 Lack of randomisation to different treatments

There was no intention that clients would be randomised to different treatments in this research. Stiles, et al., (2008) wrote ‘The best predictor of post-treatment scores is usually pre-treatment scores on the same measure’ (p. 2) and the starting scores for clients in this research were comparable with other studies, e.g. clients in this study with a subsequent CORE-OM score ($n = 79$) started with CORE-OM scores of mean 18.4 cf. mean 17.6 for Stiles, et al. (2007), mean 17.4 for Stiles et al. (2006), mean CORE equivalent 14.5 for Deale, Chalder, Marks, & Wessely (1997) and mean CORE equivalent 17.1 for Ehlers, et al. (2003).

In these articles the transformations from BDI to CORE-OM were made using Leach, et al. (2006) by Stiles, et al. (2008). In general terms clients in this study had similar mean pre-treatment scores to the benchmark studies with which they were compared, although there were some differences in standard deviations. Whilst there was no randomisation to different treatments, clients in this research were generally comparable with other studies on the most important metric; pre-treatment scores, see discussion.

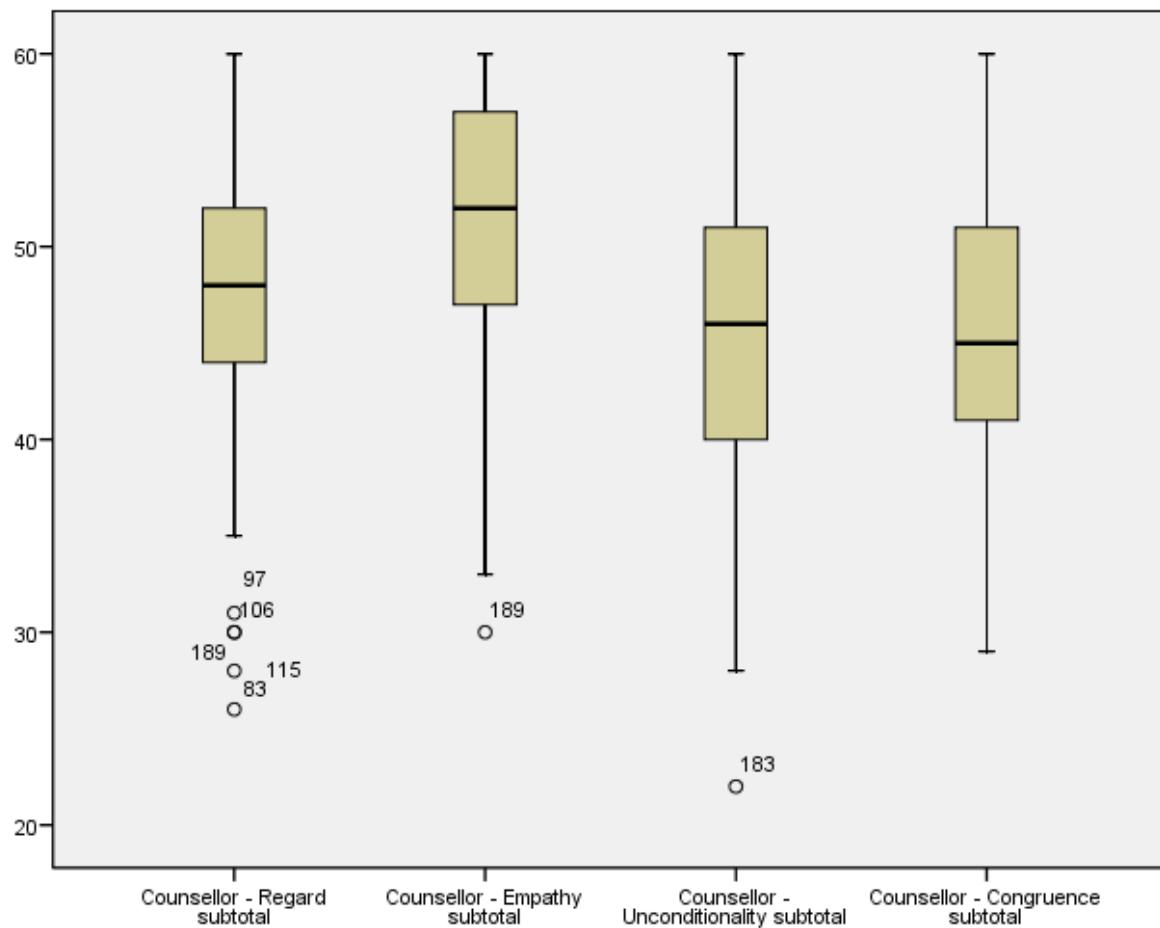
7.1.4 No evidence that treatments appropriately delivered

In their rejoinder to Clark, et al. (2007), Stiles, et al. (2008) commented that 'The magnitude of the pre-post improvements suggests the quality [of the treatments delivered] was adequate' (p. 3). This was a naturalistic study of person-centred therapy at the PP of a recently qualified person-centred therapist and at the UCS where therapists were a combination of students studying for a diploma and post-qualification therapists. There was no treatment manual, (other than the published theory statements Rogers (e.g. 1957, 1959), no research specific training in person-centred therapy (other than 'usual training' which the therapists had or were undertaking), and no independent checks on treatment fidelity or competence (other than 'supervision as usual'). It was interesting to note that whilst it was believed that 'PCT [person-centred therapy] is fairly easy to deliver within the constraints of primary care but CBT is not.' (Clark, et al., 2007, p. 4) At the PP, clients did have the option of being seen at a local (and for most clients more convenient) GP practice with which the author had a relationship. During the fieldwork for this research stretching out to nearly five years, no client asked to be seen at the GP practice and all preferred to be seen at the author's PP (and private residence). Clients appeared to prefer to be seen at the private residence rather than the GP practice; whilst this was only weak evidence it does challenge an assumption that 'PCT is fairly easy to deliver within the constraints of primary care...' if clients would prefer to be seen at a private residence than a GP practice.

For those clients who completed a BLRI this was an opportunity for clients to rate their perception of their therapist's congruent empathy and unconditional positive regard. Rogers (1967) noted that client perceptions differed from observers and the theory was about 'client perception' not about 'independent verification' (Rogers, 1957, 1959) and the impact this has on 'constructive personality change'. There were 118 client completions of BLRI with mean score 189.71 (out of maximum 240) with a range of 126 to 240 and a standard deviation of

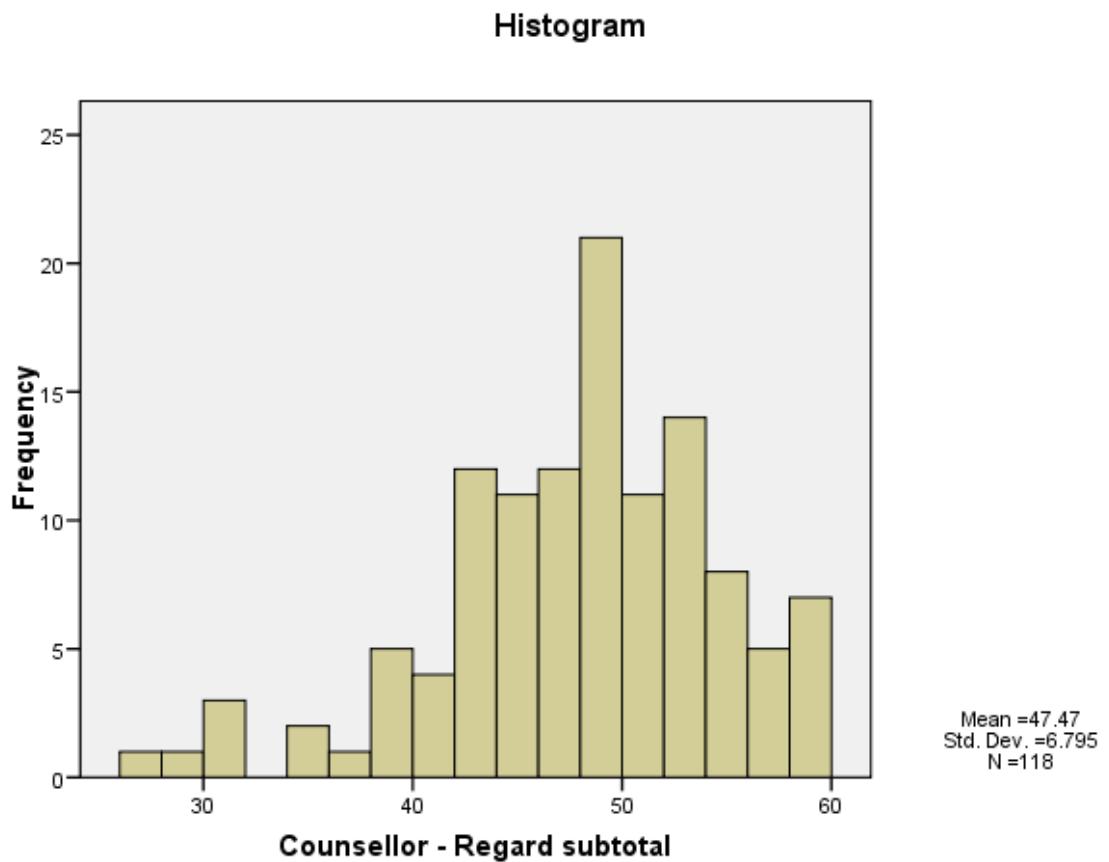
21.6. Distribution of BLRI scores was not significantly different from normal K-S D(118) = .04, $p > .2$. Scores on the BLRI subscales, regard, empathy, unconditionality and congruence had a maximum score of 60 and a box plot, see Figure 37, shows the subscales scores.

Figure 37: Box plot of BLRI subscale scores for 118 clients who completed the BLRI in this research.



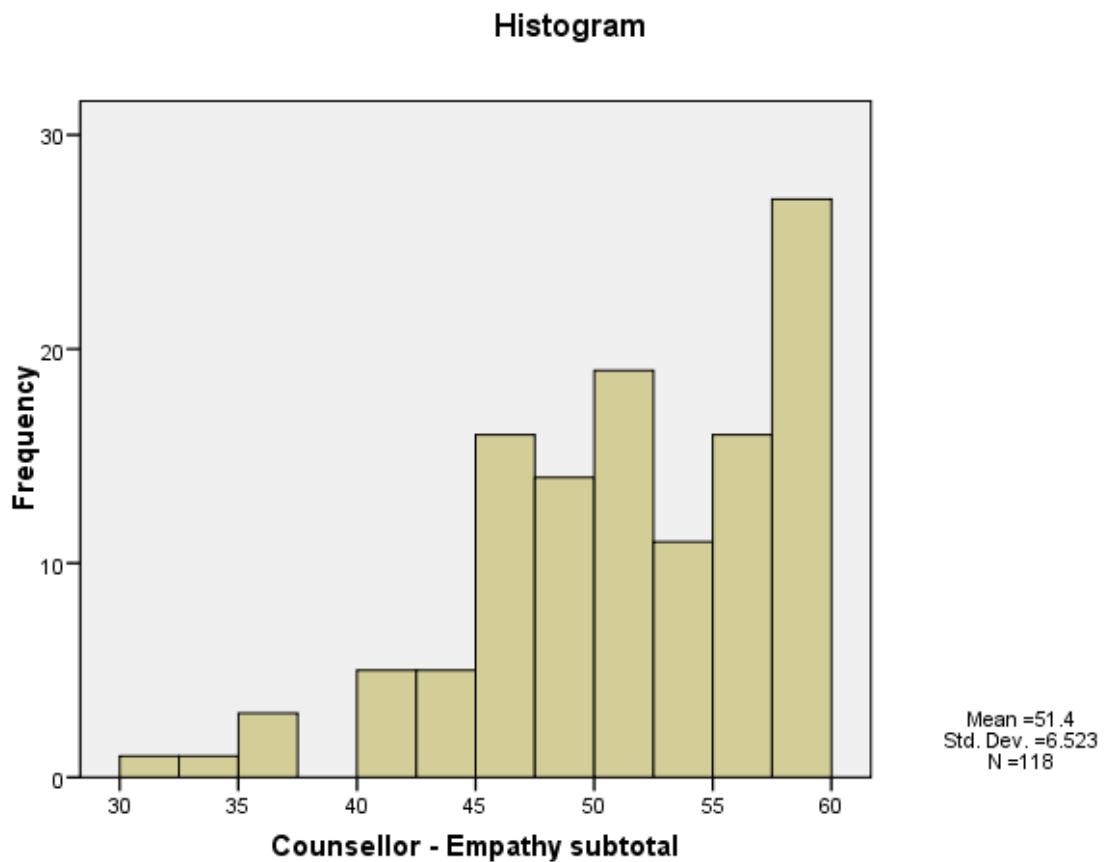
The mean score for 'regard' was 47.5 (SD 6.8, range 26-60) and this was distributed significantly different from normal, K-S D(118) = .09, p = .020, see Figure 38.

Figure 38: 'Regard' BLRI subscale scores: Distribution was significantly different from normal.



The mean score for 'empathy' was 51.4 (SD 6.5, range 30-60) and this was distributed significantly different from normal, K-S D(118) = .09, p = .013, see Figure 39.

Figure 39: 'Empathy' BLRI subscale scores: Distribution was significantly different from normal.



The mean score for 'unconditionality' was 44.0 (SD 7.9, range 22-60) and this was distributed not significantly different from normal, K-S D(118) = .06, p > .2. The mean score for 'congruence' was 45.4 (SD 6.9, range 29-60) and this was distributed not significantly different from normal, K-S D(118) = .05, p > .2.

The theory was that congruent empathy and unconditional positive regard provided the right conditions for therapeutic change and if the measures of these conditions showed an absence of these conditions this would challenge the theory. The BLRI scores provided some evidence that congruent empathy and unconditional positive regard were judged by clients to be to some extent provided. This suggested that to some extent the treatment conditions were provided and treatments may have been appropriately delivered, although there was no direct evidence from this research that these 'treatment conditions' predicted outcome, although pre-post improvements were observed.

7.2 Process-outcome research.

This section looks at some possible shortcomings to the process-outcome research. In particular the literature review of previous process-outcome studies suggested it was likely the therapeutic relationship could be a significant predictor of outcome. This research found no statistically significant evidence that therapeutic relationship did predict outcome and this section considers what, if anything, could account for this finding; whilst bearing in mind it could be the case that the therapeutic relationship was not a predictor of outcome.

This section examines in more detail the analyses for the depression, anxiety and distress process-outcome data for prediction of outcome by therapeutic relationship. This section presents those further results together with some related analyses, firstly, the extent to which the full range of client-perceived therapeutic relationships were tested by these analyses and subsequently possible sources of within therapist variation for client perceptions of the therapeutic relationship.

7.2.1 The range of client perceptions of the therapeutic relationship tested in this research.

In his criticism of process-outcome research using correlation Stiles (1988) pointed out that 'most investigations of psychotherapy process have used correlations, probably because conducting clinical trials of isolated process components is impractical' (p. 31). Furthermore Stiles described deliberately preventing therapists from providing a process component whilst holding other process components constant and concluded that 'such destructive manipulations are ethically problematic' (p. 32).

As naturalistic research this research measured client perceptions of the therapeutic relationship (BLRI) for qualified and trainee psychotherapists and this limited the range of BLRI scores available for the correlation analysis. The mean BLRI score was 190.3 (SD 21.3, variance 452.7) on an available range of zero to 240. In practice the observed range in the regression analyses was much narrower than the fully available range and was 126 to 237, occupying 46.3% of the available range. Clients saw therapists as operating above the midpoint of the BLRI scale. Therefore the findings in the process-outcome correlation analyses were limited to the situation where clients saw therapists as operating above the

mid-point of the BLRI scale and this data did not permit analysis of the situation where clients saw therapists as offering relationships scored below the mid-point of the BLRI scale.

As Stiles pointed out there could be some ethical difficulties in deliberately offering therapeutic relationships intended to be 'poor' in terms of therapeutic relationships likely to score below the mid-point of the BLRI scale. To that extent there was, and could be in future research, some difficulty in fully testing the impact of the therapeutic relationship when 'reasonably good' relationships are only practically and ethically capable of being tested. It is therefore only possible to speculate on the impact of relationships scored below the mid-point of the BLRI scale. For example there are a number of possibilities, including: 1) there could be a linear relationship between BLRI and outcome all the way to the origin of the scale (BLRI score = zero), or 2) there could be some minimal level of BLRI score required for therapy to take place. The second of these was a part of Rogers' 1957 theory 'The communication to the client of the therapist's empathic understanding and unconditional positive regard is to a minimal degree achieved' (p. 96). Rogers' approach to researching this 'sixth condition' was to observe whether a client indicated that:

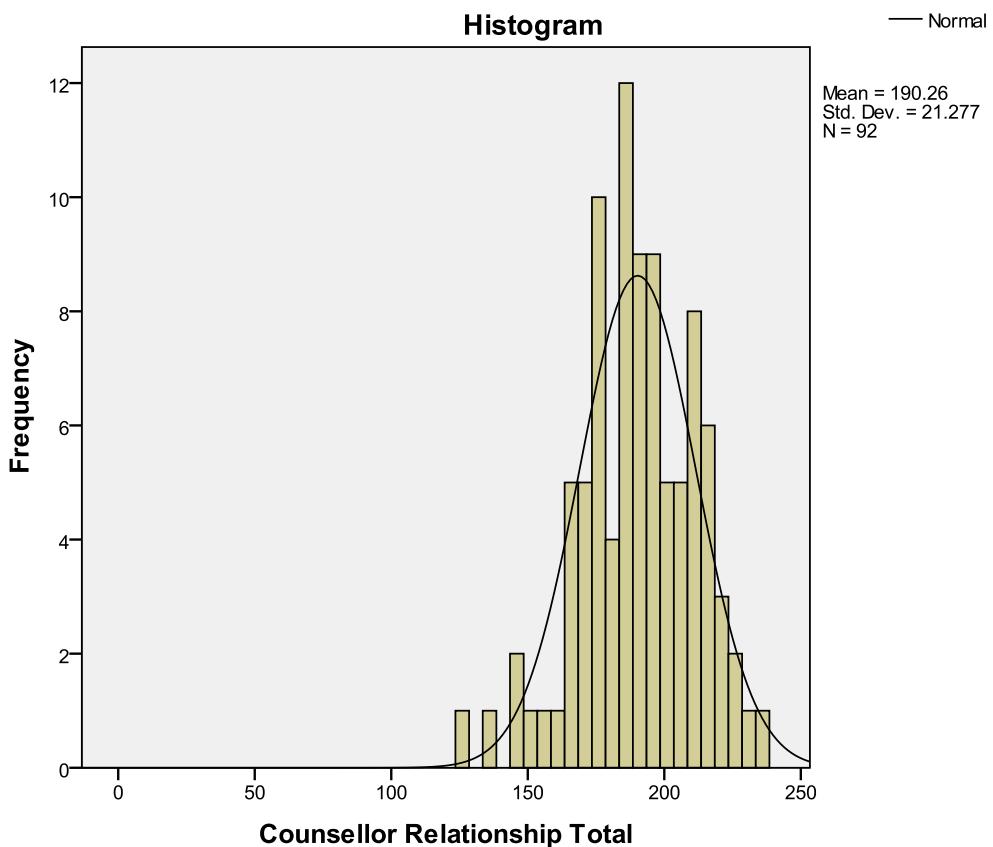
... several items descriptive of acceptance and empathy are sorted by the client as characteristic of the relationship, then this condition could be regarded as met. In the present state of our knowledge the meaning of "to a minimal degree" would have to be arbitrary' (p. 99)

It seems likely that contemporary researchers would not accept Rogers' idea that simply observing the client's recognition of these conditions as sufficient evidence of causation e.g. Elliott (2010).

In this research the distribution of BLRI scores was not significantly different from a normal distribution, K-S (D) = .05, p > .2. The lowest BLRI score of 126 was still higher than the mid-range of the scale and to that extent this research did not measure the effects of therapists perceived as being un-empathic, incongruent, conditional and offering negative regard. If therapists were performing at a fairly high level then there may not be a correlation between therapeutic relationship and outcome (D. Cramer, personal communication, 15th February 2011). This point is illustrated by Figure 40 wherein the distribution of BLRI scores has been presented along the full zero to 240 range, showing it was only a relatively tight bunch of higher-end BLRI scores that were tested by this research in practice. The properties of a normal distribution suggested two-thirds of therapeutic relationships were scored in the range

167-212 BLRI points and by definition clients needed to have a post-test to be included in the analysis, suggesting the likelihood that some 'minimal degree' of a relationship may have been attained.

Figure 40: Distribution of BLRI scores for the process-outcome research.



Note: BLRI scores shown for the depression process-outcome data as this was the largest of the three process-outcome datasets with $n = 92$.

The following sections consider the depression, anxiety and distress process-outcome analysis in some further detail.

7.2.2 The depression process-outcome data.

This section details an iterative process whereby six depression process-outcome analyses were conducted to examine individual cases as putative outliers or unduly influential cases. With an alpha level of .05 and a literature review that suggested the therapeutic relationship could be a predictor of outcome there was a risk of a Type II error. The potential error was that with a relatively small sample size to detect small-medium effects we could erroneously conclude that there was no effect within the dataset when there was an effect in the population, i.e. rejecting an effect in the dataset that exists in the population.

Multiple regression results were presented (Section 6.2.1) for the depression process-outcome data with $n = 92$ and for $n = 90$ when two cases with standardised residuals ≥ 3 were removed. Further examination identified four further cases that were outliers or leverage cases, according to the criteria described above, and these were deleted, see Table 41.

Whilst there was a negative finding for prediction of outcome (BDI-II at subsequent session) by the therapeutic relationship (BLRI) whilst controlling for depression at the start (BDI-II at first session) with the full complement of 92 cases, sequential removal of 6 cases (6.5% of the cases) on the basis that they were outliers or influential cases, resulted in a finding of prediction of depression outcome by the therapeutic relationship for analysis 6 ($n = 86$) with a small-medium effect size of $r_{\text{partial}} = .22$, $p = .047$, accounting for 4.7% of outcome variance, broadly in line with the literature. Depression at the start (pre-test) predicted subsequent depression, $r = .33$, $p = .002$.

In terms of temporal precedence (Elliott, 2010) depression outcome was assessed mean 4.6 sessions (SD 7.0) after assessment of the therapeutic relationship. In two cases the relationship measure was after the depression outcome (by 1 and 3 sessions), in 15 cases depression outcome and relationship were assessed at the same session and the remaining 69 cases (80.2%) had depression outcome measured after the relationship by between 1 and 43 sessions.

Table 41: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Depression Score (BDI-II) at First Session in Predicting Depression Score at Subsequent Session.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
n = 92 df (2, 89)								
1 (1)	Dep Subseq	Dep First	.373	.081	.439			<.001
(2)		BLRI	-.025	.040	-.059			.539
						.20	11.437	<.001
n = 91 df (2, 88)								
2 (1)	Dep Subseq	Dep First	.270	.079	.345			.001
(2)		BLRI	-.028	.037	-.075			.455
						.13	6.71	.002
n = 90 df (2, 87)								
3 (1)	Dep Subseq	Dep First	.250	.075	.337			.001
(2)		BLRI	-.036	.035	-.103			.844
						.14	6.81	.002
n = 89 df (2, 86)								
4 (1)	Dep Subseq	Dep First	.314	.075	.410			<.001
(2)		BLRI	-.039	.033	-.115			.244
						.20	10.39	<.001
n = 87 df (2, 84)								
5 (1)	Dep Subseq	Dep First	.267	.076	.350			.001
(2)		BLRI	-.074	.036	-.206			.042
						.19	10.12	<.001
n = 86 df (2, 83)								
6 (1)	Dep Subseq	Dep First	.234	.075	.319			.002
(2)		BLRI	.070	.035	-.205			.047
						.17	8.48	<.001

Note: DV = dependent variable. Dep Subseq = Depression at Subsequent Session (BDI-II). Dep First = Depression at First Session (BDI-II). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory.

7.2.3 The anxiety process-outcome data.

Multiple regression results were presented (Section 6.2.2) for the anxiety process-outcome data with $n = 75$ and for $n = 73$ when two cases with standardised residuals ≥ 3 were removed. No significant effect for therapeutic relationship was found.

It was observed that there was a statistically significant relationship between anxiety outcome and site (section 6.1.2), so site (PP or UCS) was included as a covariate in the further results regression analysis, for $n = 75$, $r_{\text{partial}} = .42$, $p < .001$. There was no significant relationship between site and depression outcome.

Further analysis showed that residuals in the multiple regression analysis were significantly different from normally distributed K-S $D(75) = .11$, $p = .021$. The decision was taken to use the transformed data, the square root of the sum of anxiety score plus one, and this gave a residuals distribution not significantly different from normal K-S $D(75) = .06$, $p > .2$.

One case was identified as an influential case and this case was deleted. There was no evidence that therapeutic relationship predicted subsequent anxiety level, see Table 42, and this was the case when number of sessions was included too (see 7.2.4).

Anxiety at the first session was a significant ($p < .001$) predictor of anxiety at the subsequent session with a large effect in the first analysis ($n = 75$), $r = .54$ and in the second ($n = 74$), $r = .51$. Site was also a significant predictor of subsequent anxiety with a large effect in the first analysis $r = .43$ and in the second $r = .42$. It seemed plausible that the presence of a covariate (site) may have left little 'room' for the effect of the therapeutic relationship in a sample of this size to appear as a statistically significant predictor with unique variance, rather than possibly variance confounded with site and pre-test (Field, 2009, pp. 397-9) and see below.

Table 42: Summary of Regression Analysis for Therapeutic Relationship (BLRI) and Transformed Anxiety Score (BAI) at First Session in Predicting Transformed Anxiety Score at Subsequent Session – Controlling for Site (PP versus UCS).

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
n = 75 df (3, 71)								
1 (1)	Anx Subseq	Anx First	.527	.098	.530			<.001
(2)		Site	.911	.226	.399			<.001
(3)		BLRI	.001	.005	.015			.876
						.35	12.62	<.001
n = 74 df (3, 70)								
2 (1)	Anx Subseq	Anx First	.523	.107	.509			<.001
(2)		Site	.906	.231	.406			<.001
(3)		BLRI	.001	.005	.014			.891
						.31	10.46	<.001

Note: DV = dependent variable. Anx Subseq = Transformed Anxiety at Subsequent Session (BAI). Anx First = Transformed Anxiety at First Session (BAI). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory. Site = Private Practice (PP) or University Counselling Service (UCS).

Finding no statistically significant relationship between the therapeutic relationship and subsequent anxiety a further iterative process was followed to identify those cases that did not fit a structure wherein the therapeutic relationship was a predictor of outcome. This was to assess the potential size of a putative partial correlation between therapeutic relationship and anxiety outcome that could be considered for sample size estimation for future research and also to identify cases that 'do not fit' and possible reasons for this (qualitative analysis of quantitative data). This analysis (Table 43) was approached by sequentially deleting the case with the highest positive standardised DFBETA for BLRI in the regression analysis. This identified the case most influencing BLRI from being a statistically significant predictor (Field, 2009, page 218).

This sequential deletion took the analysis from n = 75 to n = 68 when BLRI $r_{partial} = -.25$, $p = .046$, anxiety at first session $r_{partial} = .58$, $p < .001$ and site $r_{partial} = .52$, $p < .001$. Regression coefficients for each step of the analysis are shown in Table 43.

Table 43: Summary of Regression Analysis for Therapeutic Relationship (BLRI), Site and Transformed Anxiety Score (BAI) at First Session in Predicting Transformed Anxiety Score at Subsequent Session: Deleting highest SDBETA BLRI sequentially.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
n = 75 df (3, 71)								
1 (1)	Anx Subseq	Anx First	.527	.098	.530			<.001
(2)		Site	.911	.226	.399			<.001
(3)		BLRI	.001	.005	.015			.876
						.35	12.62	<.001
n = 74 df (3, 70)								
2 (1)	Anx Subseq	Anx First	.532	.096	.544			<.001
(2)		Site	.877	.222	.389			<.001
(3)		BLRI	-.002	.005	-.047			.629
						.36	13.07	<.001
n = 73 df (3, 69)								
3 (1)	Anx Subseq	Anx First	.550	.096	.560			<.001
(2)		Site	.934	.223	.411			<.001
(3)		BLRI	-.004	.005	-.074			.443
						.38	14.08	<.001
n = 72 df (3, 68)								
4 (1)	Anx Subseq	Anx First	.554	.094	.575			<.001
(2)		Site	.895	.218	.399			<.001
(3)		BLRI	-.006	.005	-.103			.281
						.39	14.72	<.001
n = 71 df (3, 67)								
5 (1)	Anx Subseq	Anx First	.542	.095	.557			<.001
(2)		Site	.920	.220	.409			<.001
(3)		BLRI	-.007	.005	-.125			.197
						.40	14.64	<.001
n = 70 df (3, 66)								
6 (1)	Anx Subseq	Anx First	.548	.093	.565			<.001
(2)		Site	.981	.220	.434			<.001
(3)		BLRI	-.009	.005	-.152			.116
						.42	15.90	<.001
n = 69 df (3, 65)								
7 (1)	Anx Subseq	Anx First	.535	.092	.556			<.001
(2)		Site	1.027	.217	.457			<.001
(3)		BLRI	-.010	.005	-.178			.066
						.44	16.70	<.001
n = 68 df (3, 64)								
8 (1)	Anx Subseq	Anx First	.522	.092	.541			<.001
(2)		Site	1.055	.218	.469			<.001
(3)		BLRI	-.011	.005	-.195			.046
						.44	16.82	<.001

Note: DV = dependent variable. Anx Subseq = Transformed Anxiety at Subsequent Session (BAI). Anx First = Transformed Anxiety at First Session (BAI). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory. Site = Private Practice (PP) or University Counselling Service (UCS).

Characteristics of cases that did not fit the structure wherein therapeutic relationship was a predictor of outcome together with possible reasons for 'lack of fit' are shown in Table 44.

Table 44: Characteristics of cases that did not fit the anxiety predictor structure, together with possible reason(s) for non-fit.

Case	BAI first	BAI subs ^a	Sessions	BLRI	Possible reason(s) for 'did not fit'.
33	15	12	5	213	Appeared to overscore BLRI relative to start anxiety. Relatively poor outcome despite relatively high BLRI.
57	9	9	6	223	Appeared to overscore BLRI relative to start anxiety. Relatively poor outcome despite relatively high BLRI.
95	32	13	4	237	Appeared to overscore BLRI relative to start anxiety. Premature ending - Relatively poor outcome despite relatively high BLRI.
115	23	0	12	126	Disproportionately good outcome. Appeared to underscore BLRI relative to start anxiety. Relatively good outcome despite relatively low BLRI score.
150	10	1	6	169	Appeared to underscore BLRI relative to start anxiety. Relatively good outcome despite relatively low BLRI score.
173	8	3	3	166	Appeared to underscore BLRI relative to start anxiety. Relatively good outcome despite relatively low BLRI score.
291	13	22	8	218	Premature ending - Disproportionately poor outcome. Appeared to overscore BLRI relative to start anxiety. Premature ending. Relatively poor outcome despite relatively high BLRI.

Note: ^a BAI at subsequent session.

Examining those cases that 'did not fit' suggested there was at least one variable that was perhaps influencing how clients scored the perceived therapeutic relationship, beyond the therapeutic relationship, e.g. cases 95 and 115 had widely differing views of the same therapist. This may have been an interaction effect between a consensual view of the capabilities of a particular therapist, in terms of their congruent empathy and unconditional positive regard and how particular clients scored their perceived relationship; perhaps something related to what Baldwin et al. (2007) referred to as the 'critical' or 'complimentary' client.

Beyond this there may have been at least one other factor that influenced the extent to which certain clients achieved relatively good outcomes even though the relationship had been scored at a relatively low level, and vice versa, e.g. cases 115, 150 and 173 versus 33, 57, 95 and 291. Given the research evidence for 'exposure' as an effective intervention for anxiety (Castonguay & Beutler, 2006) there may have been differences in the way that therapists 'empathised with anxiety'; either empathising with the feelings created by the troublesome or feared situation, thereby 'exposing' the client to the 'difficult' feelings in a person-centred

manner or simply empathising with the client's discomfort, without 'exposing' the client to the 'difficult' feelings.

In terms of temporal precedence (Elliott, 2010) anxiety outcome was assessed mean 4.8 sessions (SD 6.3) after assessment of the therapeutic relationship. In two cases the relationship measure was after the anxiety outcome (by 1 and 2 sessions), in 11 cases anxiety outcome and relationship were assessed at the same session and the remaining 55 cases (80.9%) had anxiety outcome measured after the relationship by between 1 and 32 sessions.

7.2.4 The distress process-outcome data.

Multiple regression results were presented (Section 6.2.3) for the distress process-outcome data with $n = 54$ and no statistically significant effect was found for the client-perceived therapeutic relationship.

It was observed there was a statistically significant relationship between distress outcome and time (Table 40) and number of sessions was included as a covariate in the regression analysis, with $r_{\text{partial}} = .39$, $p = .004$. There was no significant relationship between time and depression or anxiety outcome, and site (unlike the anxiety dataset) was not a predictor for distress outcome.

Outlier and influence statistics were analysed and the analysis was iterated with sequential deletion of five problematic cases (Table 45). The sixth iteration of the analysis ($n = 49$) found no outlier or influential cases that exceeded the cut-offs for these statistics and the analysis was halted at this stage.

For $n = 49$ there was no evidence that the therapeutic relationship significantly predicted the outcome ($r_{\text{partial}} = -.09$, $p = .543$) when controlling for the start level of distress ($r = .68$, $p < .001$) and the number of sessions ($r = -.32$, $p = .028$). Therapeutic relationship shared some collinearity with the two other large and medium-sized effects and this perhaps, in a manner similar to the anxiety data, may have left little 'room' for this to be a significant predictor.

Table 45: Summary of Regression Analysis for Therapeutic Relationship (BLRI), Number of Sessions and Distress Score (CORE-OM) at First Session in Predicting Distress Score at Subsequent Session: Deleting outliers and influential cases.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
n = 54 df (3, 50)								
1 (1)	Distress Subseq	Distress First	.500	.100	.622			<.001
(2)		BLRI	-.004	.031	-.017			.889
(3)		Sessions	-.289	.096	-.358			.004
						.37	9.77	<.001
n = 53 df (3, 49)								
2 (1)	Distress Subseq	Distress First	.502	.103	.636			<.001
(2)		BLRI	-.005	.032	-.018			.886
(3)		Sessions	-.297	.125	-.299			.022
						.35	8.69	<.001
n = 52 df (3, 48)								
3 (1)	Distress Subseq	Distress First	.492	.104	.610			<.001
(2)		BLRI	-.013	.033	-.049			.697
(3)		Sessions	-.357	.141	-.317			.015
						.36	8.79	<.001
n = 51 df (3, 47)								
4 (1)	Distress Subseq	Distress First	.556	.100	.673			<.001
(2)		BLRI	-.025	.031	-.093			.431
(3)		Sessions	-.405	.133	-.361			.004
						.44	12.26	<.001
n = 50 df (3, 46)								
5 (1)	Distress Subseq	Distress First	.590	.103	.713			<.001
(2)		BLRI	-.020	.031	-.076			.515
(3)		Sessions	-.482	.148	-.395			.002
						.45	12.67	<.001
n = 49 df (3, 45)								
6 (1)	Distress Subseq	Distress First	.625	.101	.718			<.001
(2)		BLRI	-.018	.030	-.069			.543
(3)		Sessions	-.353	.156	-.256			.028
						.49	14.51	<.001

Note: DV = dependent variable. Distress Subseq = Distress at Subsequent Session (CORE-OM). Distress First = Distress at First Session (CORE-OM). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory. Sessions = Number of sessions until subsequent distress measure.

As with the anxiety data the decision was taken to search for the likely size of the effect if the therapeutic relationship were a statistically significant predictor and to identify cases that did not fit this structure. This analysis (Table 46) was approached by sequentially deleting the case with the highest positive standardised DFBETA for BLRI; the case most influencing BLRI from being a negative predictor (Field, 2009, page 218). This sequential deletion took the analysis from $n = 54$ to $n = 49$ when BLRI was a predictor of outcome $r_{\text{partial}} = -.29$, $p = .046$. Whilst number of sessions was a significant predictor in this analysis $r = -.40$, $p = .005$, distress at the start was not $r = .29$, $p = .052$ and this appeared to provide further evidence that a sample of this size was perhaps underpowered for several medium-large effects sharing some co-linearity.

The five cases that did not fit the 'forced fit' distress predictor structure appeared to score their perception of the therapeutic relationship at a more extreme level than the other cases and were one very low severity case with an unusually low score for the therapeutic relationship and four of the more severe cases which each had evidence of suffering some kind of trauma and identified problems included borderline process, sexual abuse, suicidal ideation, eating problems and obsessive-compulsive disorder.

In terms of temporal precedence (Elliott, 2010) distress outcome was assessed mean 3.8 sessions (SD 6.2) after assessment of the therapeutic relationship. In three cases the relationship measure was after the distress outcome (by 1,2 and 3 sessions), in 11 cases distress outcome and relationship were assessed at the same session and the remaining 35 cases (71.4%) had distress outcome measured after the relationship by between 1 and 32 sessions.

Table 46: Summary of Regression Analysis for Therapeutic Relationship (BLRI), Number of Sessions and Distress Score (CORE-OM) at First Session in Predicting Distress Score at Subsequent Session: Deleting highest SDBETA BLRI sequentially.

Model	DV	Predictor (s)	B	SE B	β	R^2	F ^a	p
n = 54 df (3, 50)								
1 (1)	Distress Subseq	Distress First	.500	.100	.622			<.001
(2)		BLRI	-.004	.031	-.017			.889
(3)		Sessions	-.289	.096	-.358			.004
						.37	9.77	<.001
n = 53 df (3, 49)								
2 (1)	Distress Subseq	Distress First	.412	.105	.539			<.001
(2)		BLRI	-.017	.031	-.017			.583
(3)		Sessions	-.252	.095	-.344			.011
						.30	6.86	.001
n = 52 df (3, 48)								
3 (1)	Distress Subseq	Distress First	.379	.107	.503			.001
(2)		BLRI	-.027	.031	-.115			.385
(3)		Sessions	-.244	.094	-.342			.012
						.29	6.50	.001
n = 51 df (3, 47)								
4 (1)	Distress Subseq	Distress First	.342	.113	.454			.004
(2)		BLRI	-.040	.034	-.164			.242
(3)		Sessions	-.234	.094	-.332			.017
						.28	6.18	.001
n = 50 df (3, 46)								
5 (1)	Distress Subseq	Distress First	.302	.115	.391			.012
(2)		BLRI	-.055	.035	-.226			.121
(3)		Sessions	-.275	.098	-.366			.007
						.30	6.70	.001
n = 49 df (3, 45)								
5 (1)	Distress Subseq	Distress First	.239	.120	.307			.052
(2)		BLRI	-.075	.037	-.306			.046
(3)		Sessions	-.285	.096	-.382			.005
						.32	7.06	.001

Note: DV = dependent variable. Distress Subseq = Distress at Subsequent Session (CORE-OM). Distress First = Distress at First Session (CORE-OM). BLRI = Therapeutic Relationship assessed by client using Barrett-Lennard Relationship Inventory. Sessions = Number of sessions until subsequent distress measure. .

7.2.5 The predictive ability of within and between therapist variability.

Baldwin, et al. (2007) and Crits-Christoph, et al. (2009) both found it was between client-rated therapist variability that best predicted outcome. That is, it was the consensus view of clients about a particular therapist that predicted outcome, not whether a particular client scored the therapist high or low. Whilst it might have been interesting to compare this finding with data collected in this research this was not possible because all but two therapists saw more than four clients. Most therapists saw only one client, one saw four and three saw 2-3 clients each. This resulted in a severely distorted regression model with large numbers of outliers and highly influential cases that meant it was not possible to analyse the impact of within and between therapist variation in BLRI scores on outcome.

7.2.6 Possible sources of client 'rating tendencies' for the therapeutic relationship.

In the depression process-outcome data the two leverage cases deleted were those with the highest and lowest BLRI scores. In the anxiety process-outcome data the 'non fit' cases were those that appeared to 'under' or 'over' score their therapist in terms of BLRI scores. In the distress process-outcome data the 'non-fit' cases appeared to score their therapist in BLRI terms at more extreme levels than other cases. It would seem important to seek to understand what, if anything, was leading clients to rate their therapist, in BLRI terms, in the way that they did as this does appear to impact upon the predictive ability of the BLRI.

It seemed plausible clients may rate their therapist based upon a combination of: 1) a consensual view of the therapist's ability to convey congruent empathy and unconditional positive regard, and 2) individual client's ways of rating their therapist. For example the therapist who received the most ratings ($n = 74$) was scored as mean BLRI score 190.3 (SD 21.6), perhaps the 'consensual view' of 74 clients as to this therapist's ability to convey congruent empathy and unconditional positive regard. However, the range of scores was 114 BLRI units; from BLRI score 126 to 240. Perhaps two-thirds of clients rated this therapist in the range 169 to 212. It seems important to understand, what it anything, was leading clients to rate this same therapist in such widely differing ways.

In their discussion Baldwin et al. (2007) speculated as to the source of client 'rating tendencies' (p. 850) and gave an example of the 'critical client' as opposed to the 'complimentary client' as potential sources of client variability in scoring the same therapist.

It might be that understanding client 'rating tendencies' could be important in gaining a more 'pure' view of the impact of the therapeutic relationship upon outcome, controlling for individual client rating tendencies.

With a reasonably large number of client ratings of the same therapist together with symptom scores for several outcome questionnaires at the start of therapy this offered an opportunity to look at some possible sources of client rating tendencies by looking at whether any questionnaire at the start of therapy predicted therapeutic relationship score (Table 47).

Table 47: Regression analysis of outcome measures at the start as a predictor of the therapeutic relationship (BLRI score): Possible sources of within therapist variability.

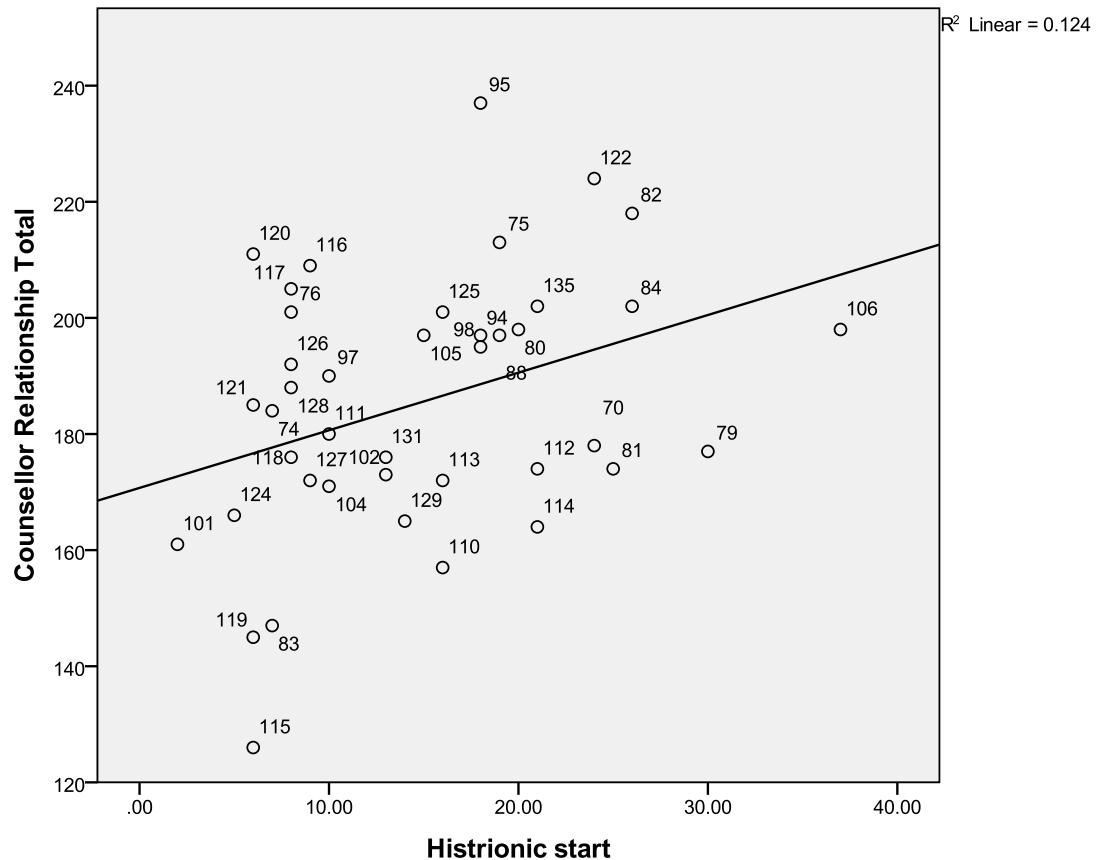
Predictor	n ^a	B	SE B	β	R ²	F ^b	p ^c
Depression	70	-.169	.260	-.079	.006	.423	.518
Anxiety	70	-.203	.262	-.093	.009	.598	.442
Distress	32	-.151	.179	-.153	.023	.717	.404
Avoidant	41	.272	.344	.126	.016	.624	.434
Dependent	41	.212	.379	.089	.008	.314	.578
Passive-Aggressive	41	.616	.417	.230	.053	2.182	.148
Obsessive-Compulsive	41	.473	.332	.222	.049	2.023	.163
Antisocial	41	.588	.828	.113	.013	.505	.482
Narcissistic	41	.707	.622	.179	.032	1.294	.262
Histrionic	41	.993	.423	.352	.124	5.514	.024
Schizoid	41	.173	.498	.055	.003	.120	.731
Paranoid	41	.123	.338	.058	.003	.133	.717
Borderline	41	.286	.394	.115	.013	.524	.473
Total PBQ score	41	.087	.061	.222	.049	2.029	.162

Note: ^a No outliers with standardised residuals ≥ 3 . ^b For depression (BDI-II) df = 1, 68. For anxiety (BAI) df = 1, 68. For distress (CORE-OM) df = 1, 30. For Personality Disorders (PBQ) df = 1, 39. ^c Two-tailed significance score.

The alpha level was not adjusted for multiple comparisons to avoid a Type II error in this exploratory analysis. Data suggested beliefs associated with histrionic personality disorder best predicted therapeutic relationship score for this therapist; a higher tendency toward 'attention-seeking' perhaps leading to a higher scoring, or 'over-estimation', of relationship conditions by the therapist (Figure 41). There was evidence for each of the predictors in Table 47 it was cubic lines that best fit the data, suggesting non-linear relationships between measures of client disturbance and BLRI scores, Crits-Christoph, et al., 2009 found a non-linear relationship between alliance scores and outcome rate, more extreme alliance scores were associated with poorer rate of improvement.

It seemed plausible that variables might be identified in future research that went at least some way towards accounting for client rating tendencies.

Figure 41: Histrionic personality disorder (PBQ-HI subscale score) as a predictor of client perception of the therapeutic relationship (BLRI score).

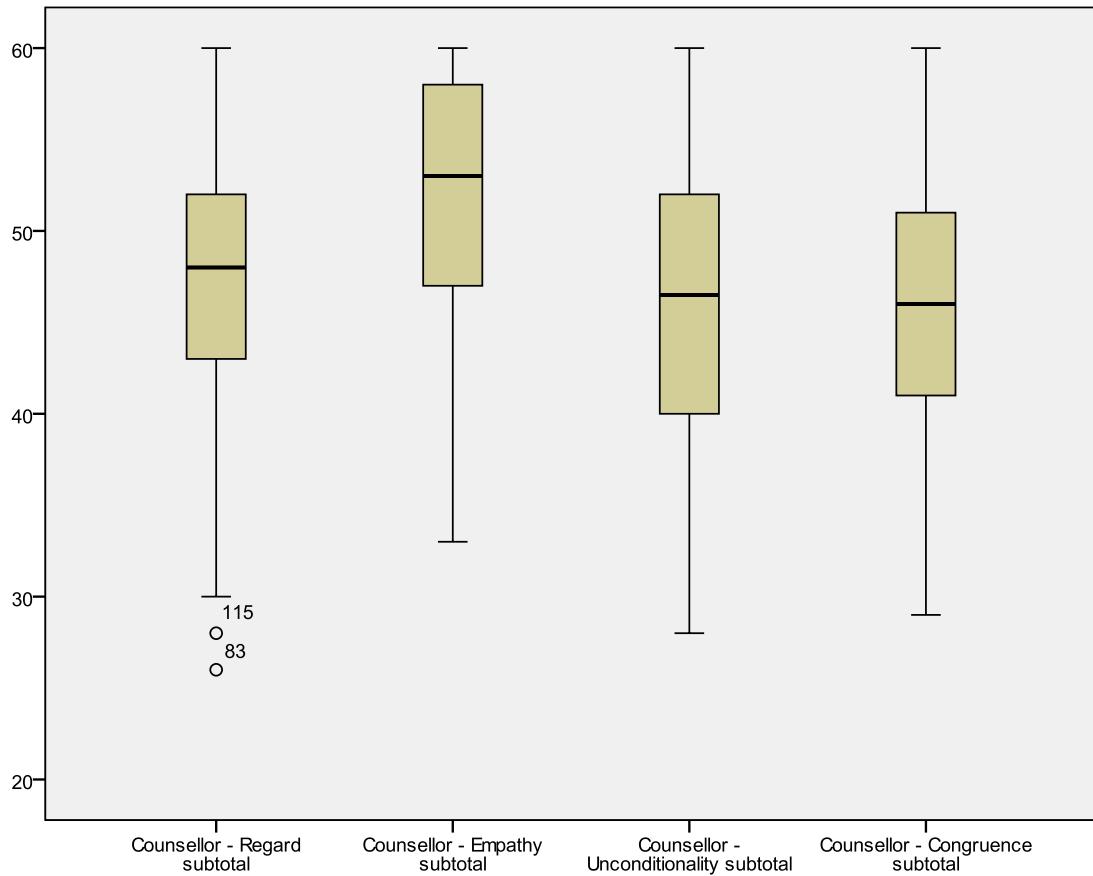


Note: n = 41.

In the same way that therapists working in an ‘outcomes management’ manner might share outcome scores with clients and seeking to be transparent the author shared all questionnaire results with clients. By so doing this provided some opportunity for qualitative discussion about the BLRI questionnaire and client ratings. For example a client with a very low score for the ‘regard’ subscale related a history wherein it seemed impossible to the client to consider another person might hold positive regard for the client. Similarly a client with a very low score for the ‘unconditionality’ subscale related a history of extreme conditionality wherein it seemed impossible to the client to consider anything other than extreme

‘conditionality’. A box plot of the subscales for the BLRI scores with the author is provided, see Figure 42.

Figure 42: Box plot of therapeutic relationship (BLRI) subscale scores.



Note: n = 74. Subscales for single therapist BLRI scores.

This pointed to an apparent contradiction in Rogers' 1957 theory. Rogers described condition two, the client's incongruence, as 'it refers to a discrepancy between the actual experience of the organism and the self picture of the organism' (p. 96) i.e. that client's, by definition, could not be relied upon to accurately report their experience, e.g. by accurately scoring the BLRI in this case. However, when it came to determining condition six, Rogers wanted the client to determine whether they perceived that the therapist had acceptance and empathy for them. Rogers defined 'incongruence' in a precise way as relating to the client's 'anxiety' and according to Rogers; incongruence would seem to be the source of 'client rating tendencies'.

Rogers himself became frustrated by the structure of his theory when he found that trained raters and 'incongruent' clients most closely agreed upon the therapeutic conditions present, more so than the supposedly 'congruent' therapists who had a 'clear divergence of... views of the relationship from the views of both judges and patients' and 'the therapists... were divergently optimistic in their assessments of the patient-therapist relationship' (1967, p. 183).

Subsequent research generally suggested that client perspectives on the therapeutic alliance were better predictors of outcome, although more recent studies have suggested that therapists' assessments later in therapy become more predictive of outcome and the degree of similarity between therapist and client rating in mid- and later phases of therapy were related to outcome (Horvath & Bedi, 2002). Considering the predictive capability of client-rated process of therapy points to both the process of therapy being rated as a predictor and also the way in which clients vary in their rating tendency, which is not yet fully understood.

In summary, further results for the three process-outcome datasets showed a statistically significant effect for therapeutic relationship on subsequent depression when controlling for start depression and illustrative findings for the effect of relationship on subsequent anxiety and distress when controlling for start severity plus number of sessions and site, respectively, see Table 48.

Table 48: Summary of further results for impact of therapeutic relationship on outcome: Significant effect on depression and illustrative results for anxiety and distress.

	BLRI ^a		Start severity ^b		Covariate ^c		% dataset m/n = % ^d
	r _{partial}	p	r _{partial}	p	r _{partial}	p	
Depression	-.22	.047	.33	.002	-	-	86/92 = 93.5%
Anxiety ^e	-.25	.046	.58	<.001	.52	<.001	68/75 = 90.7%
Distress ^e	-.29	.046	.29	.052	.40	.005	49 /54 = 90.7%

Note: a Client-perceived therapeutic relationship. BLRI = Barrett-Lennard Relationship Inventory. ^b Depression (BDI-II), transformed anxiety (BAI) and distress (CORE-OM) respectively. ^c Covariate – there was none for depression, site (PP or UCS) was a covariate in the anxiety dataset and number of session was a covariate in the distress dataset. ^d Proportion of dataset, m cases in structure that fit BLRI as a predictor out of total n cases to start with, expressed as a percentage.

Analysis of 'non-fit' cases suggested there maybe other variables to consider, including client rating tendencies and other variables that may influence outcome, including differential therapist or client effectiveness at attaining 'good' outcomes.

8. Discussion

In this section main findings are summarised, some of the strengths and weaknesses of this research are considered, the results discussed and the implications of this research considered.

8.1 Summary of findings

Over a five year period a naturalistic sample of outcomes was created from clients seen by therapists, including the author, at the University of East Anglia (UEA) and the author's Private Practice (PP). This sample included a process measure, the Barrett-Lennard Relationship Inventory, for some of the clients. There were no inclusion or exclusion criteria for the clients invited to participate in this research, although criteria were used in creating the various subsamples for the analyses, details of which were provided. The 321 clients who started in the research were mean age 31 years, 66% were female, 62% were single, 67% were not parents, 88% were British and 84% were taking no relevant medication. There was no systematic data captured for client issues that might ordinarily exclude a client from a clinical trial, e.g. eating disorder, suicidality, psychosis, self-harming, drug or alcohol abuse etc. Amongst a subset of the clients seen at the PP it was estimated that perhaps 80-85% of clients may have had a personality disorder. The sample was clients seen in routine practice at both sites.

This research found, in terms of the hypotheses stated in the introduction, on average clients with 'caseness' undergoing person-centred therapy had statistically significant large improvements for symptoms of depression (ES 1.48), anxiety (ES 1.15) and distress (ES 1.80) and that the therapeutic relationship as defined by Rogers was a statistically significant small predictor of subsequent depression outcome ($r = .22$), once outlier and influential cases were considered. Illustrative analysis suggested the therapeutic relationship could be a small predictor for both anxiety ($r = .25$) and distress ($r = .29$).

This section sets out in more detail the main outcomes from this study for depression, then anxiety and finally distress and in particular compares these with the benchmark studies, in year of publication order, as this was an aim of this research.

8.1.1 Depression

In terms of depression outcomes, 162 clients started with a clinical level of depression, on an LOCF basis $ES = 1.02$ (95% CI .85 to 1.20); of these 111 clients had a subsequent measure of their depression (31.5% were ‘missing’) with $ES = 1.48$ (95% CI 1.28 to 1.68), 29.7% had no reliable change, 70.3% reliably improved (53.2% had recovered and 17.1% had reliable improvement only). These ‘subsequent measure’ clients were not necessarily ‘completers’ in the sense that while they may have had a subsequent measurement of their depression this did not necessarily mean they had ‘completed’ therapy. For these 111 clients, mean age was 36 years, 69% were female, 39% were single, 47% were not parents, 87% were British and 78% were taking no relevant medication. Clients were seen for a mean 8.6 sessions by 16 therapists and the author saw 73.9% of these clients.

Using transformed scores to correct for a distribution of depression scores at initial measurement that was significantly different from normal gave a slightly higher $ES = 1.87$ ($n = 111$) and a slightly higher figure for reliable improvement, 71.2% of clients. The rest of this section considers data from untransformed scores.

Comparing those clients that started with a clinical level of depression ($n = 162$) and those that also completed a subsequent measurement of their depression ($n = 111$) there were 51 ‘missing cases’. Each of these 51 cases was examined and classified with the information available into one of four groups, to attempt to account for their ‘missing’ status, i.e. attended only one session ($n = 21$), probably did improve according to another measure ($n = 11$), reasons for stopping therapy probably not related to lack of progress ($n = 3$) and possible treatment failures, although no direct evidence of this ($n = 16$). On this basis estimates were made for the impact on effect size and reliable change percentages of ‘missing cases’ and inclusion or exclusion of these groups meant mean estimates for effect size ranged from 1.02 to 1.48, and reliable change percentage estimates ranged from 48.1% to 70.3%.

A subset of 36 UCS clients had wait controlled outcomes and this showed there was no significant improvement during waiting and a significant improvement during treatment, suggesting for these clients, a wait controlled $ES = 1.26$ (95% CI .82 to 1.69). Treatment (Mdn 49 days) was significantly longer than wait (Mdn 21 days), so the dynamics of change were examined and there was no significant relationship between change in time and change in depression found during either wait or treatment conditions for this subset.

Comparing the wait controlled subset of clients with those clients for whom only waiting data or treatment data were available found no significant differences on start of period or end of period outcome measures, i.e. in terms of outcome measures the wait controlled clients were representative of those other clients who only waited or were only treated.

Clients who had a subsequent measurement of their depression were significantly more depressed at the start than those who did not have a subsequent measurement of their depression.

Clients who were taking relevant medication had no significant difference in outcome compared with those who were not taking medication. There was a significant interaction for 21 clients taking anti-depressants, compared with 86 clients who were not, suggesting a small effect ($r = .19$) for those on anti-depressants. It was not clear that it was the anti-depressants themselves that were the cause of this small effect, this apparent effect may have confounded those on anti-depressants who tended to have higher depression scores (Stiles, et al., 2008).

Clients who started with a clinical level of depression were divided into 52 clients with 'severe depression', LOCF ES = 2.15 (95% CI 1.62 to 2.68) of which 41 had a subsequent measure of their depression ES = 2.73 (95% CI 2.20 to 3.25) (22.0% had no reliable change, 78.0% reliably improved; 48.8% had recovered and 29.3% had reliable improvement only) and 110 clients with 'non-severe depression', LOCF ES = 1.62 (95% CI 1.28 to 1.95) of which 70 had a subsequent measure of their depression ES = 2.50 (95% CI 2.13 to 2.88).

In terms of comparing these outcomes with the studies chosen as benchmarks (see Introduction) a recommendation in the literature was that differences in effect size of .4 or greater are considered 'interesting' (Elliott, et al., 2004) on the basis that these are effects at least midway between 'small' and 'medium' effects (Cohen, 1988), although others, (Minami, et al., 2007) argued for effect sizes of .2 or greater as the criterion to be used. For this study .4 was used as the criterion. For this comparison of effect sizes the comparator study ES was subtracted from the present study ES, such that the following signs were used throughout: + = present study effect size was higher than the comparator study, - = present study effect size was lower than the comparator study.

The measures used in the comparator studies are identified and in the case where multiple measures were used the particular measure being compared with is identified. Research has shown that measures do matter, in the sense that some measures have been found to be more

responsive to client changes than others (Shadish & Sweeney, 1991, Westen, et al., 2004, Minami, et al., 2007). Furthermore, there was some evidence that some measures may be differentially responsive to different forms of treatment (Shadish & Sweeney, 1991).

In Ward, et al. (2000) the benchmark figures chosen for comparison were ES = 1.7738 for CBT completers, ES = 1.6162 for PCT completers and ES = 1.0449 for usual GP care, all after four months measured using BDI (Table 2 of reference). Comparing depressed clients with a subsequent measure of their depression in this study (n = 111, ES = 1.4848) the differences in ES were -.2890 for CBT, -.1314 for PCT and +.4399 for usual GP care. On this basis there was no 'interesting' difference between this study and CBT and PCT samples and an 'interestingly' better outcome than usual GP care in Ward, et al.

The benchmark study offered 6-12 sessions with mean number of sessions 5.0 for CBT, 6.4 for PCT (both over four months) and 9.2 GP consultations (over twelve months, the article appears to omit mean number of GP consultations over four months) compared with 8.5 sessions for this study. The benchmark study was targeted at those with depression or mixed anxiety and depression (BDI-II equivalent at start ≥ 16 , using the conversion table in Beck, et al., 1996, p. 26) and sought to exclude those clients with other psychiatric diagnoses or who were substance misusing, self harming or suicidal or on anti-depressants (although 49% of 'usual GP care' took antidepressants and 34% of this part of the sample were referred to 'a mental health professional'); the present study did not specify inclusion or exclusion criteria, other than start BDI-II ≥ 14 and a subsequent BDI-II measurement for this part of the sample and included clients that met each of the comparator study exclusion criteria.

Consequently clients in Ward, et al. (2000) tended to be slightly more depressed than in this study. Converting from BDI to BDI-II for comparability (Beck, et al., p. 26), mean cell start scores on a BDI-II basis were of the order 28-31 compared with 27 for this study. Sample SDs were also lower for the comparator cells at 8.4 - 4.9 compared with 9.6 for this study, lower SDs tend to increase ES (see 1.2.1 b). Overall this suggested depression outcome for this study was comparable with the CBT and PCT arms of Ward, et al. and perhaps better than 'usual GP care'. The slightly longer episodes of therapy in this study may have arisen because clients were not only 'working on depression' as was the intention in the comparator study or there may have been some significant but unknown difference between the two studies that accounted for this difference.

In Watson, et al. (2003) the differences in ES between this study and the CBT and PE completer groups were -.2571 and -.3346 respectively, both insignificant differences according to the criterion set. Watson et al. used a range of different measures (SCID IV, IIP, RSE, and SCL-90-R) and the measure used for comparison was BDI. This benchmark study was a fixed protocol of 16 sessions; much longer than the mean 8.5 sessions for this study. On a comparable basis (Beck, et al., 1996, p. 26) the benchmark study clients were slightly more depressed, mean cell start scores 27-29 cf. 27 in this study.

The meta-analysis of 23 studies of depressed clients by Elliott, et al. (2004) (which included Watson et al., 2003 and Ward, et al., 2000, although referenced as King, et al., 2000) reported $ES = 1.18$, trivially lower (+.3048) than clients with a subsequent depression measure in the present study. Whilst meta-analyses have a number of strengths it was much more difficult to be certain this was a fair comparison between this study and the 23 studies as necessarily much less information about the combined study sample was provided e.g. inclusion/exclusion criteria, severity of depression at start, treatment length etc.

The next study compared was that of Missirlian, et al. (2005) who set a relatively high inclusion criterion of a minimum equivalent score of 19 BDI-II points (cf. equivalent minimum 16 BDI-II points for Ward, et al. (2000) and minimum 14 for this part of the sample in the present study). Measures used by these authors included IIP, RSE, SCL-90-R and the measure chosen for comparison with this study was BDI. The mean start depression score for Missirlian, et al. was not that different to this study, 28 cf. 27 respectively, although this gave rise to differences in start SDs, 6.08 cf. 9.55 respectively. Furthermore 14 to 20 (MdN 17.5) weekly sessions of PCT or PE were offered compared with 8.5 sessions in the present study; mean end scores were slightly lower at 11 cf. 13 in this study. The difference in $ES = -1.053$, was on the face of it a large difference, although to what extent this was a 'real' difference in effectiveness was unclear. The change in depression scores was -14.18 BDI-II points this study and -15.43 BDI points in the comparator study. If the present study had been more tightly defined, perhaps with a comparable starting SD to the comparator sample (6.08 cf. 9.55 present study) there would have been a trivial difference of $ES = -.2055$ (using a hypothetical $ES = 2.3322$ for the present study, estimated using SD 6.08 instead of 9.55 to attempt to control for differences in variability). If a higher inclusion criterion had been set for the present study this difference may not have arisen, for example, comparing the 'severe depression' sample from this study with Missirlian, et al. led to a trivial difference of +.1935. The present study had no inclusion/exclusion criteria (other than start $BDI-II \geq 14$

and a subsequent BDI-II measurement for this part of the sample) and this comparator study sought to include only non-functionally impaired depressed people and excluded those with three or more previous episodes of depression, current drug or alcohol abuse, current eating disorder, antisocial or borderline personality disorder, bipolar or psychotic disorder, a past history of incest, recent suicide attempts, loss of a significant other in the past year or involvement in an ongoing violent relationship; clients with all of these issues were included in the present study. Whilst there may be a 'real' difference in effectiveness between these two samples, this apparent difference could also be due to methodological and statistical artefacts.

The present study was compared with Dimidjian, et al. (2006) who sought to test 'the efficacy of behavioural activation (BA) by comparing it with cognitive therapy (CT) and antidepressant medication (ADM) using BDI-II in a randomised placebo-controlled (PLA) design in adults with major depressive disorder (n = 241)' (pp. 658-659). In this comparator study clients were recruited mainly through the media, a substantial minority were referred from local agencies and the rest by word of mouth or other sources cf. the present study was clients presenting for counselling in routine practice. The comparator study had a number of exclusions such as those with BDI-II less than 20, a lifetime diagnosis of psychosis or bipolar disorder, substantial and imminent suicide risk, current drug or alcohol abuse or dependence, anorexia, bulimia, some personality disorders (antisocial, borderline or schizotypal), were known non-responders to cognitive therapy or the antidepressant used in the trial, had particular physical health problems or were pregnant, lactating or capable of becoming pregnant and not using 'suitable contraception' cf. the present study that did not exclude clients with these characteristics. The comparator study suffered with significant differences in rates of attrition; whilst relatively low rates of attrition were experienced for the therapy and placebo arms of the trial (BA 16.3%, CT 13.3%, and PLA 22.6%), 44.0% taking antidepressants did not complete the full acute phase of the protocol cf. the present study where 31.5% (51/162) of clients who started with a clinical level of depression did not have a subsequent measurement of their depression. Comparing those clients in the present study who started with BDI-II ≥ 14 and a subsequent measure of their depression with those 'lower severity' clients in the comparator study the relative ESs were BA -2.3757, CT -1.0609, ADM -2.0213 after sixteen weeks of the trial cf. mean 8.5 sessions in the present study. Whilst the mean starting BDI-II scores were similar between the studies BA 28.7, CT 27.3 and ADM 27.8 (cf. 26.9 the present study), because the comparator study divided clients

into 'lower severity' and 'higher severity' groups the standard deviations were somewhat different BA 4.59, CT 6.89 and ADM 5.67 (cf. 9.55 in the present study). If the present study had an SD similar to the BA cell (hence speculative $ES = 3.0893$ for the present study) the differences in ES would be changed to BA -.7711, CT +.5436 and ADM -.4167. If the comparison were made at eight weeks of the trial (cf. 8.5 sessions for $ES = 1.4848$ in the present study) the cell differences in ES would be reduced to BA -1.4324, CT -.5993, ADM -.9666 and PLA -.7085 (PLA was stopped at eight weeks). Combining these two phenomena, differences in SD and treatment length, the differences between the comparator study and the present study (at speculative $ES = 3.0893$) became trivial at BA +.1721 and 'interesting' at, CT +1.0052, ADM +.6379 and PLA +.8960. Comparing the 'severe depression' group from the present study (mean start BDI-II 36.9, $ES = 2.7313$) with the 'higher severity' group in the comparator study (mean start BDI-II BA 36.7, CT 34.1 and ADM 35.6) there were non-trivial advantages in active treatment at eight weeks (cf. mean 8.8 sessions this study) for BA -.6291 and at sixteen weeks for BA -1.9945 and ADM -1.1719. If the present study had been more tightly defined with an SD comparable to the BA cell of 5.91 (cf. 7.41 present study, hence speculative $ES = 3.4263$) there would have been no 'interesting' advantage in the comparator study at eight weeks (cf. mean 8.8 sessions in the present study) but there would have been apparent advantages in active treatments for the comparator cells at sixteen weeks, BA -1.2995 and ADM -.4769. The comparator study recruited patients with the intention of completing a sixteen week trial; there was no intended 'commitment' to the present study, although some clients had more than 16 sessions, some had fewer and some had particular limits on the numbers of sessions set e.g. by their attendance at UCS, by an employer, etc. Whilst there were apparent outcome advantages in the comparator study to some extent these can be set against what may be methodological and statistical artefacts e.g. inclusion/exclusion criteria, length and commitment to trial period, tightness of definition of sample populations, etc. On the other hand there may have been some real advantages in the comparator study which concluded that BA and ADM were particularly effective among more severely depressed patients. From a therapeutic point of view, if these findings were confirmed, there may be some merit in person-centred therapists emphasising the more 'behavioural' parts of their own practice (emphasising client's exposure to and awareness of context for their 'difficult' feelings) for more severely depressed clients. For example this could be achieved by bringing client's attention to (perhaps by empathising with) 'the relationship between [their] activity and mood and the role of contextual changes associated with decreased access to reinforcers that may serve an antidepressant function' (Dimidjian, et

al., p. 660). In a behavioural therapy this would include ‘self-monitoring, structuring and scheduling daily activities, rating the degree of pleasure and accomplishment experienced during engagement in specific daily activities, exploring alternative behaviours related to achieving participant goals and using role-play to address specific behavioural deficits’ (p. 660). Most or all of these ‘behavioural’ emphases can be accomplished from a person-centred/process-experiential perspective by empathising with the feelings clients’ experience through their lived experience and drawing connections between activities and mood and changes in each of these. This comparator study served to illustrate some of the difficulties in making comparisons between studies, the importance of head-to-head comparisons of therapies in researcher-allegiance balanced trials, the benefits that may come from such ‘dismantling’ studies and the means by which therapists may modify their approach based on research evidence.

The final comparator study was the meta-analysis of 29 studies by Minami, et al. (2007) to construct depression benchmarks for studies to compare with. These authors described benchmarks with a number of measures and this comparison was based on the results for BDI. Comparing clients in the present study with BDI-II ≥ 14 at the start and a subsequent measure of their depression with the ‘completers’ part of the comparator study (typical duration 15 weeks cf. 8.5 sessions this study) showed a mean difference of -.3742, a trivial difference using the standard criteria of .4 ES for an ‘interesting’ difference between studies (Elliott, et al., 2004) although Minami, et al. (2008) argued for a smaller minimum criteria of .2 ES. Contributions made by Minami, et al. (2007) and Minami, et al. (2008) were to demonstrate a technique (using non-central t) to statistically compare studies and to examine statistical evidence for five moderators that may influence outcome (initial severity, treatment type – CBT versus other, modality – individual versus group, weeks in treatment and sample size), concluding that initial severity and weeks in treatment were moderators when completer outcome was measured using a depression outcome measure. This seems like an important step forward in overcoming some of the difficulties highlighted above in making comparisons between studies.

8.1.2 Anxiety

In terms of anxiety outcomes, 156 clients started with a clinical level of anxiety, on an LOCF basis ES = .69 (95% CI .54 to .85); of these, 91 clients had a subsequent measure of their

anxiety (41.7% were 'missing') with $ES = 1.15$ (95% CI .95 to 1.35), none reliably deteriorated, 70.3% had no reliable change, 29.7% reliably improved (18.7% had recovered and 11.0% had reliable improvement only). 'Subsequent measure' clients were not necessarily 'completers'. For these 91 'subsequent measure' clients, mean age was 35 years, 64% were female, 45% were single, 53% were not parents, 88% were British and 78% were taking no relevant medication. Clients were seen for a mean 8.1 sessions by 16 therapists and the author saw 64.8% of clients.

Using transformed scores to correct for a distribution of anxiety scores at initial measurement that was significantly different from normal gave a slightly higher $ES = 1.56$ ($n = 91$) and a higher figure for reliable improvement, 46.2% of clients. The rest of this section considers data from untransformed scores.

A subset of 36 UCS clients had wait controlled outcomes and this showed there was a small and significant improvement during waiting ($ES = .18$) and a significant improvement during treatment ($ES = .57$), suggesting for these clients, a wait controlled $ES = .38$. Treatment (Mdn 49 days) was significantly longer than wait (Mdn 20.5 days), so the dynamics of change were examined. This suggested a significant improvement of 5.3 BAI units after the exploratory session and a gradual and significant deterioration as waiting progressed. After the first counselling session there was a significant improvement (4.7 BAI units) and there was no significant relationship between changes in time (days) and change in anxiety found during treatment for this subset. Unlike outcomes for depression, there was no evidence of a 'premature endings effect' at the UCS.

Comparing the wait controlled subset of clients with those clients for whom only waiting data or treatment data were available found no significant differences on start of period or end of period outcome measures, i.e. in terms of outcome measures the wait controlled clients were representative of those other clients who only waited or were only treated.

Clients who had a subsequent measurement of their anxiety were significantly more anxious at the start than those who did not have a subsequent measurement of their anxiety.

Clients who were taking relevant medication had no significant difference in outcome compared with those who were not.

These effects were compared with the studies chosen as benchmarks (see Introduction) with the criterion of a difference in effect size of .4 or greater considered 'interesting'.

Clients in the present study with start BAI ≥ 8 and a subsequent BAI score (ES(d) = 1.1569) were compared with the meta-analysis of Borkovec and Whisman (1996) for generalised anxiety disorder and only trivial differences emerged between both BT +.2569 and CBT +.1469, although as described above it was difficult to know the validity of a comparison with a meta-analysis of a group of studies, and see below re wait-controlled effect sizes. Borkovec and Whisman compiled studies that used the State-Trait Anxiety Inventory, Hamilton Rating Scale of Anxiety and Diagnostic assessors rating of GAD severity.

The meta-analysis of Gould, et al. (1997) sought studies only with controlled effect sizes, the vast majority of which were wait-list or attention placebo controlled. These authors did not identify the individual measures used to compile their overall evaluation. The present study identified statistically significant improvements for clients during waiting, following an exploratory session with an experienced therapist and on average this improvement deteriorated as waiting continued. Whilst it appeared that treatment offered only an incremental benefit to waiting, the dynamics of change were different during treatment, with no significant relationship between passage of time and improvement in anxiety. The picture was further complicated by the finding that whilst there was a significant improvement during waiting and treatment there was no significant difference between changes during waiting and those during treatment, perhaps because the sample was underpowered. This calls into question what is the correct comparison to make from the present study, the ES 1.1569 found for clients suffering with some level of anxiety and a subsequent measurement of their anxiety or whether this should perhaps be reduced by some factor to take account of the significant reduction found during wait (ES = .18), or perhaps the wait-controlled ES = .38. However, as waiting offered only a 'small' improvement this was put to one side for the time being. The comparator meta-analysis found controlled effect sizes for relaxation training with biofeedback .34 (+.8169), behaviour therapy .51 (+.6469), cognitive therapy .59 (+.5669), relaxation training .64 (.5169) and CBT .91 (+.2469). Whilst there appeared to be some non-trivial advantages in the present study this may have been because the present study was inadequately controlled and these apparent differences may not be 'real' when the comparison was with a wait period of similar length to the treatment period. As a meta-analysis limited information was given about the precise characteristics of the subjects and in general terms it appeared that most studies excluded those with alcohol or substance abuse, psychotic disorders and suicidality (the present study did not exclude clients on these grounds and clients with each of these issues were included within the sample). It seemed in general

terms the studies that made up the comparator meta-analysis did not exclude those with co-morbid personality disorders. It would seem that the most cautious interpretation of this comparison was there was no significant difference between the present study and the meta-analytic findings of Gould, et al. for generalised anxiety disorder (GAD). In general terms this conclusion would seem to have some support from the earlier Borkovec & Whisman (1996) GAD meta-analysis, finding that CBT failed to demonstrate any superiority over non-directive therapy.

The next comparison made was that with Bryant, et al. (1998) who studied clients who had suffered civilian trauma and who were suffering with acute stress disorder (ASD), a known precursor to post-traumatic stress disorder (PTSD). Clients were offered five 1½ hour sessions of CBT within two weeks of experiencing a motor vehicle or industrial accident. The contact time of 7.5 hours was comparable with the mean 8.1 sessions for those clients with anxiety and a subsequent measurement of their anxiety in the present study. These authors used a range of measures (Clinician administered PTSD measure, Acute Stress Disorder Interview, Impact of Events Scale and BDI) and the State-Trait Anxiety measure used for this comparison. Comparing effect sizes between the comparator study and the present study showed trivial differences for the STAI State anxiety measurement -.3931 and +.2969 STAI Trait anxiety measurement.

A subsequent Bryant study (Bryant, et al., 1999) used similar measures to their previous study in this study for acute stress disorder suffered by clients following motor vehicle accidents or non-sexual assault; when compared with the present study found trivial differences on STAI Trait anxiety across all conditions; prolonged exposure and anxiety management -.2931, prolonged exposure -.2331 and ‘supportive counselling’ (psychological placebo) +.3069.

The study of Barrowclough, et al. (2001) sought to compare CBT and counselling for anxiety symptoms in older adults using BAI. Whilst there appeared to be a significant advantage for CBT in this comparator study ES -.5231, this advantage was probably due to the higher mean BAI start scores. When the comparison was with a sample with start BAI ≥ 19 and a subsequent BAI measurement from this study (ES = 2.2222) the apparent advantage disappeared in favour of a slight advantage for the present study, +.5422. The Barrowclough, et al. study used only one counsellor for the ‘supportive counselling’ arm of the study, the present study showed an advantage over this arm, both for the regular anxiety clients +.4469

and for the more anxious subset ($BAI > 19$) of the present sample $+1.5122$. There were significant differences in the client populations studied for the comparator study, as a study of CBT for anxiety with older adults, for example, comparator study mean age 72 years ($SD 6.2$) and the present study, mean age 35 years for the anxiety ‘subsequent measure’ clients.

Comparing the present study with the anxiety arm of the Elliott, et al. (2004) meta-analysis pointed to a trivial difference of $-.1431$. This meta-analysis and the following one did not identify the measures used.

On the other hand the comparison with the Westen and Morrison (2001) meta-analysis pointed to a large advantage for CBT $-.9331$ compared with the present study. However, these authors questioned the validity of their meta-analytic findings and were particularly critical of reliance upon effect sizes, pointing to the importance of additional metrics such as percent improved, post-treatment symptomatology and follow-up for two years and beyond. One of the themes of this meta-analysis that emerged was the correlation between percentage of patients excluded in each study that made up the meta-analysis and the percentage of patients that improved with treatment, i.e. studies that excluded high percentages of patients had high percentages of patients that improved. The authors questioned the assumption that researchers appeared to have made, that improved internal validity (maximisation of diagnostic homogeneity) was worth diminished external validity (e.g. when researchers exclude 70% of anxious patients from a study they cannot legitimately generalise to any but a minority of anxious patients). On this basis it was not clear that there was any real advantage from the studies included in this comparator meta-analysis with the present study, although there may have been. The present study appears strong on external validity, as this was a sample of clients seen in routine practice, by Westen and Morrison’s argument.

8.1.3 Distress

In terms of distress outcomes, 130 clients started with a clinical level of distress, on an LOCF basis an $ES = 1.1159$ (95% CI $.89$ to 1.34); of these 79 clients had a subsequent measure of their distress (39.2% were ‘missing’) with $ES = 1.80$ (95% CI 1.53 to 2.06), 21.5% had no reliable change, 78.5% had reliable change (55.7% were recovered and 22.8% were improved only). ‘Subsequent measure’ clients were not necessarily ‘completers’. For these 79 ‘subsequent measure’ clients, mean age was 33 years, 63% were female, 60% were single, 65% were not parents, 89% were British and 85% were taking no relevant medication.

Clients were seen for a mean 8.8 sessions by 16 therapists and the author saw 60.8% of clients.

A subset of 37 UCS clients had wait controlled outcomes and this showed there was a small and significant improvement during waiting ($ES = .22$) and a significant improvement during treatment ($ES = 1.45$), suggesting for these clients, a wait controlled $ES = 1.22$. Treatment (Mdn 49 days) was significantly longer than wait (Mdn 20 days), so the dynamics of change were examined. This suggested a significant improvement of 1.9 CORE-OM units after the exploratory session and no significant change as waiting progressed. After the first counselling session there was a significant improvement (5.4 CORE-OM units) and there was a significant relationship between time (days) and improvements in distress found during treatment for this subset, i.e. longer treatment, lower distress. When the analysis of improvements in distress was based on number of sessions there was a trend towards a 'premature endings effect' at the UCS ($p = .054$).

Comparing the wait controlled subset of clients with those clients for whom only waiting data or treatment data were available found no significant differences on start of period or end of period outcome measures, i.e. in terms of outcome measures the wait controlled clients were representative of those other clients who only waited or were only treated.

Clients who had a subsequent measurement of their distress were not significantly more or less distressed at the start than those who did not have a subsequent measurement of their distress.

Clients who were taking relevant medication had no significant difference in outcome compared with those who were not.

As described above this study ($ES(d) = 1.8007$) was compared with the outcomes for the studies chosen as benchmarks (see Introduction) with the criterion that differences in effect size of .4 or greater were considered 'interesting' (Elliott, et al., 2004). Apart from the meta-analysis of Elliott et al. all of these comparator studies used CORE-OM as the outcome measure.

The outcomes for distress were compared with the overall outcomes reported in the Elliott, et al. (2004) meta-analysis (127 treatment groups) on the basis that CORE-OM is to some extent a measure of 'all round' psychological well-being (Barkham, et al., 2001) and well suited to such a comparison. The differences between the present study and the comparator

meta-analysis point to a non-trivial advantage for this study +.8107 when compared with group mean and +.9807 when compared with mean weighted by sample size. However, as noted above, without detailed information about the inclusion and exclusion criteria for the studies included in the meta-analysis it was difficult to determine the validity of such comparisons. The comparator meta-analysis included 42 controlled studies and suggested the mean ES for untreated conditions was $ES = .11$. The present study found a small ($ES = .22$) and significant improvement during waiting that was significantly less than the improvement during treatment. Again, it was not clear whether the effect size should be controlled to take account of improvement during wait, but in any case the uncontrolled effect size for distress showed a large advantage over the meta-analysis controlled effect size of +.9107 when compared with the group mean and +1.0207 when compared with the sample size weighted group mean; this suggested an advantage over these studies even if the effect size was controlled for untreated improvement e.g. $ES = .22$. Without detailed information about the studies included in the meta-analysis it was difficult to determine the validity of such comparisons and whether there was a 'real' advantage from the present study.

Using CORE-OM, Stiles, et al. (2006) reported an average $ES = 1.36$ for 1,309 patients who received CBT, PDT or PCT at one of 58 NHS primary or secondary care sites. These authors found no significant outcome differences between treatments. Both this comparator study and the present study looked at patients found in routine care; females accounted for 71% and 63% of the samples respectively and so there were some differences in the study populations. The present study had a slightly higher mean start score 18.4 cf. 17.4 and a lower SD 5.0 cf. 6.5. There was a small and non-trivial advantage for the present study +.4407, although to what extent this was a 'real' advantage was unclear, if unreported differences in patient populations were taken into account. Interestingly the comparator study had nearly half of the population (49%) taking psychotropic medications cf. 15% in the present study, although the present study found no statistically significant advantage from taking psychotropic medication for distress.

Mullin, et al. (2006) and Stiles, et al. (2007) reported naturalistic CORE-OM outcomes for 11,953 and 5,613 patients respectively seen at 32 NHS primary care settings where CBT, PDT and PCT were offered; again no statistically significant outcome differences between therapy approaches were found and ES of 1.42 and 1.39 respectively were found. The first of these studies had a trivially lower ES than the present study +.3807 and the second of these had a non-trivial but lower ES +.4107. The first of these comparator studies provided little

demographic information and the second, compared with the present study, was 71% female cf. 63%, mean age 40.7 years (SD 12.7) cf. 32.5 years (SD 10.7) and 53.3% were taking medications cf. 15%. The first of these comparator studies also provided recovery, improvement and deterioration benchmarks. Amalgamating the benchmark comparisons suggested the reliable change results from the present study were about average cf. 11,953 clients at 32 NHS primary care counselling and psychological therapy services.

In summary, the effect sizes for depression, anxiety and distress were compared with other studies and on the whole found to be broadly comparable, although without a formal statistical comparison (e.g. non-central t). Analysis of reliable change suggested the observed improvements were not just restricted to a small proportion of cases improving.

Whilst there was no randomisation to different treatments in this research, clients in this research were of severities broadly comparable with other studies on the most important predictor of outcome; pre-treatment scores. Whilst there was no direct evidence that the person-centred psychotherapy was appropriately delivered in compliance of a treatment manual, for a sub-group of clients, client perceptions of congruent empathy and unconditional positive regard were measured. The BLRI scores did suggest that on the whole clients perceived that congruent empathy and unconditional positive regard were offered towards them by their therapist. For the UCS part of the research there was some evidence of a 'premature endings' effect and some of the clients seen at the UCS may have ended therapy prematurely; perhaps because of the relatively unstable nature of a university setting where clients and trainees leave periodically. At the PP clients were more regularly monitored for progress by way of 'outcomes management'. A high proportion of the therapists in the UCS arm of the research were pre-qualification trainees and this together with the possibility of a 'premature endings' effect and the absence of 'outcomes management' at this setting suggested the outcome effect sizes at the UCS were not as high as they might otherwise have been and that a greater proportion of clients may have improved than was recorded by a subsequent outcome measurement.

Compared with some of the criticisms that have been made of this type of naturalistic outcome research further analysis of the context for these outcomes suggested it was unlikely that (cf. Clark, et al., 2007):

- There were lots of ‘missing cases’ that were really ‘treatment failures’
- These findings were restricted to the ‘easiest’ clients in the research
- These findings were restricted to the ‘best’ outcomes
- These findings were simply ‘regression to the mean’
- These findings were simply ‘natural recovery’
- These findings were ‘attributable to concurrently administered medications’

Instead it seemed plausible person-centred psychotherapy in this research was at least partly responsible for some of the observed improvements.

8.2 Process-outcome findings

8.2.1 Introduction

There was some evidence that outcomes for depression were predicted by the therapeutic relationship ($r = .22$); ‘illustrative’ results suggested the therapeutic relationship could have predicted outcomes for anxiety ($r = .25$) and distress ($r = .29$), although there may have been some issues to do with the sample wherein there were alternative predictors of outcome, site and number of sessions, respectively.

The Norcross (2010) meta-analytic review of key elements of the therapy relationship identified a range of effect sizes from the lowest, $r = .19$ (alliance in child and adolescent psychotherapy) to the highest, $r = .56$ (managing counter transference). It was not known whether these estimates of effect size were significantly different.

The significant and illustrative findings from this research are of a similar order to the findings found in the literature, of which the preceding is simply an example.

Elliott (2010), citing DeRubeis (2007), described the case for the upper bounds for process-outcome correlations likely being in the range .2 to .4 and that the meta-analytic finding for empathy of $r = .32$ was likely as large as that obtainable (Elliott, et al., 2010). Elliott (2010) described three reasons for this ‘ceiling effect’, firstly, to do with client variance across a number of variables (e.g. disordered personality process, pre-therapy severity and complexity, resilience, psychological fragility, life circumstances, etc.), secondly to do with the restricted range of therapist-provided predictor variables (e.g. empathy) and thirdly to do with measurement error (stability and consistency of measures).

8.2.2 Depression

This research found evidence that outcomes for depression were predicted by the therapeutic relationship ($r = .22$) once six cases were removed from the analysis. These six cases were not described in detail because of client confidentiality and at summary level these were in accordance with Elliott’s first reason for a ceiling effect in terms of client variance across a number of variables. In terms of Elliott’s second reason the restricted range of therapist-provided relationships was noted in section 7.2.1 above. In terms of the third reason, measurement error for the instruments used was noted in section 5.2, as was regression to the mean effects (section 6.1.2) and can be added to this were variability in the measurement of the relationship with BLRI because only one measurement was taken (Barnett, et al 2005) plus likely regression dilution effects in the correlation (Barnett, et al.). Beyond the reasons cited by Elliott there was some evidence on an interaction effect between reasons 1 and 3, that is there appeared to be some client-based variable(s) that impacted the assessment of the therapeutic relationship with BLRI and this seemed to fit with observations by Baldwin, et al. (2007) and Crits-Christoph, et al. (2009) and this may have further reduced the ability of the methodology to detect a significant effect. This phenomenon may go some way to explaining the lack of a significant finding for the anxiety and distress analyses, see below.

To what extent this process-outcome finding supported the case for a claim of causal inference is considered below.

8.2.3 Anxiety

Initially this research found no evidence that outcomes for anxiety were predicted by the therapeutic relationship. Analysis for the outcomes part of the research found that there was a statistically significant relationship between anxiety outcome and site, clients at the PP had significantly better outcomes than at the UCS. It was not possible to know what accounted for this difference, possible explanations included: a) factors to do with site e.g. private home, private practice, etc.; b) factors to do with the approach of the therapist e.g. process-directive empathic approach to the difficult feelings associated with 'exposure' to any difficult stimuli in a client-responsive manner; c) factors to do with the outcomes management approach e.g. it is not known to what extent different client concerns may be differentially impacted by outcomes monitoring, anxious clients might find this particularly reassuring; d) other unknown factors; e) some combination of one or more of these, or; f) none of these.

However, with a covariate this reduced any unique variance for any effect of the therapeutic relationship, with potentially confounded variance for site and the effect of the therapeutic relationship (Field, 2009, pp. 397-9). Furthermore, the sample size for the anxiety dataset (maximally, $n = 75$) was smaller than that for depression (maximally, $n = 92$). Calculation of sample size for regression is complicated (Field, 2009, pp. 222-3) and perhaps of critical importance Field suggested increasing the number of predictors from 2 to 3 for medium effects increased the required sample size from around 70 to around 80. It seemed plausible that the absence of a statistically significant finding could be due to a lack of power, together with the confounded variables. Future research would perhaps benefit from a bigger sample size and the absence of a 'site' effect or other similar confounding variables.

Seeking to discover perhaps what reasons had led to a non-significant finding and to attempt to estimate any apparent effect size for future research an iterative process was followed wherein the case with the highest DFBETA for BLRI was deleted from the regression analysis. Removal of the seven cases most influencing BLRI from being a statistically significant predictor suggested an 'illustrative' effect size of $r = .25$ ($n = 68$).

Using Elliott's (2010) structure in a similar manner to the analysis for prediction of depression outcome by the therapeutic relationship suggested client variance across a number of variables could have been problematic (e.g. disordered personality process, pre-therapy severity and complexity, resilience, psychological fragility, life circumstances, etc.), together with the restricted range of therapeutic relationship predictor variables, measurement error and perhaps some interaction between some unknown client variable and systematic error in

the measurement of the relationship e.g. histrionic personality process (section 7.2.6) or other client variable.

With consideration of the foregoing it seemed plausible that therapeutic relationship could be a significant predictor of anxiety outcome to the order of $r = .25$ suggested by the analysis and in accordance with the literature.

To what extent this process-outcome finding supported the case for a claim of causal inference is considered below.

8.2.3 Distress

Initially this research found no evidence that outcomes for distress were predicted by the therapeutic relationship. Analysis for the outcomes part of the research found that there was a statistically significant relationship between distress outcome and time. This could have been because the distress dataset as a percentage had a higher proportion of UCS clients, some of whom finished at the end of term, had interrupted processes due to end of term breaks, had therapists leave at the end of term, etc. This appeared to be a dose-effect relationship.

Number of sessions was entered as a statistically significant covariate. In a similar manner to the anxiety process-outcome dataset this covariate reduced any unique variance for any effect of the therapeutic relationship, with potentially confounded variance for number of sessions and the effect of the therapeutic relationship (Field, 2009, pp. 397-9). Furthermore, the sample size for the distress dataset (maximally, $n = 54$) was smaller than that for both depression (maximally, $n = 92$) and anxiety (maximally, $n = 75$). As with the anxiety dataset for three, rather than two predictors, even of medium size increases the required sample size from around 70 to around 80. It seemed plausible that the absence of a statistically significant finding could be due to a lack of power, together with the confounded variables. Future research would perhaps benefit from a bigger sample size and the absence of a 'session' effect or other similar confounding variables.

A similar process was followed as previously described for anxiety and removal of the five cases most influencing BLRI from being a statistically significant predictor suggested an 'illustrative' effect size of $r = .29$ ($n = 49$).

Again, using Elliott's (2010) structure in a similar manner to the analysis for prediction of depression and anxiety outcome by the therapeutic relationship suggested client variance

across a number of variables could have been problematic (e.g. disordered personality process, pre-therapy severity and complexity, resilience, psychological fragility, life circumstances, etc.), together with the restricted range of therapeutic relationship predictor variables, measurement error and perhaps some interaction between some unknown client variable and systematic error in the measurement of the relationship e.g. histrionic personality process (section 7.2.6) or other client variable.

With consideration of the foregoing it seemed plausible that therapeutic relationship could be a significant predictor of distress outcome to the order of $r = .29$ suggested by the analysis and in accordance with the literature.

To what extent this process-outcome finding supported the case for a claim of causal inference is considered below.

8.3 The case for causal inference based on the outcomes and process-outcomes findings.

Elliott (2010) reviewed psychotherapy change process research and compared four different groups of change process research designs. From the literature (Haynes & O'Brien, 2000, Cook & Campbell, 1979) Elliott identified seven criteria to consider as 'causal inference criteria'. The author pointed out that each of the four research designs being reviewed had their strengths and weaknesses, and that others may disagree with his analysis of these, and he concluded that each of the designs had a part to play in making causal inferences and that none of the designs by themselves were sufficient to satisfy all of the criteria for causal inference.

Seeking to rise to Elliott's challenge that 'psychotherapy researchers are sometimes shy about aspiring to causal inference' (2010, p. 133) the present author attempted to consider what the present research (outcomes, process-outcomes and literature findings) might offer towards causal inference of the therapeutic relationship 'causing' outcomes in person-centred psychotherapy.

Analysing causal inference is complicated, as Cook and Campbell recognised when they wrote 'external validity and construct validity are so highly related that it was difficult for us

to clarify some of the threats as belonging to one validity type or another' (1979, p. 82). This is a brief overview.

Taking each of the seven causal inference criteria in turn:

1. Document temporal precedence

In this research the majority of clients within the regression analyses completed their assessment of outcome after they completed their perception of the therapeutic relationship.

For the depression analysis ($n = 86$, $r = .22$) 80.2% of clients had temporal precedence, overall mean 4.6 sessions, for anxiety ($n = 68$, illustrative $r = .25$) 80.9% of clients had temporal precedence, overall mean 4.8 sessions and for distress ($n = 49$, illustrative $r = .29$) 71.4% of clients had temporal precedence, overall mean 3.8 sessions. Whilst to some extent temporal precedence was shown this was something that could be improved in future research, and see literature review.

2. Provide plausible explanation

This research did not provide a plausible explanation, or logical mechanism, for the hypothesised causal relationship between congruent empathy and unconditional positive regard and client outcomes. The techniques used in this research were not suited to a consideration of this causal inference criterion (although see literature review) and other techniques were better suited to this type of consideration e.g. helpful factors, sequential process or significant events (Elliott, 2010).

3. Show covariation

In this research depression outcome was shown to covary with the therapeutic relationship as defined by Rogers, whilst controlling for start depression, for 86 clients $r = .22$, $p = .047$. In this research there was a null finding for covariation of anxiety and distress outcome with the therapeutic relationship as defined by Rogers, whilst controlling for the pre-test. There was some suggestion of issues to do with the sample and illustrative analyses suggested for 68 clients with anxiety outcomes $r = .25$ was plausible and for 49 clients with distress outcomes $r = .29$ was plausible. This was something that could perhaps be assessed again in future research; and see literature review.

4. Consider alternate causes

In this research 111 clients with depression had large improvements in their depression (ES = 1.48, 95% CI 1.28 to 1.68), 70.3% had reliable improvement and 53.2% of the 111 depressed clients had recovered. There was some suggestion that a proportion of these clients may have had co-morbid disordered personality processes. Analysis suggested 7.2% of the observed improvement in depression may have been due to regression to the mean effects. Analysis suggested for a subset of the clients that they did not improve whilst waiting and did improve during treatment and that there was no significant relationship between passage of time and recovery from depression. Of the 111 clients with depression, 21 were taking anti-depressants and the analysis suggested those taking anti-depressants had a larger improvement in their depression, a small effect ($r = .19$) for presence of antidepressant, although not necessarily due to the antidepressant because those taking antidepressants started with higher depression scores. Of the 111 clients with depression 83 completed a measure of the therapeutic relationship as defined by Rogers and the mean score of 188.40 (SD 20.27, range 126-237) suggested that at least for this subset of 74.8% the conditions hypothesised to cause outcome were present, as perceived by the client. This research considered some, but not all alternate causes, and it seemed plausible that person-centred psychotherapy was at least partly responsible for at least some of the observed improvements.

In this research 91 clients with anxiety had large improvements in their anxiety (ES = 1.15, 95% CI .95 to 1.35), 29.7% had reliable improvement and 18.7% of the 91 anxious clients had recovered. There was some suggestion that a proportion of these clients may have had co-morbid disordered personality processes. Analysis suggested 9.6% of the observed improvement in anxiety may have been due to regression to the mean effects. Analysis suggested for a subset of the clients that they did improve whilst waiting and did improve during treatment. There was a significant relationship between passage of time and change in anxiety whilst waiting; longer wait was associated with worsening anxiety. There was no significant relationship between passage of time and recovery from anxiety during treatment. Of the 91 clients with anxiety, 20 were taking some form of psycho-active medication (of which 16 were taking anti-depressants) and the analysis suggested those taking psychoactive drugs/anti-depressants had no significant difference to their improvement in anxiety compared with those not taking medications. Of the 91 clients with anxiety 68 completed a measure of the therapeutic relationship as defined by Rogers and the mean score of 188.24 (SD 21.64, range 126-237) suggested that at least for this subset of 74.7% the conditions

hypothesised to cause outcome were present, as perceived by the client. This research considered some, but not all alternate causes, and it seemed plausible that person-centred psychotherapy was at least partly responsible for at least some of the observed improvements.

In this research 79 clients with distress had large improvements in their distress (ES = 1.80, 95% CI 1.53 to 2.06), 74.7% had reliable improvement and 54.4% of the distressed clients had recovered. There was some suggestion that a proportion of these clients may have had co-morbid disordered personality processes. Analysis suggested 10.6% of the observed improvement in distress may have been due to regression to the mean effects. Analysis suggested for a subset of the clients that they did improve whilst waiting and did improve during treatment. There was no significant relationship between passage of time and change in distress whilst waiting. There was a significant relationship between passage of time and recovery from distress during treatment and this dose-response effect may have been due to interruptions in therapy processes at the UCS. Of the 79 clients with distress, 12 were taking some form of psycho-active medication (of which 7 were taking anti-depressants) and the analysis suggested those taking psychoactive drugs/anti-depressants had no significant difference to their improvement in distress compared with those not taking medications. Of the 91 clients with anxiety 21 completed a measure of the therapeutic relationship as defined by Rogers and the mean score of 148.5 (SD 40.8, range 87-233) suggested that at least for this subset the conditions hypothesised to cause outcome were present, as perceived by the client. Of the 79 clients with distress 45 completed a measure of the therapeutic relationship as defined by Rogers and the mean score of 188.24 (SD 16.38, range 135-228) suggested that at least for this subset of 57.0% the conditions hypothesised to cause outcome were present, as perceived by the client. This research considered some, but not all alternate causes, and it seemed plausible that person-centred psychotherapy was at least partly responsible for at least some of the observed improvements.

Future research of this kind could consider other plausible alternate causes (and see literature review) and Elliott (2010) pointed out that significant events research was also suited a consideration of this causal inference criterion.

5. Demonstrate construct validity of cause/effect

This research did not demonstrate the construct validity of cause/effect and future research of this kind could consider this causal inference criterion (see literature review) and Elliott

(2010) pointed out that significant events research was also suited to a consideration of this causal inference criterion.

6. Direct relevance to clinical practice (generalisability)

Elliott (2010) wrote that process-outcome research was not by itself directly relevant to clinical practice (the generalisability criterion) and that helpful factors, sequential process and significant events research were better suited to a consideration of this causal inference criterion (see literature review).

7. Provide direct causal evidence

This research did not provide direct causal evidence of the hypothesised link between client perceptions of congruent empathy and unconditional positive regard with client outcomes (and see literature review). Elliott (2010) pointed out that helpful factors, sequential process and significant events research could lead to a consideration of this causal inference criterion.

In summary, in addition to the findings in the literature, this research appeared to make some limited contributions to 3 of the seven causal inference criteria and suggested specific areas for further research to test the proposition that the provision of congruent empathy and unconditional positive regard *caused* at least some part of the observed improvements in clients' wellbeing.

8.4 Strengths and weaknesses of this research

Like most research of this type, this research had some strengths and some significant weaknesses. The research was under-supervised, the author inexperienced and much time and effort was used. The research design evolved during the study and much time and effort was wasted analysing findings to seek to understand and control for methodological imperfections. In some ways a randomised controlled trial could have made the analysis much easier, although practically in the settings where therapy was delivered it would have been difficult to deliver a comparator treatment condition. The research was carried out at two sites with subtly different client groups and subtly different ways of working and this served to increase the complexity of both the conduct and analysis of the research. Whilst post hoc ‘further analysis’ has sought to address some of the criticisms of naturalistic research these criticisms still remain, pending a more thorough trial. This section describes some further weaknesses (and some strengths) and discusses what could be done in future research to mitigate effects of these.

8.4.1 Questionnaires and related issues

The author designed a ‘client demographics’ questionnaire to capture data to characterise the sample as recommended by TREND (Des Jarlais, et al., 2004); a strength. Whilst client demographic information was captured there was a problem with capturing information about ethnic status, which was left to clients to complete as a ‘free format’ response. Whilst this was ‘person-centred’ it meant this was not captured in a regular way, a weakness. In future it would make more sense to ask for this data in an approved structured format, e.g. in a format recommended by the Equalities and Human Rights Commission.

The ‘client demographics’ questionnaire sought information about medication usage and this was interpreted by a Member of the Royal College of General Practitioners by categorising information into different groups of relevant psycho-active drugs. It was important to note that what a client recorded on their questionnaire did not necessarily equate with what was really going on. Clients may have under or over-reported their prescription medications and simply recording a drug was prescribed did not mean that it was being taken or that it was taken at therapeutic levels in accordance with a treatment protocol. As a naturalistic experiment there was no control of medication usage and an attempt was made to control for this, to the extent that was possible with the information available. The numbers of clients

taking medication was relatively small, typically 15-22%. The information captured about concurrent medication use may have been sufficient to rule this out as a cause of the observed improvement, however given the apparently small effect sizes for medications compared with therapy there was insufficient client numbers to assess the effect size of the medications taken.

For depression, anxiety and distress outcomes there was no statistically significant interaction between stage of therapy (start of therapy measure, subsequent measure) and medication status (relevant medication taken or not), apart from a small and significant interaction between stage of therapy for depression and antidepressant medication status (antidepressant taken or not). This was a small effect for anti-depressant medication and whilst this difference was significant it did not mean that the medication was necessarily responsible for this small effect difference; it could have been a confound that more depressed clients were more likely to have been given anti-depressants and because their starting depression score was higher their improvement was larger. Another explanation could be that those clients on anti-depressants were also receiving some 'therapeutic input' from the prescriber of their medication, since there is some suggestion in the literature that it is the therapeutic relationship with the prescriber to a greater extent than the medication, that is responsible for outcome (see Introduction). A fairer test would be to randomise clients to medication-taking status and this was not possible in this research.

The finding of small effect sizes relative to therapy appeared to accord with findings in the literature for anti-depressant medications; wherein Kirsch and colleagues appear to have found evidence for *statistically* significant effects, but not *clinically* significant effects, apart from for the most extremely severe depression (e.g. Kirsch, et al., 2008). Given the large amounts of money spent in the UK on anti-depressants (e.g. Weston & Weston, 2008) for seemingly relatively small effect this would appear to support a case for increasing provision of psychological therapies; and this would appear to be the rationale underlying the UK Government's IAPT programme.

The outcome questionnaires used in this research, particularly BDI-II, BAI and CORE-OM were apparently effective measures of outcome based upon finding good reliabilities and were generally easily completed by clients.

The test-retest stabilities of the BDI-II and CORE-OM were high, .93 and .90 respectively and this led to relatively low reliable change criteria of 35% and 31% of the mean start scores

(Table 2). However, the test-retest stability of the BAI was relatively low at .75 and this led to a relatively high reliable change criterion of 85% of the mean start score. This may have gone some way to accounting for the relatively low reliable change percentages observed for anxiety where 30% of clients had reliable change compared with depression and distress, where 70% and 75% of clients had reliable change, respectively. The low test-retest stability of BAI was not helped by finding a distribution of start scores that were significantly different from a normal distribution and transformation went some way to improving the reliable change criterion for BAI to 42% of the mean start score. Transformation also helped reduce the reliable change criteria for BDI-II to 16% of the mean start score. With transformed scores the reliable change percentages were 71% for BDI-II and 46% for BAI, compared with 75% for CORE-OM (untransformed scores). Whilst the effect sizes were smaller for anxiety 1.56 (with transformed scores) than depression 1.87 (with transformed scores) and distress, 1.80 (untransformed scores) future research might benefit from an anxiety questionnaire with higher test-retest stability.

Interestingly all three of the outcome questionnaire reliabilities measured and compared with their authors published reliabilities were lower, but often only very slightly, than those reported by their authors.

All of the reliabilities found in this research exceeded the recommended cut-off of .7 (Field, 2005) and by quite some way.

To this author's knowledge, the PBQ has not been widely used and its provenance as a screen has not been confirmed by researchers other than its authors.

It was inconvenient that the PBQ subscales passive-aggressive, histrionic and schizoid did not have clinical mean scores.

For future research other personality questionnaires could be considered. It would be ideal, although expensive and in some cases impracticable (e.g. the private practice) to employ a psychiatrist to conduct the personality assessments on intake and subsequently.

Given Rogers' theory that it is client perceptions of congruent empathy and unconditional positive regard to avoid shared method error ideally pre and post-tests would be conducted by or at least supplemented by observer measurements.

8.4.2 Frequency of measurement and related issues

It is not unusual for a naturalistic study involving several therapists to have a number of clients who start the research and ‘disappear’ from the research for various different reasons, including premature endings before post-measures completed, therapists forgetting to ask for post-measure completion etc. (Stiles, et al., 2006, Stiles, et al., 2007). This aspect of naturalistic research has been criticised (Clark, et al., 2007) and it is an inevitability of effectiveness (cf. efficacy) research that could limit the generalisability of the findings (external validity) but not the value of this type of research (Stiles, et al., 2008).

The fact remains too many clients started in the research and did not complete subsequent outcome measures and there were too many ‘missing cases’. The UCS site in particular was affected in this way and this could have been for a number of reasons including therapists deciding to opt out of the research part way through, clients opting out, etc. At the PP site there were only five clients with clinical depression scores, who did more than one session and who did not complete a subsequent depression questionnaire; there were more clients with clinical anxiety scores at the PP site who did not complete a subsequent measure of their anxiety and this was because completion of a depression questionnaire was seen as a priority over and above completion of an anxiety questionnaire (Newman, et al., 2006).

It would seem appropriate to monitor client progress frequently, perhaps with completion of at least one questionnaire per session, at least for the first six sessions or so, and then at regularity determined by the client and therapist to make sense to them both. There is a clear need to balance the needs of the researcher with the needs and wishes of the client. To some extent these needs can overlap, especially if the therapist shares the results of the questionnaires with the client, so that quantitative progress, or its lack, is part of the therapeutic dialogue. As described above, there is evidence that ‘outcomes management’ can improve outcomes (e.g. the work of Lambert and others) and that mutual understanding of the goals and tasks of therapy is both a predictor and mediator of outcome in therapy (e.g. the work of Horvath and others). It was a strength of this research that outcome measures were used reasonably frequently at the PP and a weakness that they were used only at the last session at the UCS. It was a weakness of the research that there were two different ways of using outcome questionnaires within the same trial.

Whilst many therapists and researchers would agree with the idea of using outcome measures reasonably frequently as described in the preceding paragraph it is important to note that not

all would. There are therapists and clients who would not want to 'impose' upon the therapeutic relationship in the way described above and this does not invalidate the wishes of these clients and or their therapists. However, it would seem appropriate that this choice is offered, perhaps through training therapists to be able to do this, so that therapists and clients can make an informed choice to 'opt out' of this way of working, if they wish.

Monitoring client progress through therapy can provide some practice-based evidence of effectiveness (or otherwise); a strength of this research. However, this 'strength' is illusory if the bodies that review the evidence for therapeutic approaches do not consider practice-based evidence as admissible or that the outcome measures used are invalid (NICE, 2009a, 2009b). For example a large amount of effort has gone into implementing the CORE system in the UK (Barkham, et al., 2006) and whilst the CORE-OM has been shown to be a valid means of diagnosing depression (Gilbody, et al., 2007) this has not been accepted by the reviewers (NICE, 2009a, 2009b). To that extent there would appear to be a disconnection between those who would seek to monitor the effectiveness of their therapy service and those who would pronounce upon the efficacy/effectiveness of therapy and there would appear to be scope to improve this connection to the benefit of clients, therapists, researchers and reviewers. To that extent the use of practice-based research using CORE-OM was evidence of two weaknesses, from the perspective of NICE, rather than using 'more acceptable' RCT methodology and a 'more acceptable' outcome measurement.

Using BDI-II and BAI was a strength, as these have both been considered by NICE as valid measures of depression and anxiety respectively (NICE, 2004b, 2004a). It is reasonably widely accepted that depression and anxiety frequently co-exist (Beck, et al., pp. 27-28, Beck & Steer, 1993, p. 1). With a desire to monitor outcome throughout therapy and to capture practice-based evidence a therapist is faced with the problem of what questionnaires to give to clients and when. This author's approach to this was to take the guidance in the literature, that since CBT for anxiety was found to be more effective with non-depressed clients then depression should be treated (hence monitored) as a priority (Newman, et al., 2006). Having a rationale for this approach was a strength, however the result that fewer clients at the PP had subsequent anxiety measures was a weakness.

This issue is far wider than simply two competing outcome measures. It was perfectly possible that a client could have been regularly completing questionnaires about their depression, anxiety, panic attacks, suicidal ideation, obsessive-compulsive problems,

posttraumatic stress disorder, and so on. Clearly this is a nonsensical situation wherein the monitoring takes over the therapy. However, not regularly monitoring outcomes gives rise to 'missing cases' and attendant criticisms from other researchers. There is a question about what a therapist who is seeking to be 'responsible' to clients (current and future), other researchers and evidence-based reviewers ought to do, to satisfy the interests of all parties. In particular the economic costs and benefits of doing 'empirically supported therapy' are high. Authors, researchers and practitioners of 'evidence-based' therapies stand to do well in terms of money, prestige, power and so on; those who aren't won't. It was a strength of this research that outcomes were monitored for a large number of clients and a weakness that there were missing cases.

In the early part of this research this author did not appreciate the importance of not just 'clinical significance' but also 'reliable change'. There were around ten clients who started with mild depression and subsequently had non-clinical scores; however, the change in scores was not big enough to be reliable (Jacobson & Truax, 1991). This could distort the results, in that clients could have been asked to complete a subsequent depression questionnaire that might have provided evidence of a further reduction in depression to make this not just clinically significant but also reliable change. This was a weakness of the research. This pointed up an education need for therapists, about understanding 'reliable change' (not covered in any of the questionnaire manuals) and also for therapists to be able to effectively communicate this to their clients; otherwise clients could be left wondering why they are still completing depression questionnaires when their score is in the non-clinical range. There is the potential for an ethical issue in this area; in whose interests is the questionnaire completion, is this about establishing that the client is no longer depressed, or about proving the worthiness of the therapist and therapeutic approach. A related issue is that if therapists are judged on their outcomes there is a disincentive to take on 'difficult cases' that risk reducing apparent effect sizes for a particular therapist (e.g. very severely depressed borderline personality disordered clients who have been sexually abused, are actively suicidal, have an eating problem, expect not to get better, are not psychologically minded, have poor attachment style, do not want to be in therapy, are homeless and misusing substances, Clarkin & Levy, 2004).

A further weakness of this research was that there was no reported follow-up of clients to check if treatment gains were maintained (Westen et al., 2004). A small proportion of clients did receive follow-up and the methodology was to post outcome questionnaires out to clients,

with their prior consent, at three, six, nine and twelve months post-completion. Whilst the unreported results of this endeavour suggested treatment gains were being maintained the author did not have the resources to continue this practice and this was a weakness in the research. With appropriate resources similar research in future could benefit from monitoring post-completion. Furthermore it could be an appropriate part of the therapy service for clients to receive regular post-completion monitoring, a potential strength.

As described above, in the predictor analysis, there may have been an issue about when the subsequent outcome measure was taken that impacted the finding for prediction of outcome by the therapeutic relationship; the 'premature endings' effect, especially for the relationship-distress dataset. This was a weakness in this part of the research. In addition to the outcomes part of the research the predictor part of the research could have benefitted from monitoring outcome at every session. This could have enabled a measurement of the outcome curve slope for higher and lower levels of the perceived therapeutic relationship using Structural Equation Modelling (SEM), e.g. Zuroff and Blatt (2006). This analysis would have been beyond the capability of this particular author (a weakness). However use of such techniques could further the analysis of causal inference criteria, criterion 5, demonstrate construct validity of cause/effect has been achieved by some researchers with this type of approach e.g. Burns and Nolen-Hoeksema (1992) showed that empathy directly caused recovery from depression in cognitive-behavioural therapy, Zuroff and Blatt (2006) showed that client perceptions of therapist regard, empathy and congruence caused recovery from depression with CBT, IPT, medication and pill placebo conditions.

8.4.3 Measurement of the therapeutic relationship and related issues

The therapeutic relationship as defined by Rogers was assessed by clients as per the theoretical statements (1957, 1959). This part of the research was in line with the theory and a clear strength of the research cf. earlier criticisms of research of the theory, e.g. Lambert, et al. (1978), Watson, (1984). This author judged the BLRI to be cumbersome to complete (weakness) and sought to improve the ease of completion, with Barrett-Lennard's consent and assistance (strength). However, the questionnaire as used in this research was not identical to that used in previous research and only limited reliance could be placed upon previous reliability data for the questionnaire (Gurman, 1977). This was a weakness, although

the alpha coefficient was assessed for the revised questionnaire (a strength) and found to be adequate (a further strength).

Barrett-Lennard's advice was taken upon which session to measure the relationship (a strength) and his recommendation was followed that this should be the start of the fifth session (a strength). During the course of the research it became apparent that some clients were not completing the BLRI because they did not get to five sessions. This reduced the data available about therapeutic relationships (a weakness). Potentially clients could have been leaving therapy before the fifth session because the relationship was poor and this missing data could have affected the results (weakness). Furthermore it came to the attention of the author during the research that completion of BLRI after the first session had been shown to be a mediator of outcome (Zuroff & Blatt, 2006). The decision was taken to ask clients to complete the BLRI after the first session. It is a strength to modify an action in the light of new information, however, a weakness to change the methodology part-way through this experiment, in terms of this research.

Asking clients to evaluate a relationship early on meant this could be based more on perception than experience (a weakness). Some could argue this was too early (after one session) to evaluate the therapeutic relationship, before it had had chance to develop (weakness). However, clients do make judgements about whether to work with a particular therapist early-on, based upon their perceptions and this serves to strengthen the case for this approach. There was a statistically significant difference in the evaluation of the relationship based upon when the questionnaire was completed and this suggested a client may be making different judgements after different time periods. There is likely merit in making both measurements, perhaps termed, 'early perceptions of relationship' and 'relationship as experienced to date'. Relationship perceptions could be measured at every session and whilst this could improve the granularity of data this could also serve to put clients off participation in the research and limit the generalisability of the findings, i.e. only valid for clients who enjoy completing questionnaires.

Making only one measurement of the BLRI per client increased the variance of these scores and contributed to regression dilution bias that may have reduced the link between relationship-outcome. Future research may consider addressing this issue e.g. making more than one measurement and using a mean score.

There appeared to be some unmeasured client-based variable that may have been influencing the way in which BLRI forms were completed and this phenomenon has perhaps been observed by other researchers e.g. Baldwin et al. (2007) and Crits-Christoph, et al. (2009) understanding this link, if there is one, may help to refine the estimates of relationship-outcome (Elliott, 2010).

8.4.4 Sampling – Clients, and related issues

It was a strength that the sample numbers were relatively large compared with other research reported in the literature, especially for an individual researcher with modest resources. This served to show that it was possible to do research of this kind and achieve reasonable numbers of clients in a sample. There were particular weaknesses in the nature of the client sample and these were that the UCS and PP were different client groups, attending different counselling services, operating under different ways of working (although both person-centred). Furthermore there was not the level of client diagnostic information that would be expected with a reasonably good RCT, e.g. data on co-morbidity for personality disorder, substance misuse etc. This was a weakness that brought the comparability of these samples into question with the published literature, especially RCTs. Routinely clients who are actively suicidal, have personality disorders or are misusing substances are excluded from RCTs for depression and anxiety. This research included clients with these difficulties and whilst this may be more like ‘counselling in the real world’ (good external validity, a strength) this may have resulted in lower effect sizes because of including the types of clients who may routinely have poorer outcomes (Clarkin & Levy, 2004).

Realising that there was no measure of personality disorder, prevalence and outcome, and doing something about this was a strength. Although, as described above, changing a research methodology part way through is a weakness. The merits and otherwise of the PBQ are discussed elsewhere. It was a strength to estimate the likely prevalence of personality disordered clients in the sample and a weakness that this was only for one small part of the research.

There was some evidence that the relationship-outcome samples were underpowered to detect small effects, especially where outcome was confounded with other variables. Future research would benefit from larger samples and non-confounded outcomes.

8.4.5 Sampling – Therapists, and related issues

This research included outcomes for a reasonably large number of therapists (a strength) and a large number of outcomes for one therapist (another strength). It is also clear from the method and results sections that one therapist saw too high a proportion of the clients (weakness), too many therapists chose not to be involved in the research (weakness) and there were too many trainee therapists and not enough highly experienced therapists (weakness).

It would seem that there is an issue about some person-centred therapists and their attitude towards quantitative research, perhaps because of Rogers' 1985 paper, despite his earlier work (see Introduction); a weakness in 'an evidence-based world'. Too many UCS therapists did not participate or may have subsequently dropped out (weaknesses). However, some UCS therapists did participate and appeared to find a way to do this that was congruent with their way of working (strength). This research suggested it was possible for person-centred therapists to participate in quantitative research (a strength).

Inclusion of trainee therapists in the research may have impacted on BLRI scores and on outcomes and these effects were not examined in this research (weakness).

As described above there was limited treatment specificity and no check on adherence (both weaknesses) although not unusual in research of this kind (Stiles, et al., 2008). Therapists were trusted to do person-centred therapy and there was no check on this, other than completion of the BLRI that measured the quality of the therapeutic relationship. As described above, based on the outcomes data, the quality of the therapy appeared adequate.

As described above there were lots of 'missing cases' (a weakness), perhaps because some of the therapists were not fully enrolled in the research (a weakness and an issue to consider in future) and attempts were made to estimate the impact of this in the analysis (a strength) although others may dispute the methodology (weakness).

Whilst the predictor analysis was not confined to just one therapist (strength), for future research it would make sense to have a more even distribution of clients amongst therapists. Other researchers (Baldwin et al., 2007, Crits-Christoph, et al., 2009) have found between-therapist variation predicted outcome and this analysis was not possible in this research because many therapists saw small numbers of clients and only two saw more than five clients. A large amount of data about one therapist gave an opportunity to consider possible

sources of within therapist variability (a strength). Future research may wish to consider variables predicting client perceptions of therapeutic relationships.

Researching the impact of therapeutic relationships requires the inclusion of therapists who may not offer a 'good' therapeutic relationship to clients, so as to increase the range of relationships examined e.g. Watson (1984). This research did include some trainee therapists (a strength) who some may not expect to provide such a 'good' therapeutic relationship as an expert therapist, although the literature tends to 'cast doubt on the validity of the suggestion that specific training in psychotherapy, even when unconfounded with general experience, may be related to therapeutic success or skill' (Beutler, et al., 2004, p. 239). There could be an ethical issue about researching a range of therapeutic relationships wherein 'poor' therapists are offered to clients for the purposes of discovering lack of progress or deterioration. In this naturalistic research there was a range of BLRI scores and by this measure some 'worse' and some 'better' relationships and no deliberate intent to offer 'poor' therapists.

8.4.6 Methodology and related issues

This research was an uncontrolled naturalistic experiment that is a recognised form of inquiry (strength). Some would argue this form of research is not as valid as an RCT, a weakness, see for example (NICE, 2009a); and there are others who would argue for greater validity for this form of inquiry e.g. Des Jarlais, et al. (2004); a strength.

As this was naturalistic research it was difficult to control for other causes of recovery (a weakness), unlike an RCT. Nevertheless some attempt was made in the analysis to examine and control for this. Furthermore there was no randomisation to different treatments (a weakness), although none was intended.

There was a large variation in 'dose' (weakness) that came about from varying whether or not clients had exploratory sessions, the length of the exploratory session (none or 15-60 minutes), the waiting period before counselling (0-158 days), the numbers of sessions (1-102 sessions), the duration of sessions (50-60 minute therapy sessions), the intervals between sessions (weekly, fortnightly, monthly etc.) and the duration of treatment (0-852 days). Future research may wish to consider issues around 'dose'. There was some attempt to characterise some of the forces at play in relation to 'dose' issues.

Different clients were taking different mediations, as discussed above, and whilst this was a weakness, there was some attempt to consider the impact of this (strength).

Taking all of these issues together the appeal of an RCT that seeks to control for all variables is apparent. This author's experience has been that it was extremely hard work to seek to consider many variables that may impact upon an experiment and RCTs appear much more straight-forward, although costly. For example section 7.1 of this report on outcomes would likely not be required for an RCT; and there are lots of shortcomings in the analysis in section 7.1. However, reading journal articles about RCTs, it is apparent that RCTs are not the panacea they appear to be, in theory. The NIMH TDCRP is likely one of the most thorough and analysed experiments in the history of psychotherapy and yet it still had plenty of shortcomings e.g. Imber, et al., (1990). Elliott (2010) pointed out that RCTs are not well suited to supporting work on causal inference with complex treatment packages, unless only very precise therapeutic elements are varied, to attain conceptual clarity.

One of the difficulties of RCTs is that there is an assumption that all relevant variables will be evenly distributed across treatment cells. This was not always the case and researchers only know to check for even distribution of variables known to impact outcome; the problem of the 'unknown unknowns'. For example discovering that treatment preferences affect alliances which in turn affects outcome (Iacoviello, et al., 2007) arguably makes the case that all prior research that did not control for treatment preferences across cells is invalid and should be discarded; an inconvenient but possible truth.

An issue that arises from dependence upon RCTs is that only therapies/researchers with access to sufficient funding for an RCT will have their therapies adequately tested (a weakness). There is a strength in testing therapy as it is practised by therapists and experienced by clients.

Assessing outcomes and processes with client completed questionnaires has shared method variance (Watson, 1984); a weakness. Ideally only one of these variables should be client assessed. Since Rogers' theory is about *client perceptions* of relationships this has to be client assessed and therefore outcomes ideally need to be assessed by an observer; to avoid the shared method variance problem. Employing suitably qualified persons to assess start and subsequent measures on a multi-diagnostic multi-axis system (American Psychiatric Association, 2000) is expensive and 'intrusive' in that clients need to be interviewed at least

twice by a consultant psychiatrist or similar. This research did not have access to this level of resource (a weakness).

8.4.7 Analysis and related issues

A clear weakness in this research was that the author was inexperienced at statistical analysis when the research began (September 2004) and was seeking to learn how to do this during the course of the research and analysis, with little support. A strength was the provision of access to an expert in statistical analysis (July 2008), although this expert also had lots of competing demands and a further expert in statistical analysis was provided (December 2010).

As described above there were very many variables in this naturalistic research and it was difficult to understand the impact of these upon the outcomes and predictor hypotheses.

A strength of this research was the comparison with published outcome studies, although the comparisons could have been more compelling if they were carried out in a statistical manner (non-central t) and by comparing the same measures across studies, since measures have been found to be differentially responsive to client change (Shadish & Sweeney, 1991, Minami, et al., 2007). As described above effect sizes can be ‘manipulated’ through tightly defining study population e.g. Westen et al. (2004), Dimidjian, et al. (2006). A less naïve researcher might have considered more tightly specifying samples e.g. dividing up the client sample into ‘low severity clients’ and ‘high severity clients’, although see Westen et al. (2004).

After the fact, another approach to predictor/mediator analysis (e.g. SEM) might have been better than the SPSS procedure used.

8.4.8 Summary of strengths and weaknesses

Although this research had some significant weaknesses it also had some strengths, not least that the author got on and did the research with a relatively low level of resources and captured outcomes data for a relatively large sample of clients. This was probably one of the largest outcome and process research projects using person-centred therapy in recent years. It was found that the outcomes were comparable with those in the literature for a relatively large number of clients; this is difficult, although not impossible to ignore. Whilst NICE may

recommend against 'counselling' for depression and anxiety, this research provides an alternative perspective that at least for some clients good outcomes were possible with person-centred psychotherapy. As a naturalistic study this research had good external validity; because this was a test of bona fide clients at actual counselling practices this does suggest the outcomes were generalisable, not least to similar clients at these practices. Whilst criticisms have been made of this type of methodology, data had been captured that provided a means to examine in some detail the evidence-base for such criticisms within this research. Having the data to examine and rebut criticisms is an important strength. Whilst flaws in the research became apparent because this was an evolving methodology it was possible to seek to correct and address flaws as they became apparent e.g. absence of a process measure, absence of NICE-recognised outcome measure, absence of check on prevalence of personality disorders etc. (Although changing the methodology part way through an experiment is never ideal).

To the present author's knowledge this was the first study to look at both 'congruent' and 'incongruent' clients in receipt of congruent empathy and unconditional positive regard that controlled for the pre-test level of depression and showed covariance between therapeutic relationships, as defined by Rogers, and depression outcome.

On balance, given these strengths and weaknesses, it would seem that some reliance can be put upon these outcome findings; for some clients, at least in the settings tested, it was possible for some clients of person-centred therapists to have comparatively good outcomes for some client presenting problems including depression, anxiety and distress and that the therapeutic relationship as defined by Rogers predicted depression outcome and may also have made a contribution to anxiety and distress outcome.

8.5 Discussion of results

In terms of considering the methodology and results of this research there are a number of points to discuss more broadly and they are as follows: Firstly, outcomes and related issues, secondly process-outcome correlation with the therapeutic relationship and related issues and thirdly, person-centred theory and related issues. Finally the implications of this research are discussed.

8.5.1 Outcomes and related issues

This research sought to test the clinical effectiveness of person-centred therapy using BDI-II, BAI and CORE-OM. It is important to note that clients don't just go to therapy for 'depression', 'anxiety', etc. Within the research were clients who had lots of sessions, yet they had started with 'non-clinical' scores on all outcome measures. Psychotherapy isn't just for 'ill' people with a DSM diagnosable condition. Often clients don't know what 'the problem is' and part of the role of therapy can be in seeking to understand what 'the problem' is. Whilst there is lots of evidence for the efficacy of CBT, the mode of operation for CBT is to have a treatment manual for a defined problem. Perhaps one of the strengths of person-centred therapy is in offering a way of working with clients who have no defined 'problem' and also a way of working with clients *simultaneously* upon many (perhaps non-defined) 'problems'. This research looked at a group of clients from several different perspectives, in terms of their 'problems' with depression, anxiety and distress; and found that many of those 'problems' were being 'worked on' simultaneously.

This research appeared to show, or perhaps 'rediscover', that person-centred therapy is amenable to quantitative research and in a manner this author believes *respected* clients. Offering clients a therapy that is not backed up by quantitative research evidence is perhaps *not respecting* clients.

Within the CBT literature there has been an *excitement* about 'sudden gains' as perhaps a route to demonstrating that cognitive change is responsible for improvements in depression e.g. Tang and DeRubeis (1999), Tang, et al. (2002), Tang, et al. (2005), Tang, et al. (2007); although the evidence seemed to suggest that 'sudden gains' are not unique to CBT and it

seems unlikely that a sudden cognitive change in CBT *causes* sudden change in depression symptoms as a mechanism unique to CBT, Jarrett, et al. (2007).

The phenomenon of 'sudden change' has been reported elsewhere: 'Patients with major depression did not change after the pre-therapy interviews, whereas anxiety patients changed substantially and significantly faster than the depressed patients' (Svartberg, Seltzer, Choi, & Stiles, 2001, p. 201).

Whilst mainly monitoring a similar group of clients from the different symptom group perspectives (depression, anxiety and distress) this research found evidence of 'sudden gains' and of different dynamics of change for the different groups of symptoms.

In terms of depression it was found that on average clients who had clinical depression scores did not get better whilst waiting for treatment and this finding appeared to coincide with findings in the literature, that relatively small changes occur in untreated samples (Elliott, et al., 2004, Minami, et al., 2007). There was evidence to suggest that on average clients experienced a large 'sudden gain' upon starting treatment (9.2 BDI-II units) and overall there was no significant relationship between time (days) spent in treatment and change in depression score. Taken together this suggested that some improvement process was underway in therapy (but not in wait) that was not necessarily time-dependent yet resulted in improvement in depression. In contrast there appeared to be some different phenomena underway with respect to changes in anxiety scores.

Clients with clinical anxiety scores had on average a statistically significant improvement during the wait period and the analysis suggested on average an improvement immediately following the *exploratory* session (5.3 BAI units) that deteriorated as waiting progressed. Following the first therapy session there was on average an improvement immediately following this session (4.7 BAI units) and there was no statistically significant relationship between passage of time (days) and change in BAI scores during treatment. Perhaps some improvement process was underway in therapy (but not in wait).

Clients with clinical distress scores had on average a statistically significant improvement during the wait period and the analysis suggested on average an improvement immediately following the exploratory session (1.9 CORE-OM units) and that there was no statistically significant relationship between passage of time (days) and change in CORE-OM scores during waiting. Following the first therapy session there was on average a large improvement

immediately following this session (5.4 CORE-OM units) and there was a statistically significant relationship between passage of time (days) and change in CORE-OM scores during treatment; longer treatment/more improvement. This may have been because of a ‘premature endings’ effect at the UCS that showed a trend towards statistical significance, perhaps because of a small sample size, and that this may have been because these UCS clients made up a greater proportion of the CORE-OM sample than the BDI-II or BAI samples. Perhaps some improvement process was underway in therapy (but not in wait)

Beck (1976, p. 309) criticised Rogers’ therapeutic approach on the basis that it ‘avoids a comprehensive model of psychopathology’; although Beck referred to Rogers, 1951 book and not the later theory statement (1957) or the more detailed theory statement that included a detailed description of ‘the process of breakdown and disorganisation’ (1959, pp. 228-229). The position Rogers and his team took earlier (1954) on researching the outcomes from ‘client-centred therapy’ was that:

...we should be able to make a series of statements of this order: “Client-centred therapy, operationally defined in this way, tends to produce changes a, b, d and f in clients. No change is found in characteristics c and e.” When such a series of statements is available, the profession and the lay public will be in a position to make a sound value judgement as to whether they regard as a “success” a process which produces these changes. (p. 31)

This research appeared to suggest that there were different dynamics of change for different symptoms, see also Stiles, et al. (1998). In some senses Beck and Rogers were both right and wrong. Beck was correct that Rogers, 1951, did not contain a comprehensive model of psychopathology, although incorrect that this was the case for ‘client-centred therapy’ because Rogers’ theory did attempt a statement of psychopathology (1959) and this was neither ‘comprehensive’ nor now a credible model of psychopathology cf. American Psychiatric Association (2000). In a sense Rogers was correct to say in a sense ‘let the market decide if they want client-centred therapy’ (although this assumes the Economist’s perfect flow of information, e.g. existence of comprehensive quantitative research) and the dynamics of psychological change appear to be much more complex than Rogers’ somewhat naïve view. For example Rogers’ theory stated that ‘For constructive personality change to occur, it is necessary that these conditions exist and continue over a period of time...no other conditions are necessary’ (1957, p. 96), yet this theory ignores variables that are suggested to

produce the majority of outcome (40% extratherapeutic change and 15% client expectancy, Lambert & Barley, 2002).

8.5.2 Process-outcome correlations with therapeutic relationship and related issues

This research sought to examine the therapeutic relationship (BLRI) as a putative predictor of outcome for depression, anxiety and distress. The findings for all three outcomes were initially non-significant and upon further examination a more complex picture emerged wherein depression at the subsequent session was significantly predicted by the therapeutic relationship, with some suggestion that this could also have been the case for anxiety and distress outcome. For future research it would perhaps be helpful to have a wider range of therapist abilities with a more even spread of clients per therapist and perhaps a different approach to the analysis, e.g. multi-level modelling. It is important to note that therapeutic relationships could be *causative* without satisfying prediction criteria, see for example Stiles (1988) and subsequent debate.

There was a significant finding for prediction by BLRI of depression outcome and some suggestion that prediction of anxiety and distress outcomes were 'close' or at least plausible. This is not the same as statistically significant evidence of a difference in outcomes, perhaps some of which *are* caused by the therapeutic relationship and some which *are not*. Nevertheless there appears the possibility that some symptoms are more readily facilitated by the therapeutic relationship than others; perhaps a more nuanced version of Rogers' idea (Rogers & Dymond, 1954) that different therapies may produce different outcomes, or even that different aspects of the relationship may facilitate different aspects of outcome, see Stiles et al. (1998). Although there is a vast amount of literature showing the equivalence paradox and as yet there is only limited evidence to suggest unique effects or mechanisms e.g. Siev and Chambless (2007), Connolly Gibbons, et al. (2009). As Stiles, et al. (1986) suggested this could be a consequence of inadequate research methods and measures, different techniques sharing a common core of processes, or therapy outcome and process studies have so far failed to observe differential effectiveness at the micro-process level. The results of dismantling studies appear to suggest the component parts of different therapies may have different effects (Castonguay & Beutler, 2006). As described above, person-centred therapy *now is not just* about the relationship, for example encouraging a client to *focus* on a feeling

is a form of behavioural/exposure therapy, empathising and repeating back something a client has said can lead to cognitive change, etc.

For the therapeutic relationship to 'predict', for example, depression outcome, there has to be some differential outcome between clients, otherwise there is no variance to explain (Stiles, 1988), i.e. therapeutic relationship significantly predicts subsequent depression. Rogers' theory does not state this, Rogers' theory was about 'constructive personality change'. Furthermore, Rogers' theory contains an apparent contradiction between clients being 'incongruent' and clients being the ones to judge the extent to which the facilitative conditions were offered (1967). Rogers sees *contact* as binary, it is or isn't in place. Subsequent person-centred theory and practice suggests 'contact' exists as a continuum and that gaining 'contact' is itself a therapeutic issue and that contemporary PCT offers approaches to addressing that issue (Wyatt & Sanders, 2002). Rogers' theory does state that *if* the relationship conditions are present *then* constructive personality change will occur; he did not state the terms in which outcomes would differ according to the relationship e.g. better relationship/quicker change or better relationship/lower depression at end, etc. Instead Rogers used the phrase 'If all six conditions are present then the greater the degree to which Conditions 2 to 6 exist, the more marked will be the constructive personality change in the client' (1957, p. 100). This implies both the extent and rate of change will be greater for greater degrees of conditions 2 to 6, although this was not made clear. The historical statement of Rogers' theory is not readily amenable to testing with contemporary research methodologies, although as Stiles, et al. (1998) pointed out, the absence of theory makes interpretation of results difficult, there maybe a need for a more updated theory statement in terms that contemporary researchers can respond to.

In general the best predictor of post-therapy scores is pre-therapy scores and in statistical terms a putative third variable cause has 'an awful lot of work to do' to be a predictor of outcome e.g. 20-26% of outcome variance was explained by starting scores. A different way of looking at this could be to do a median split of the sample based on BLRI scores and compare improvement on outcome scores, although Field (2009) cautions against this type of analysis. Wampold has played a key role in looking at the statistical techniques used to analyse the impact of the therapeutic relationship (e.g. Wampold & Brown, 2005, Baldwin, et al., 2007). The present study entered therapists in an ordinary least squares (OLS) regression analysis treating therapists as a fixed factor, however, 'if therapists are treated as fixed, the results are conditioned on the particular therapists included in the clinical trial' (Wampold &

Brown, 2005, p. 914). On this basis these authors recommended therapists should be treated as a random factor, randomly selected from a population of therapists, so that results could be generalised to therapists in general. There were examples of different approaches in the literature, for example Huppert, et al. (2001) used OLS and treated therapists as a fixed factor, whereas Blatt, et al. (1996) segregated therapists into groups and looked at group differences. Reviewing these different methods Wampold and Brown (2005) recommended using multi-level modelling, as used for example by Burns and Nolen-Hoeksema (1992) and Zuroff and Blatt (2006) to look at the impact of therapist factors. Wampold and Brown (2005) used multilevel modelling with patients nested within therapists, with therapists considered a random factor, taking into account that patients were not randomly assigned to therapists and found that about 8% of the variance in outcomes was due to the differences between therapists. Taking account of variability in outcomes due to initial severity this estimate reduced to 5% of outcome variance due to differences between therapists at an average level of initial severity. In regression analysis observations are intended to be independent and this cannot be the case for a clients nested within therapists design. In the later paper (Baldwin, et al., 2007) multilevel modelling was used to look at the relationship between therapeutic alliance and outcome and in particular to examine both within and between therapist variability, finding that therapist variability (between therapist variability) accounted for the alliance-outcome correlation, 'therapists who, on average, formed stronger alliances with their patients showed statistically significantly better outcomes than therapists who did not form as strong alliances... furthermore, within the caseload of a given therapist, the strength of the alliance did not significantly predict outcome, which suggests that patient variability in the alliance may be unimportant to outcome' (p. 849). Taken together this suggested research of this kind, rather than using OLS regression may be better served using multilevel modelling to analyse a sample, perhaps with a more even spread of clients to therapists.

8.5.3 Person-centred theory and related issues

This research suggested that good outcomes were possible for a wide range of client-presented symptoms. Yet some of the flaws in person-centred theory as stated (Rogers, 1957, 1959) have been described above and this theory no longer entirely covers common practice e.g. Sanders (2004). It would appear to be insufficient to rely on a fifty year old theory and Rogers thought this too. Rogers wrote 'What I shall offer is not a series of research findings,

but only the first step in that gradual approximation which we call science, a description of some observed phenomena which appears to be significant, and some highly tentative explanations of these phenomena.' (1947, p. 359). It was clear that Rogers was an empiricist (at least in 1947 and for some years afterwards) and he did not intend for research, theory and practice to stagnate and become dogmatic.

This research suggested 'counselling' was an effective intervention for depression cf. NICE (2009a). Giving a warning that 'there is no evidence for the effectiveness of counselling' potentially undermines the therapy and makes poor outcomes more likely (Lambert & Barley, 2002). This research suggested person-centred therapy was worth spending money on in further research, perhaps a 'person-centred control' condition could be offered to all psychotherapy research funded by the NHS in researcher allegiance balanced trials, given the large numbers of BACP member person-centred/humanistic trained therapists.

The importance of therapeutic relationships in psychotherapy, psychiatry, medicine and healthcare were discussed in the introduction. Person-centred psychotherapies have much to offer about the theory and practice of 'relationships' in many different spheres. For example the work of Kirsch and colleagues and others highlights the impact of interpersonal relationships upon pharmacotherapy and the need to control for relationship effects in pharmaceutical RCTs. There is a risk of systematic error in an RCT if warm, empathic, psychologically-minded psychiatrists give out medications and cold, un-empathic, biologically-minded psychiatrists give out placebos, e.g. Lambert (2004, pp. 234-239), Kaptchuk, et al. (2008), Kirsch, et al, (2008). Crits-Christoph, et al. (2009) suggested training therapists to form better alliance.

As described in the introduction large amounts of money are spent in the UK each year on treatment for mental health problems and upon NHS and non-NHS funded research into mental health. It is a competitive world to get money, power and influence and whilst the person-centred psychotherapies do appear to have much to offer, at present the influence of person-centred psychotherapies is relatively small. Despite the fact that in the UK there are large numbers of therapists trained in person-centred psychotherapies. The person-centred psychotherapies have fared poorly in evidence-based reviews of therapies and yet there is evidence of efficacy and effectiveness. It would appear legitimate to expect NICE to control for the amount of money spent on research, to control for reviewer allegiance and so on

(Scriven, 1998). Yet person-centred psychotherapies have more to offer than simply ‘squabbling over’ NICE reviews.

Within this research the author has used an ‘outcomes management approach’ that is person-centred and this appeared to have the benefits of transparently making progress, or its lack, a part of the therapeutic dialogue. One of the problems with the ‘good enough level’ of outcome (Barkham, et al., 2006) is that clients with low self esteem and personality issues may leave therapy ‘early’ and their ‘good enough level’ may not be ‘enough’; a therapeutic issue. There was some evidence of this at the UCS site where some clients had what therapists described as a ‘mutually agreed and satisfactory ending’ and yet the outcome measures showed these clients still had clinical level scores, and in some cases suggestions of suicidal ideation. This is just one area, of many, that person-centred psychotherapy could make an important contribution to the effectiveness of psychotherapies e.g. ‘clinically effective counselling’. Person-centred psychotherapies have a unique perspective that have value and deserve to exert a greater influence than at present. The lack of willingness by some person-centred therapists to participate in research of this kind may in fact hold back the development of person-centred therapy as a credible and contributing approach to therapy.

8.6 Implications of this research

The main implication of this research is that for a relatively large number of clients there was a good outcome and that there was some limited evidence that this may have been at least partly to do with the therapeutic relationship. This suggested person-centred therapy was both capable of being researched and worth researching. Whilst there has been relatively little money spent on researching person-centred therapy, this research suggested that there may be some merit in following up this research with a randomised controlled trial, perhaps an allegiance-controlled trial versus CBT. Future attempts at researching the therapeutic relationship as a putative mediator of outcome would appear to benefit from having a more standardised method for assessing 'outcome' e.g. all outcomes measured upon completion, all outcomes measured after a set period of time or perhaps ongoing outcome measurement and 'latent growth curve' modelling or similar using structural equation modelling (SEM).

This research began because the author wondered whether person-centred therapy was effective. Rather than it being 'the therapy' that was effective it appears there was considerable scope for a 'therapist' to be effective, despite or because of their theoretical model; an 'empirically supported therapist' (Weston, 2005, Lambert, 2006). It has become clear to me that at times I can be highly effective and at other times not; the challenge for me is to be the most effective I can be at what I do. Within a 'person-centred' approach there are times when I am doing what CBT therapists also do, albeit in perhaps a slightly different way, for example activating behaviour, challenging beliefs, building awareness of the inner dialogue, exposing clients to unpleasant feelings and so on, whilst empathising with and accepting them. Whilst my approach may come from a slightly different theoretical and methodological background, the practice may have some similarities and the outcome maybe somewhat similar. Person-centred psychotherapies (and therapists) can be clinically effective, although more research is required, particularly upon the possibility of a causative role for the therapeutic relationship.

9. References

- Ablon, J. S., & Marci, C. (2004). Psychotherapy process: The missing link: Comment on Westen, Novotny and Thompson-Brenner (2004). *Psychological Bulletin* , 130 (4) 664-668.
- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual 4th Edition, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association.
- American Psychological Association. (2001). *Publication Manual of the American Psychological Association (Fifth Edition)*. Washington DC: American Psychological Association.
- American Psychological Association, Division of Clinical Psychology, Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically validated psychological treatments. *The Clinical Psychologist* , 48 3-23.
- Antonuccio, D. O., Burns, D. D., & Danton, W. G. (2002). Antidepressants: A triumph of marketing over science. *Prevention & Treatment* , 5 (1). No pagination specified.
- Arntz, A., Dreesen, L., Schouten, E., & Weertman, A. (2004). Beliefs in personality disorders: A test with the personality disorder questionnaire. *Behaviour Research and Therapy* , 42 1215-1225.
- Assum, A. L., & Levy, S. J. (1948). Analysis of a nondirective case with followup interview. *Journal of Abnormal and Social Psychology* , 43 (1) 78-89.
- Axline, V. M. (1948). Play therapy and race conflict in young children. *Journal of Abnormal and Social Psychology* , 43 (3) 300-310.
- Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology* , 75 (6) 842-852.
- Barkham, M., Connell, J., Stiles, W. B., Miles, J. N., Margison, F., Evans, C., et al. (2006). Dose-effect relations and the responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology* , 74 (1) 160-167.

- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, J., et al. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, 69, 184-196.
- Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE system. *Counselling and Psychotherapy Research*, (1) 3-15.
- Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 215-220.
- Barrett-Lennard, G. T. (1998). *Carl Rogers' helping system: Journey and substance*. London: Sage.
- Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs*, 76 (43, Whole No. 562).
- Barrett-Lennard, G. T. (1978). The relationship inventory: Later development and adaptions. *JSAS Catalogue of selected documents in psychology*, 8 68 MS 1732.
- Barrett-Lennard, G. T. (1986). The relationship now: Issues and advances in theory, method and use. In L. S. Greenberg, & W. M. Pinsof (Editors), *The psychotherapeutic process: A research handbook* (pp. 439-476). New York: Guilford Press.
- Barrowclough, C., King, P., Colville, J., Russell, E., Burns, A., & Tarrier, N. (2001). A randomised trial of the effectiveness of cognitive-behavioural therapy and supportive counselling for anxiety symptoms in older adults. *Journal of Consulting and Clinical Psychology*, 69 (5) 756-762.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. London: Penguin.
- Beck, A. T., & Steer, R. A. (1993). *Beck Anxiety Inventory Manual*. San Antonio, USA: The Psychological Corporation Harcourt Brace & Company.
- Beck, A. T., Butler, A. C., Brown, G. K., Dahlgard, K. K., Newman, C. F., & Beck, J. S. (2001). Dysfunctional beliefs discriminate personality disorders. *Behaviour research and therapy*, 1213-1225.

- Beck, A. T., Freeman, A., Davis, D. D., & Associates. (2004). *Cognitive therapy of personality disorders (Second Edition)*. London: The Guilford Press.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *The Beck Depression Inventory - Second Edition (BDI-II) Manual*. San Antonio: The Psychological Corporation.
- Bergin, A. E. (1977). Introduction. In A. S. Gurman, & A. M. Razin (Editors), *Effective psychotherapy: A handbook of research* (pp. xv-xvi). Oxford: Pergamon Press.
- Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., et al. (2004). Therapist variables. In (M. Lambert, *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change 4th Edition* (pp. 227-306). New York: John Wiley and Sons.
- Blatt, S. J., Zuroff, D. C., Quinlan, D. M., & Pilkonis, P. (1996). Interpersonal factors in brief treatment of depression: Further analyses of the NIMH treatment of depression collaborative research programme. *Journal of Consulting and Clinical Psychology*, 64, 162-171.
- Bockman (Editor), L. (2000). *Validity and social experimentation*. London: Sage.
- Bohart, A. C. (2007). An alternative view of concrete operating procedures from the perspective of the client as active self-healer. *Journal of Psychology Integration*, 17 (1) 125-137.
- Bohart, A. C., & Tallman, K. (1999). *How clients make therapy work: The process of active self-healing*. Washington DC: American Psychological Association.
- Bohart, A. C., Elliott, R., Greenberg, L. S., & Watson, E. C. (2002). Empathy. In J. C. Norcross (Editor), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 89-108). New York: Oxford University Press.
- Bohart, A. C., O'Hara, M., & Leitner, L. (1998). Empirically violated treatments: Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research*, 8, 141-157.
- Borkovec, T. D., & Whisman, M. A. (1996). Psychosocial treatment for generalised anxiety disorder. In M. Mavissakalian, & R. Prien (Editors), *Long-term treatment of anxiety disorders*. Washington DC: American Psychiatric Association.

British Association for Counselling and Psychotherapy. (2009). *About BACP*. Retrieved October 12, 2009, from BACP: http://www.bacp.co.uk/about_bacp/profile_history.php

British Association for Counselling and Psychotherapy. (2004a). *Ethical guidelines for good practice in counselling and psychotherapy*. Rugby: British Association for Counselling and Psychotherapy.

British Association for Counselling and Psychotherapy. (2004b). *Ethical guidelines for researching counselling and psychotherapy*. Rugby: British Association for Counselling and Psychotherapy.

Brown, G. S. (2007). Implications of therapist effects for employers and health plans. *American Psychological Association Convention*. Presentation slides. San Francisco: Centre for Clinical Informatics.

Brown, W. A. (2002). Are antidepressants as ineffective as they look? *Prevention & Treatment*, 5 (1). No pagination specified.

Bryant, R. A., Harvey, A. G., Dang, S. T., Sackville, T., & Basten, C. (1998). Treatment of acute stress disorder: A comparison of cognitive-behavioural therapy and supportive counselling. *Journal of Consulting and Clinical Psychology*, 66 (5) 862-866.

Bryant, R. A., Sackville, T., Dang, S. T., Moulds, M., & Guthrie, R. (1999). Treating acute stress disorder: An evaluation of cognitive behavioural therapy and supportive counselling techniques. *American Journal of Psychiatry*, 156 (11) 1780-1786.

Burlingame, G. M., Theobald McClenon, D., & Alonso, J. (2010). Cohesion in group therapy. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter5>.

Burns, D. D., & Nolen-Hoeksema, S. (1992). Therapeutic empathy and recovery from depression in cognitive-behavioural therapy: A structural equation model. *Journal of Consulting and Clinical Psychology*, 60 (3) 441-449.

Burns, D. D., & Spangler, D. L. (2000). Does psychotherapy homework lead to improvements in depression in cognitive-behavioural therapy or does improvement lead to

increased homework compliance? *Journal of Consulting and Clinical Psychology* , 68 (1) 46-56.

Burton, A. (1972). *Twelve therapists*. San Francisco: Jossey-Bass.

Butler, A. C., Brown, G. K., Beck, A. T., & Grisham, J. R. (2002). Assessment of dysfunctional beliefs in borderline personality disorder. *Behaviour Research and Therapy* , 1231-1240.

Cambell, D. T., & Russo, M. J. (1999). *Social Experimentation*. London: Sage.

Carrol, L. (1865/1994). *Alice's adventures in Wonderland - Penguin Popular Classics Edition*. London: Penguin.

Castonguay, L. G., & Beutler (Editors), L. E. (2006). *Principles of therapeutic change that work*. Oxford: Oxford University Press.

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology* , 64 (3) 497-504.

Chambless, D. L. (1996). In defence of dissemination of empirically supported psychological interventions. *Clinical Psychology: Science and Practice* , 3 (3) 230-235.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology* , 66 7-18.

Chodorkoff, B. (1954). Self-perception, perceptual defence and adjustment. *Journal of Abnormal and Social Psychology* , 49 (4) 508-512.

Clark, D. M., Fairburn, C. G., & Wessely, S. (2007). Psychological treatment outcomes in routine NHS services a commentary on Stiles et al (2007). *Psychological Medicine* , Published online: 9 October. doi: 10.1017 S0033291707001869.

Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Editor), *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change* (pp. 194-226). New York: John Wiley and Sons.

Cloitre, M., Chase, S.-M. K., Miranda, R., & Chemtob, C. M. (2004). Therapeutic alliance, negative mood regulation and treatment outcome in child abuse related post traumatic stress disorder. *Journal of Consulting and Clinical Psychology* , 72 (3) 411-416.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences (Second edition)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Connolly Gibbons, M. B., Crits-Christoph, P., Barber, J. P., Wiltsey Stirman, S., Gallop, R., Goldstein, L. A., et al. (2009). Unique and common mechanisms of change across cognitive and dynamic therapies. *Journal of Consulting and Clinical Psychology* , 77 (5) 801-813.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Cooper, M. (2008). *Essential findings in counselling and psychotherapy: The facts are friendly*. London: Sage.

Cooper, M., O'Hara, M., Schmid, P. F., & Wyatt, G. (2007). *The handbook of person-centred psychotherapy and counselling*. New York: Palgrave Macmillan.

Cowen, E. L., & Combs, A. W. (1950). Followup study of 32 cases treated by nondirective psychotherapy. *Journal of Abnormal and Social Psychology* , 45 232-258.

Crits-Christoph, P., Gallop, R., Temes, C. M., Woody, G., Ball, S. A., Martino, S., et al. (2009). The alliance in motivational enhancement therapy and counselling as usual for substance use problems. *Journal of Consulting and Clinical Psychology* , 77 (6) 1125-1135.

Crits-Christoph, P., Wilson, G. T., & Hollon, S. D. (2005). Empirically supported psychotherapies: Comment on Westen, Novotny and Thompson-Brenner (2004). *Psychological Bulletin* , 131 412-417.

Deacon, B. J., & Abramowitz, J. S. (2004). Cognitive and behavioural treatments for anxiety disorders: A review of meta-analytic findings. *Journal of Clinical Psychology* , 60 (4) 429-441.

Deale, A., Chalder, T., Marks, I., & Wessely, S. (1997). Cognitive behaviour therapy for chronic fatigue syndrome: A randomised controlled trial. *American Journal of Psychiatry* , 154 408-414.

- Dekeyser, M., Prouty, G., & Elliott, R. (2008). Pre-therapy process and outcome: A review of research instruments and findings. *Person-centred and experiential psychotherapies* , 37-55.
- Des Jarlais, D. C., Lyles, C., Crepaz, N., & TREND Group. (2004). Improving the reporting quality of nonrandomised evaluations of behavioural and public health interventions: The TREND statement. *American Journal of Public Health* , 94 (3) 361-366.
- DiMatteo, M. R., Sherbourne, C. D., Hays, R. D., Ordway, L., Kravitz, R. L., McGlynn, E. A., et al. (1993). Physicians' characteristics influence patients' adherence to medical treatment: Results from the medical outcomes study. *Health Psychology* , 12 (2) 93-102.
- Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., et al. (2006). Randomised trial of behavioural activation, cognitive therapy and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology* , 74 (4) 658-670.
- Dobson, K., & Craig, K. (1998). *Empirically supported therapies: Best practice in professional psychology*. London: Sage.
- Dollard, J., & Mowrer, O. H. (1947). A method of measuring tension in written documents. *Journal of Abnormal and Social Psychology* , 42 (1) 3-32.
- Ehlers, A. C., Hackmann, A., McManus, F., Fennell, M., Herbert, C., & Mayou, R. (2003). A randomised controlled trial of cognitive therapy, a self help booklet and repeated assessments as early interventions for posttraumatic stress disorder. *Archives of General Psychiatry* , 60 1024-1032.
- Elkin, I., Shea, M. T., Watkins, J. T., & Imber, S. D. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Programme: General Effectiveness of Treatments. *Archives of General Psychiatry* , 46 (11) 971-982.
- Elliott, R. (2010). Psychotherapy change process research: Realising the promise. *Psychotherapy Research* , 20 (2) 123-135.
- Elliott, R. (2008, April Tuesday 8th). *Research*. Retrieved January Ssturday 23rd , 2010, from Robert Elliott's PE-EFT Blog: <http://pe-eft.blogspot.com/2008/04/individualized-reliable-change-using.html>

- Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2010). Empathy. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online resource, available at <http://174.140.153.165/Norcross.aspx#chapter6>.
- Elliott, R., Freire, E., & Cooper, M. (2008). Empirical Support for Person-Centred/Experiential Psychotherapies: Meta-analysis Update 2008. *14th BACP Research Conference Research and Regulation - Towards an Evidence-Based Profession* (p. 23). Cardiff: British Association for Counselling and Psychotherapy.
- Elliott, R., Greenberg, L. S., & Lietaer, G. (2004). Research on Experiential Psychotherapies. In M. J. Lambert (Editor), *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change* (pp. 493-539). New York: John Wiley & Sons.
- Evans, C., Connell, J., Barkham, M., Margison, F. M., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry* , 180 51-60.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology* , 16 319-324.
- Farber, B. A., & Doolin, E. M. (2010). Positive regard and affirmation. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter8>.
- Farber, B. A., & Lane, J. S. (2002). Positive Regard. In J. C. Norcross (Editor), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 175-194). New York: Oxford University Press.
- Faw, V. (1949). A psychometric method of teaching psychology. *American Psychologist* , 4 104-109.
- 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* , 67 578-582.

- Fiedler, F. E. (1949). *A comparison of therapeutic relationships in psychoanalytic, non-directive and Adlerian therapy*. Unpublished doctoral dissertation: University of Chicago.
- Fiedler, F. E. (1950). A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *Journal of Consulting Psychology* , 14 (6) 436-445.
- Field, A. (2005). *Discovering statistics using SPSS (and sex, drugs and rock 'n' roll) (Second Edition)*. London: Sage.
- Field, A. (2009). *Discovering statistics using SPSS Third Edition*. London: Sage.
- Foa, E. B., & Kozak, M. J. (1991). Emotional processing: Theory, research and clinical implications for anxiety disorders. In J. D. Safran, & L. S. Greenberg (Editors), *Emotion, psychotherapy and change* (pp. 21-49). London: Guilford Press.
- Frank, J. D. (1973). *Persuasion and healing*. Baltimore: Johns Hopkins University.
- Friedlander, M. L., Escudero, V., Heatherington, L., & Diamond, G. M. (2010). Alliance in couple and family therapy. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter4>.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland* , 15 246-263.
- Garfield, S. L., & Bergin, A. E. (1986). *Handbook of psychotherapy and behaviour change third edition*. New York: Wiley.
- Gendlin, E. (1962). *Experiencing and the creation of meaning: A philosophical and psychological approach to the subjective*. Illinois: Northwestern University Press.
- Gendlin, E. (1978). *Focusing: How to gain direct access to your body's knowledge*. London: Bantam.
- Gendlin, E. (1996). *Focusing-oriented psychotherapy: A manual of the experiential method*. New York: Guilford Press.
- Gilbody, S., Richards, D., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *British Journal of General Practice* , 57 650-652.

- Goldman, R., Greenberg, L. S., & Pos, A. E. (2005). Depth of emotional experience and outcome. *Psychotherapy Research*, 15 (3) 238-249.
- Goldman, R., Greenberg, L., & Angus, L. (2006). The effects of adding emotion-focused interventions to the therapeutic relationship in the treatment of depression. *Psychotherapy Research*, 16 537-549.
- Gomes-Schwartz, B. (1978). Effective ingredients in psychotherapy: Prediction of outcome from process variables. *Journal of Consulting and Clinical Psychology*, 46 (5) 1023-1035.
- Gone, J. P., & Alcantra, C. (2007). Identifying effective mental health interventions for American Indians and Alaska natives: A review of the literature. *Cultural Diversity and Ethnic Minority Psychology*, 13 (4) 356-363.
- Gould, R. A., Otto, M. W., Pollack, M. H., & Yap, L. (1997). Cognitive behavioural and pharmacological treatment of generalised anxiety disorder: A preliminary meta-analysis. *Behaviour Therapy*, 28 285-305.
- Grant, B. (1990). Principled and instrumental non-directiveness in person-centred and client-centred therapy. *Person-Centred Review*, 5 (1) 77-88.
- Greenberg, L. S. (2002). *Emotion-focused therapy: Coaching clients to work through their feelings*. Washington DC: American Psychological Association.
- Greenberg, L. S., Rice, L. N., & Elliott, R. (1993). *Facilitating emotional change: The moment-by-moment process*. New York: The Guilford Press.
- Greenberg, R. P. (2002). Reflections on the emperor's new drugs. *Prevention & Treatment*, 5 (1). No pagination specified.
- Gurman, A. S. (1977). The patient's perception of the therapeutic relationship. In A. S. Gurman, & A. M. Razin (Editors), *Effective Psychotherapy: A handbook of research* (pp. 503-543). Oxford: Pergamon Press.
- Haigh, G. (1949). Defensive behaviour in client-centred therapy. *Journal of Consulting Psychology*, 13 (3) 181-189.
- Hanlon, T. E., Hofstaetter, P. R., & O'Connor, J. P. (1954). Congruence of self and ideal self in relation to personality adjustment. *Journal of Consulting Psychology*, 18 (3) 215-218.

- Hart, J. T., & Tomlinson, T. M. (1970). *New directions in client-centred therapy*. Boston: Houghton-Mifflin.
- Haugh, S., & Merry (Editors), T. (2001). *Rogers' therapeutic conditions: Evolution, theory and practice: Volume 2: Empathy*. Ross-on-Wye: PCCS Books.
- Hawkins, P., & Shohet, R. (2006). *Supervision in the helping professions (Third Edition)*. Maidenhead, Berkshire: McGraw-Hill.
- Hawley, L., Ho, M.-H. R., Zuroff, D. C., & Blatt, S. J. (2006). The relationship of perfectionism, depression and therapeutic alliance during treatment for depression: Latent difference score analysis. *Journal of Consulting and Clinical Psychology*, 74 (5) 930-942.
- Hayes, A. M., Castonguay, L. G., & Goldfried, M. R. (1996). The study of change in psychotherapy: A reexamination of the process-outcome correlation paradigm. Comment on Stiles and Shapiro (1994). *Journal of Consulting and Clinical Psychology*, 64 (5) 909-914.
- Hayes, J. A., Gelso, C. J., & Hummel, A. M. (2010). Managing countertransference. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter12>.
- Haynes, S. N., & O'Brien, W. O. (2000). *Principles of behavioural assessment: A functional approach to psychological assessment*. New York: Plenum.
- Hays, R. D., Ordway, L., & Di Matteo, R. (1992). Antecedents of adherence to medical recommendations: Results from the medical outcomes study. *Journal of Behavioural Medicine*, 15 (5) 447-468.
- Hill, A., & Brettle, A. (2008). Counselling in primary care: A systematic review. *14th Annual BACP Research Conference*. Cardiff: BACP.
- Hill, C. E., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. J. Lambert (Editor), *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change Fifth Edition* (pp. 84-135). New York: John Wiley & Sons, Inc.
- Hoffman, A. E. (1949). A study of reported behaviour changes in counselling. *Journal of Consulting Psychology*, 13 (3) 190-195.

Hollon, S. D., DeRubeis, R. J., Shelton, R. C., & Weiss, B. (2002). The emperor's new drugs: Effect size and moderation effects. *Prevention & Treatment*, 5 (1). No pagination specified.

Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Editor), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37-70). New York: Oxford University Press.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counselling Psychology*, 36 (2) 223-233.

Horvath, A. O., Del Re, A. C., Fluckiger, C., & Symonds, D. (2010). Alliance in individual psychotherapy. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter2>.

House, E. R. (1980). *Evaluating with validity*. Beverly Hills: Sage.

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables and cognitive-behavioural therapy outcome in a multicentre trial for panic disorder. *Journal of Consulting and Clinical Psychology*, 69 (5) 747-755.

Iacoviello, B. M., McCarthy, K. S., Barrett, M. S., Rynn, M., Gallop, R., & Barber, J. P. (2007). Treatment preferences affect the therapeutic alliance: Implications for randomised controlled trials. *Journal of Consulting and Clinical Psychology*, 75 (1) 194-198.

Imber, S. D., Pilkonis, P. A., Sotsky, S. M., Elkin, I., Watkins, J. T., Collins, J. F., et al. (1990). Mode-specific effects among three treatments for depression. *Journal of Consulting and Clinical Psychology*, 58 (3) 352-359.

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.

Jarrett, R. B., Vittengl, J. R., Doyle, K., & Clark, L. A. (2007). Changes in cognitive content during and following cognitive therapy for recurrent depression: Substantial and enduring but not predictive of change in depressive symptoms. *Journal of Consulting and Clinical Psychology*, 75 (3) 432-446.

Kaptchuk, T. J., Kelley, J. M., Conboy, L. A., Davis, R. B., Kerr, C. E., Jacobson, E. E., et al. (2008). Components of placebo effect: Randomised controlled trial in patients with irritable bowel syndrome. *British Medical Journal*, Published online 3 April 2008. doi: 10.1136/bmj.39524.439618.25.

Kazantzis, N., Deane, F. P., & Ronan, K. R. (2000). Homework assignments in cognitive and behavioural therapy: A meta-analysis. *Clinical Psychology: Science and Practice*, 7 189-202.

Kazantzis, N., Ronan, K. R., & Deane, F. P. (2001). Concluding causation from correlation: Comment on Burns and Spangler (2000). *Journal of Consulting and Clinical Psychology*, 69 (6) 1079-1083.

Kiesler, D. J., Klein, M. H., Mathieu, R. L., & Schoeninger, D. (1967). Constructive personality change for therapy and control patients. In C. R. (Editor), *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics* (pp. 251-294). Madison: University of Wisconsin Press.

King, M., Sibbald, B., Ward, E., Bower, P., Lloyd, M., Gabbay, M., et al. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care [Monograph]. *Health Technology Assessment*, 4 (19) 1-84.

Kirsch, I. (2002). Yes, there is a placebo effect, but is there a powerful antidepressant drug effect? *Prevention & Treatment*, 5 (1). No pagination specified.

Kirsch, I., & Sapirstein, G. (1998). Listening to prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment*, 1 (2). No pagination specified.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration. *PLoS*, e45. doi:10.1371/journal.pmed.0050045.

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The Emperor's New Drugs: An analysis of antidepressant medication submitted to the US Food and Drug Administration. *Prevention and Treatment* 5 (1), No pagination specified.

Kirsch, I., Scoboria, A., & Moore, T. J. (2002). Antidepressants and placebos: Secrets, revelations and unanswered questions. *Prevention & Treatment* , 5 (1). No pagination specified.

Kirschenbaum, H. (1979). *On becoming Carl Rogers*. New York: Delacorte.

Kirschenbaum, H. (2007). *The life and work of Carl Rogers*. Ross-on-Wye: PCCS Books.

Kirschenbaum, H., & Land Henderson, V. (1990). *The Carl Rogers Reader*. London: Constable.

Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., et al. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology* , 71 (6) 997-1006.

Klein, M. H., Kolden, G. G., Michels, J. L., & Chisholm-Stockard, S. (2002). Congruence. In J. C. Norcross (Editor), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 195-216). New York: Oxford University Press.

Klein, M. H., Mathieu, P. L., Gendlin, E. T., & Kiesler, D. J. (1969). *The experiencing scale: A research and training manual (volume 1)*. Wisconsin: Wisconsin Psychiatric Institute.

Kolden, G. G., Klein, M. H., Wang, C.-C., & Austin, S. B. (2010). Congruence/Genuineness. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter9>.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomised clinical trials. *Archives of General Psychiatry* , 59 877-883.

Lambert (Editor), M. J. (2004). *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change Fifth Edition*. New York: John Wiley & Sons, Inc.

Lambert, M. J. (2006). Effective methods for reducing treatment failure and enhancing patient outcome: The importance of tracking treatment response. *Maine Consortium*.

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17 (1) 1-14.

Lambert, M. J. (1992). Psychotherapy outcome research: Implications for integrative and eclectic therapists. In J. C. Norcross, & M. R. Goldfried (Editors), *Handbook of psychotherapy integration* (pp. 94-129). New York: Basic Books.

Lambert, M. J., & Barley, D. E. (2002). Research summary on the therapeutic relationship and psychotherapy outcome. In J. C. Norcross (Editor), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 17-32). New York: Oxford University Press.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. (Editor), *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change Fifth Edition* (pp. 139-193). New York: Wiley.

Lambert, M. J., & Shimokawa, K. (2010). Collecting client feedback. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter10>.

Lambert, M. J., DeJulio, S. S., & Stein, D. M. (1978). Therapist interpersonal skills: Process, outcome, methodological considerations and recommendations for future research. *Psychological Bulletin*, 85 (3) 467-489.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69 (2) 159-172.

Lambert, M. J., Shapiro, D. A., & Bergin, A. E. (1986). The effectiveness of psychotherapy. In S. L. Garfield, & A. E. Bergin (Editors), *Handbook of psychotherapy and behaviour change (3rd Edition)* (pp. 157-212). New York: Wiley.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. L. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy. *Psychotherapy Research*, 11 49-68.

Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., et al. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy* , 9 91-103.

Lawrence, S. (2005). Anxiety guidelines: Why counselling doesn't get a look in. *Counselling and Psychotherapy* , 16 (2) 34-37.

Leach, C., Lucock, M., Barkham, M., Stiles, W. B., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology* , 45 153-166.

Lesser, W. M. (1961). The relationship between counselling progress and empathic understanding. *Journal of Counselling Psychology* , 8 330-336.

Levant, R. F., & Shlien (Editors), J. M. (1984). *Client-centred therapy and the person-centred approach: New directions in theory, research and practice*. Westport. Connecticut: Praeger.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice* , 6 95-106.

Luborsky, L., Singer, J., & Luborsky, L. (1975). Comparative studies of psychotherapy. *Archives of General Psychiatry* , 32 995-1008.

Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology* , 66 (2) 400-410.

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Consulting Psychology* , 54 (1) 32-39.

Maguire, U. (1973). Counselling effectiveness: A critical discussion. *British Journal of Guidance and Counselling* , 1 (1) 38-50.

Malik, M., Beutler, L. E., Alimohamed, S., & Gallagher-Thompson, D. (2003). Are all cognitive therapies alike? A comparison of cognitive and noncognitive process and implications for the application of empirically supported treatments. *Journal of Consulting and Clinical Psychology* , 71 (1) 150-158.

- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68 (3) 438-450.
- Matarazzo, R. G. (1971). Research on the teaching and learning of psychotherapeutic skills. In A. E. Bergin, & S. L. Garfield (Editors), *A handbook of psychotherapy and behaviour change: An empirical analysis*. New York: Wiley.
- McKay, K. M., Imel, Z. E., & Wampold, B. E. (2006). Psychiatrist effects in the pharmacological treatment of depression. *Journal of Affective Disorders*, 92 287-290.
- McLeod, J. (2003). *Doing counselling research (Second edition)*. London: Sage.
- Mearns, D. J. (1999). Person-centred therapy with configurations of self. *Counselling*, 10 125-130.
- Mearns, D. J., & Cooper, M. (2005). *Working at relational depth in counselling and psychotherapy*. London: Sage.
- Mearns, D. J., & Thorne, B. (2000). *Person-centred therapy today: New frontiers in theory and practice*. London: Sage.
- Mearns, D., & McLeod, J. (1984). A person-centred approach to research. In R. F. (Editors), *Client-centred therapy and the person-centred approach: New directions in theory, research and practice* (pp. 370-389). New York: Praeger.
- Meltzoff, J., & Kornreich, M. (1970). *Research in psychotherapy*. New York: Atherton Press.
- Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology*, 76 (1) 116-124.
- Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology*, 75 (2) 232-243.

- Missirlian, T. M., Toukmanian, S. G., Warwar, S. H., & Greenberg, L. S. (2005). Emotional arousal, client perceptual processing and the working alliance in experiential psychotherapy for depression. *Journal of Consulting and Clinical Psychology* , 73 (5) 861-871.
- Mitchell, K. M., Bozarth, J. D., & Krauft, C. (1977). A reappraisal of the therapeutic effectiveness of accurate empathy, non-possessive warmth and genuineness. In A. M. A S Gurman & Razin, *Effective psychotherapy: A handbook of research* (pp. 482-502). New York: Pergamon Press.
- Moerman, D. E. (2002). 'The loaves and the fishes': A comment on 'The emperor's new drugs: An analysis of anti-depressant medication data submitted to the US Food and Drug Administration'. *Prevention & Treatment* , 5 (1). No pagination specified. .
- Moncrieff, J. (2006). Psychiatric drug promotion and the politics of neoliberalism. *British Journal of Psychiatry* , 180 301-302.
- Moncrieff, J., & Kirsch, I. (2005). Efficacy of antidepressants in adults. *British Medical Journal* , 331 16 July 2005 155-157.
- Mullin, T., Barkham, M., Mothersole, G., Bewick, B. M., & Kinder, A. (2006). Recovery and improvement benchmarks for counselling and psychological therapies in routine care. *Counselling and Psychotherapy Research* , 6 (1) 68-80.
- Munoz, R. A. (2002). Comment on Kirsch, Moore, Scoboria and Nicholls (2002). *Prevention & Treatment* , 5 (1). No pagination specified.
- Newman, F. L. (1994). Disabuse of the drug metaphor: Introduction. *Journal of Consulting and Clinical Psychology* , 62 (5) 941-941.
- Newman, M. G., Stiles, W. B., Janeck, A., & Woody, S. R. (2006). Integration of therapeutic factors in anxiety disorders. In L. G. Castonguay, & L. E. Beutler (Editors), *Principles of therapeutic change that work* (pp. 187-200). Oxford: Oxford University Press.
- NICE. (2004a). *Anxiety: management of anxiety (panic disorder, with or without agoraphobia, and generalized anxiety disorder) in adults in primary, secondary and community care CG022*. London: <http://www.nice.org.uk/page.aspx?o=cg022&c=mental> downloaded on 10th May 2006.

NICE. (2005a). *Anxiety: Management of post-traumatic stress disorder in adults in primary, secondary and community care CG026*. London:

<http://www.nice.org.uk/page.aspx?o=CG026&c=mental> downloaded on 10th May 2006.

NICE. (2008). *Borderline Personality Disorder (draft)*. London.

NICE. (2004b). *Depression – management of depression in primary and secondary care CG022*. London: <http://www.nice.org.uk/page.aspx?o=cg023&c=mental> downloaded on 10th May 2006.

NICE. (2002a). *Depression and anxiety – computerised cognitive behavioural therapy (CCBT) CG051*. London: <http://www.nice.org.uk/page.aspx?o=appraisals.completed> downloaded on 10th May 2006.

NICE. (2009b). *Depression in adults (update): consultation comments and table of responses*. London: NICE downloaded from <http://www.nice.org.uk/guidance/index.jsp?action=download&o=44991> on 22 July 2009.

NICE. (2009a). *Depression in Adults (Update): Draft full guideline for consultation*. London: <http://www.nice.org.uk/guidance/index.jsp?action=download&o=43311> downloaded on Tuesday 3rd March 2009.

NICE. (2005b). *Depression in children and young people – identification and treatment in primary, community and secondary care CG028*. London:

<http://www.nice.org.uk/page.aspx?o=cg028&c=mental> downloaded on 10th May 2006.

NICE. (2004c). *Eating disorders – core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders CG009*. London:

<http://www.nice.org.uk/page.aspx?o=cg009&c=mental> downloaded on 10th May 2006.

NICE. (2005c). *Obsessive-compulsive disorder – core interventions in the treatment of obsessive compulsive disorder and body dysmorphic disorder CG031*. London:

<http://www.nice.org.uk/page.aspx?o=cg031&c=mental> downloaded on 10th May 2006.

NICE. (2002b). *Schizophrenia: Core interventions in the treatment and management of schizophrenia in primary and secondary care CG001*. London:

<http://www.nice.org.uk/page.aspx?o=CG001&c=mental> downloaded on 10th May 2006.

NICE. (2004d). *Self harm: the short term physical and psychological management and secondary prevention of self harm in primary and secondary care CG016*. London: <http://www.nice.org.uk/page.aspx?o=cg016&c=mental> downloaded on 10th May 2006.

Nielsen, T., Karpatschof, B., & Kreiner, S. (2005). Regression to the mean effect: When to be concerned and how to correct for it. *Nordic Psychology*, 59 (3) 231-250.

Norcross (Editor), J. C. (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients*. New York: Oxford University Press.

Norcross, J. C. (2010). *Evidence-based therapy relationships*. Retrieved January 19, 2011, from National Registry of Evidence-based Programmes and Practices (NREPP): <http://www.nrepp.samhsa.gov/Norcross.aspx>

Norcross, J. C. (2011). *Psychotherapy relationships that work (2nd Edition)*. New York: Oxford University Press.

Norcross, J. C., & Goldfried, M. R. (1992). *Handbook of psychotherapy integration*. New York: Oxford University Press.

Norcross, J. C., & Lambert, M. J. (2010). Evidence-based therapy relationships. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter1>.

Norcross, J. C., & Wampold, B. E. (2010). Adapting the relationship to the individual patient. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter13>.

Norris, N. (1990). *Understanding educational evaluation*. New York: St Martin's Press.

Orlinsky, D. E., & Howard, K. L. (1986). Process and outcome in psychotherapy. In S. L. Garfield, & A. E. Bergin (Editors), *Handbook of psychotherapy and behaviour change third edition* (pp. 311-381). New York: Wiley.

Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Process and outcome in psychotherapy. In M. J. Lambert (Editor), *Bergin and Garfield's handbook of psychotherapy and behaviour change (Fifth Edition)* (pp. 307-389). New York: Wiley.

- Patterson, C. H. (1984). Empathy, warmth and genuiness in psychotherapy: A review of reviews. *Psychotherapy* , 21 431-438.
- Polanyi, M. (1958). *Personal knowledge*. Chicago: University of Chicago Press.
- Quinn, R. D. (1950). *Psychotherapists expressions as an index to the quality of early therapeutic relationships established by representatives of the non-directive, Adlerian and psychoanalytic schools*. Unpublished doctoral dissertation: University of Chicago.
- Qureshi, N. N., Hatcher, J., Chaturvedi, N., & Jafar, T. H. (2007). Effect of general practitioner education on adherence to antihypertensive drugs: Cluster randomised controlled trial. *British Medical Journal* , 335 1030-1033.
- Raimy, V. C. (1948). Self reference in counselling interviews. *Journal of Consulting Psychology* , 12 (3) 153-163.
- Raskin, N. J. (1949a). An analysis of six parallel studies of the therapeutic process. *Journal of Consulting Psychology* , 13 (3) 206-220.
- Raskin, N. J. (1949b). The development of the 'parallel studies' project. *Journal of Consulting Psychology* , 13 (3) 154-156.
- Rehm, L. P. (2002). How can we better disentangle placebo and drug effects? *Prevention & Treatment* , 5 (1). No pagination specified.
- Rogers (Editor), C. R. (1967). *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics*. Madison, Milwaukee and London: The University of Wisconsin Press.
- Rogers, C. R. (1949). A coordinated research in psychotherapy: A nonobjective introduction. *Journal of Consulting Psychology* , 13 (3) 149-153.
- Rogers, C. R. (1959). A theory of therapy, personality and interpersonal relationships as developed in the client-centred framework. In S. Koch (Editor), *Psychology: A study of science Volume 3: Formulations of the person and social context* (pp. 184-256).
- Rogers, C. R. (1980). *A way of being*. Boston: Houghton Mifflin.
- Rogers, C. R. (1978). *Carl Rogers on personal power: Inner strength and its revolutionary impact*. London: Constable.

Rogers, C. R. (1951). *Client-centred therapy: Its current practice, implications and theory*. London: Constable.

Rogers, C. R. (1942). *Counselling and Psychotherapy*. Boston: Houghton Mifflin.

Rogers, C. R. (1931). *Measuring personality adjustment in children aged nine to thirteen years of age*. New York: Teachers College.

Rogers, C. R. (1961). *On becoming a person: A therapist's view of psychotherapy*. Boston: Houghton Mifflin.

Rogers, C. R. (1974). Remarks on the future of client-centred therapy. In D. N. Wexler, & L. N. Rice (Editors), *Innovations in client-centred therapy* (pp. 7-13). New York: Wiley.

Rogers, C. R. (1947). Some observations on the organisation of personality. *American Psychologist* , 2 (9) 358-368.

Rogers, C. R. (1939). *The clinical treatment of the problem child*. Boston: Houghton Mifflin.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology* , 21(2) 95-103.

Rogers, C. R. (1940). The processes of therapy. *Journal of Consulting Psychology* , 4 (5) 161-164.

Rogers, C. R. (1985). Toward a more human science of the person. *Journal of Humanistic Psychology* , 25 (4) 7-24.

Rogers, C. R., & Carson, C. W. (1930). Intelligence as a factor in camping activities. *Camping Magazine* , 3 (3) 8-11.

Rogers, C. R., & Dymond, R. F. (1954). *Psychotherapy and personality change: Co-ordinated research studies in the client-centred approach*. Chicago: University of Chicago Press.

Rogers, C. R., & Stevens, B. (1967). *Person to person: The problem of being human*. New York: Real People Press.

Rogers, C. R., & Wallen, J. R. (1946). *Counselling with returned servicemen*. New York: McGraw-Hill.

- Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, 6 412-415.
- Rowland, N., Godfrey, C., Bower, P., Mellor-Clark, J., Heywood, P., & Hardy, R. (2000). Counselling in primary care: A systematic review of the evidence. *British Journal of Guidance and Counselling*, 28 (2) 215-231.
- Sachse, R., & Elliott, R. (2002). Process-outcome research in client-centred and experiential therapies. In D Cain and J Seeman (Editors), *Humanistic psychotherapies: Handbook of research and practice* (pp. 83-115). Washington, DC: American Psychological Association.
- Safran, J. D., Muran, C., & Eubanks-Carter, C. (2010). Repairing alliance ruptures. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter11>.
- Salamone, J. D. (2002). Antidepressants and placebos: Conceptual problems and research strategies. *Prevention & Treatment*, 5 (1). No pagination specified. .
- Sanders (Editor), P. (2004). *The tribes of the person-centred nation: An introduction to the schools of therapy related to the person-centred approach*. Ross-on-Wye: PCCS Books.
- Schwartz, S. M., Trask, P. C., Shanmugham, K., & Oswald Townsend, C. (2004). Conducting psychological research in medical settings: Challenges, limitations and recommendations for effectiveness research. *Professional Psychology: Research and Practice*, 35 (5) 500-508.
- Schwebel, M., & Asch, M. J. (1948). Research possibilities in nondirective teaching. *Journal of Educational Psychology*, 39 (6) 359-369.
- Scriven, M. (1998). The meaning of bias. *Proceedings of the Stake symposium on educational evaluation*. Downloaded from www.eric.ed.gov on 18th October 2007 (pp. 37-48). Illinois: University of Illinois.
- Sechrest, L. (1994). Recipes for psychotherapy. *Journal of Consulting and Clinical Psychology*, 62 (5) 952-954.
- Seeman. (1949). A study of the process of nondirective therapy. *Journal of Consulting Psychology*, 13 (3) 157-168.

Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology*, 59 (6) 883-893.

Sheerer, E. T. (1949). An analysis of the relationship between acceptance and respect for self and acceptance of and respect for others in ten counselling cases. *Journal of Consulting Psychology*, 13 (3) 169-175.

Shick Tryon, G., & Winograd, G. (2010). Goal consensus and collaboration. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter7>.

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) 298-311.

Shirk, S. R., & Karver, M. (2010). Alliance in child and adolescent psychotherapy. In J. C. Norcross (Editor), *Evidence-based therapy relationships*. National Registry of Evidence-based Programmes and Practices (NREPP): Online Resource, available at: <http://174.140.153.165/Norcross.aspx#chapter3>.

Siev, J., & Chambless, D. L. (2007). Specificity of treatment effects: Cognitive therapy and relaxation for generalized anxiety and panic disorders. *Journal of Consulting and Clinical Psychology*, 513-522.

Silberschatz, G. (1994). Abuse and disabuse of the drug metaphor in psychotherapy research: Hold onto the baby as you throw out the bath. *Journal of Consulting and Clinical Psychology*, 62 (5) 949-951.

Simpson, J. A., Orinda, M. M., & Ickes, W. (2003). When accuracy hurt and when it helps: A test of the empathic accuracy model in marital interaction. *Journal of Personality and Social Psychology*, 85 (5) 881-893.

Smith, M. L., & Glass, G. V. (1977). Meta analysis of psychotherapy outcome studies . *American Psychologist*, 32 752-760.

- Snow, R. E., & Wiley (Editors), D. E. (1991). *Improving inquiry in social science: A volume in honour of Lee J Cronbach*. London: Lawrence Erlbaum Associates.
- South, S. C., Turkheimer, E., & Oltmanns, T. F. (2008). Personality disorder symptoms and marital functioning. *Journal of Consulting and Clinical Psychology* , 76 (5) 769-780.
- Spinhoven, P., Giez, A.-B. J., van Dyck, R., Kooiman, K., & Arntz, A. (2007). The therapeutic alliance in schema-focused therapy and transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology* , 75 (1) 104-115.
- Stiles, W. B. (2002). Assimilation of problematic experiences. In J. C. Norcross (Editor), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 357-365). New York: Oxford University Press.
- Stiles, W. B. (1994). Drugs, recipes, babies, bathwater and psychotherapy process-outcome relations. *Journal of Consulting and Clinical Psychology* , 62 (5) 955-959.
- Stiles, W. B. (1988). Psychotherapy process-outcome correlations may be misleading. *Psychotherapy* , 25 (1) 27-35.
- Stiles, W. B. (1996). When more of a good thing is better: Reply to Hayes et al. (1996). *Journal of Consulting and Clinical Psychology* , 64 (5) 915-918.
- Stiles, W. B., & Shapiro, D. A. (1994). Disabuse of the drug metaphor: Psychotherapy process-outcome correlations. *Journal of Consulting and Clinical Psychology* , 62 (5) 942-948.
- Stiles, W. B., Agnew-Davies, R., Hardy, G. E., Barkham, M., & Shapiro, D. A. (1998). Relations of the alliance with psychotherapy outcome: Findings in the second Sheffield psychotherapy project. *Journal of Consulting and Clinical Psychology* , 66 (5) 791-802.
- Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service settings. *Psychological Medicine* 36 , 555-566.
- Stiles, W. B., Shapiro, D. A., & Elliott, R. (1986). Are all psychotherapies equivalent? *American Psychologist* , 41 165-180.

Stiles, W., Barkham, M., Mellor-Clark, J., & Connell, J. (2007). Effectiveness of Cognitive-Behavioural, Person-Centred and Psychodynamic Therapies in UK Primary Care Routine Practice: Replication in a Larger Sample. *Psychological Medicine* , doi:10.1017/S0033291707001511.

Stiles, W., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Routine psychological treatment and the Dodo verdict: A rejoinder to Clark et al 2007. *Psychological Medicine* , 38 1-6.

Stock, D. (1949). An investigation into the interrelations between the self concept and feelings directed toward other persons and groups. *Journal of Consulting Psychology* , 13 (3) 176-180.

Strauss, J. L., Hayes, A. M., Johnson, S. L., Newman, C. F., Brown, G. K., Barber, J. P., et al. (2006). Early alliance, alliance ruptures and symptom change in a nonrandomised trial of cognitive therapy for avoidant and obsessive-compulsive personality disorders. *Journal of Consulting and Clinical Psychology* , 74 (2) 337-345.

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology* , 75 (6) 864-874.

Stulz, N., Thase, M. E., Klein, D. N., Manber, R., & Crits-Christoph, P. (2010). Differential effects of treatments for depression: A latent growth model reanalysis. *Journal of Consulting and Clinical Psychology* , 78 (3) 409-419.

Svartberg, M., Seltzer, M. H., Choi, K., & Stiles, T. C. (2001). Cognitive change before, during and after short-term dynamic and nondirective psychotherapies: A preliminary growth modelling study. *Psychotherapy Research* , 11 (2) 201-219.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioural therapy for depression. *Journal of Consulting and Clinical Psychology* , 67 (6) 894-904.

Tang, T. Z., De Rubeis, R. J., Hollon, S. D., Amsterdam, J., & Shelton, R. (2007). Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *Journal of Consulting and Clinical Psychology* , 75 (3) 404-408.

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions and sudden gains in cognitive-behavioural therapy for depression. *Journal of Consulting and Clinical Psychology*, 73 (1) 168-172.

Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive-behavioural therapy? *Journal of Consulting and Clinical Psychology*, 70 (2) 444-447.

Tarrier, N., Kinney, C., McCarthy, E., Humphreys, L., Wittkowski, A., & Morris, J. (2000). Two-year follow-up of cognitive-behavioural therapy and supportive counselling in the treatment of persistent symptoms in chronic schizophrenia. *Journal of Consulting and Clinical Psychology*, 68 (5) 917-922.

Teusch, L., Bohme, H., Finke, J., & Gastpar, M. (2001). Effects of client-centred psychotherapy for personality disorders alone and in combination with psychopharmacological treatment. *Psychotherapy and Psychomatics*, 70 (6) 328-336.

Thase, M. E. (2002). Antidepressant effects: The suit may be small, but the fabric is real. *Prevention & Treatment*, 5 (1). No pagination specified.

Thase, M. E., & Callan, J. A. (2006). The role of homework in cognitive behaviour therapy of depression. *Journal of Psychotherapy Intergration*, 16 (2) 162-177.

Thorne, B. (2003). *Carl Rogers (Second edition)*. London: Sage.

Truax, C. B., & Mitchell, K. M. (1971). Research on certain therapist interpersonal skills in relation to process and outcome. In A. E. (Editors), *Handbook of psychotherapy and behaviour change: An empirical analysis*. New York: Wiley.

Victora, C. G., Habicht, J.-P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomised trials. *American Journal of Public Health*, 94 400-405.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods and findings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73 (5) 914-923.

Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., et al. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care for patients with depression. I: Clinical effectiveness. *British Medical Journal* , 321 1383-1388.

Watson, J. C. (1996). The relationship between vivid description, emotional arousal and in-session resolution of problematic reactions. *Journal of Consulting and Clinical Psychology* , 64 (3) 459-464.

Watson, J. C., & Bedard, D. L. (2006). Clients' emotional processing in psychotherapy: A comparison between cognitive-behavioural and process-experiential psychotherapies. *Journal of Consulting and Clinical Psychology* , 74 (1) 152-159.

Watson, J. C., Gordon, L. B., Stermac, L., Kalogerakos, F., & Steckley, P. (2003). Comparing the effectiveness of process-experiential with cognitive-behavioural psychotherapy in the treatment of depression. *Journal of Consulting and Clinical Psychology* , 71 (4) 773-781.

Watson, N. (1984). The empirical status of Rogers' hypotheses of the necessary and sufficient conditions for effective psychotherapy. In R. F. Levant, & J. M. Shlien (Editors), *Client-centred therapy and the person-centred approach: New directions in theory, research and practice* (pp. 17-40). Westport, Connecticut: Praeger.

Weisz, J. R., Weersing, V. R., & Henggeler, S. W. (2005). Jousting with straw men: Comment on Westen, Novotny and Thompson-Brenner (2004). *Psychological Bulletin* , 131 418-426.

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic and generalised anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology* , 69 875-899.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2005). EBP ≠ EST: Reply to Crits-Christoph et al. (2005) and Weisz et al. (2005). *Psychological Bulletin* , 131 (3) 427-433.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings and reporting in controlled clinical trials. *Psychological Bulletin* , 130 (4) 631-663.

- Weston, T. (2008b). Clinical effectiveness of the person-centred approach: A preliminary study. *PCE 2008: Present strengths and future challenges: 8th world conference for person-centred and experiential psychotherapy and counselling* (p. 140). Norwich: University of East Anglia.
- Weston, T. (2008a). Clinical effectiveness of the person-centred approach: A preliminary study. *14th BACP Research Conference Research and Regulation - Towards an Evidence-Based Profession* (p. 60). Cardiff: British Association for Counselling and Psychotherapy.
- Weston, T. (2005). *The clinical effectiveness of the person-centred psychotherapies: A preliminary inquiry including literature review, CORE-OM questionnaires, client session recordings and client feedback*. Norwich: University of East Anglia.
- Weston, T. (2008c). The therapeutic relationship: A neglected variable in NHS England research and treatment. *PCE 2008: Present strengths and future challenges: 8th world conference for person-centred and experiential psychotherapy and counselling* (p. 139). Norwich: University of East Anglia.
- Weston, T., & Weston, S. (2008). The therapeutic relationship: A neglected variable in NHS England research and treatment. *Royal College of General Practitioners Annual National Primary Care Conference* (p. 64). Bournemouth: Royal College of General Practitioners.
- Wyatt, G. (2001). *Rogers' therapeutic conditions: Evolution, theory and practice: Volume 1 - Congruence*. Ross-on-Wye: PCCS Books.
- Wyatt, G., & Sanders, P. (2002). *Rogers' therapeutic conditions: Evolution, theory and practice: Volume 4 - Contact and perception*. Ross-on-Wye: PCCS Books.
- Zuroff, D. C., & Blatt, S. J. (2006). The Therapeutic Relationship in the Brief Treatment of Depression: Contributions to Clinical Improvement and Enhanced Adaptive Capacities. *Journal of Consulting and Clinical Psychology*, 74 (1) 130-140.

10. Appendices

Appendix 1: Research Ethics Committee submission

Doctor of Philosophy (PhD) in Counselling

– Research Proposal for consideration by Ethics Committee

Tony Weston

February 2006

Word count = 5,473 minus appendices etc. 2,899 = 2,574 words

This is a proposal for research in the University Counselling Service by the Centre for Counselling Studies with the intention of using this as a basis upon which to bid for research funding.

1. What is the idea, the issue, theme, focus or the question that you want to inquire into?

To inquire into the clinical effectiveness of the person centred psychotherapies using data from the University Counselling Service.

Briefly, the background to this is as follows:

Increasingly the providers and funders of psychological services are questioning the clinical effectiveness of therapies that are offered to clients. In the UK the National Institute for Clinical Excellence (NICE) have commissioned reports on the clinical effectiveness of different psychological therapies for treating a range of diagnoses (NICE 2004a-h). In general terms, NICE consider that the data they have examined have failed to prove the effectiveness of the person centred psychotherapies, although findings at this University (Weston 2005a) and those of others (Elliott, Greenberg and Lietaer 2004) suggest that the conclusions NICE have reached are erroneous (Weston 2005b). This research is designed to capture further clinical research data to contribute to this debate.

2. What is the aim of your inquiry?

The aim of this inquiry is to capture data on the clinical effectiveness of the person centred psychotherapies.

3. What is the utility of the inquiry you propose?

The utility of the inquiry we propose is:

1. To improve the clinical effectiveness of the therapy that individual clients receive.
2. To continue to contribute to the evidence base for the clinical effectiveness of the person centred psychotherapies (Elliott, Greenberg and Lietaer 2004) and to participate in the International Project on the Effectiveness of Psychotherapy and Psychotherapy Training (IPEPPT).

4. How does it fit with the existing system in which you practice?

One of the themes of the UEA Counselling Diploma and the British Association for Counselling & Psychotherapy (BACP) guidelines is for practitioners to be *reflective*. In some respects this research is creating a more formalised quantitative means of being reflective.

Through the use of professional supervision we review the work that we do and seek to be more effective.

The process that we propose is as follows:

1. Every client attending the University Counselling Service undertakes an Exploratory session with an experienced (post qualification) counsellor. As much as is practicable we will seek to make prospective clients aware that a research project is underway at the Counselling Service through the use of information provided at enquiry e.g. handout, e-mail, laminated cards etc. Whilst this is not achieving consent prior to the Exploratory session this is building awareness of this potential issue so that clients anticipate being informed about the research (see below).
2. At present when a client arrives for their Exploratory session the counselling receptionist asks the client to complete some standard paperwork (contact details etc.) before the client meets the counsellor doing the Exploratory. We propose that the research paperwork is given to the client by the counsellor (Research Information and Consent Form – Appendix 1). This will contain a brief explanation of the research, consent and demographics forms plus three standard psychological questionnaires. The counsellor judges whether the client is competent to consider giving their consent.
3. It is important that data is captured at this stage because we intend to use a ‘wait list control’ as part of our research methodology. However, we also need to give the client time

to consider if they wish to take part in the research. The proposed process is that the client completes the psychological questionnaires at the start of the Exploratory session and if they wish to take part in the research they bring the completed consent form to their first counselling session. It is made clear that data will not be used without consent and that the client may withhold consent after the psychological questionnaires have been completed at the Exploratory stage.

4. The consent paperwork makes it clear that the client will not be identified/identifiable in anything that is written about the research and that they need not consent to inclusion in the research and their therapy will continue as normal.

5. If the client is interested in participating in the research then they may take away the consent form for further consideration and then take this to their first therapy session. Counsellors may offer to retain the Consent and Demographics Forms on behalf of the client, with the client's notes, if this would be helpful to the client.

6. Clients may spontaneously offer their consent at the Exploratory stage and, provided that this is unprompted by the counsellor and that the counsellor judges the client competent to give their consent, the counsellor may accept the completed forms.

7. At the first therapy session the client may spontaneously offer the completed consent form, or, spontaneously ask to sign it. To avoid any sense of coercion the counsellor may inquire after the client's consent in the negative (e.g. 'You've decided not to participate in the research?') in a manner acceptant of this, in case this is the client's decision. Clients may delay giving their consent until the end of their sessions if they wish.

8. The counsellor may have a spare copy of the Consent Form to hand at the first session, in case the client has lost it, this must not be used in a 'coercive' manner.

9. If a period of six weeks or more has elapsed since the Exploratory session, at the first session the counsellor will ask the client to complete similar psychological questionnaires to the exploratory session and this is important for the ‘wait list control’.
10. At the start of the last session (usually the sixth session) the counsellor will ask the client to re-complete the three psychological questionnaires and complete a relationship questionnaire.
11. At any stage the counsellor may, if they wish, and especially in response to a direct question from the client, share the information from the psychological questionnaires with the client. As person centred therapists working without making diagnoses and offering unconditional positive regard for clients it is important to use non-judgemental language to describe test results and how these compare with the normative data. E.g. we wouldn’t say ‘you have severe depression’ but rather ‘your score was 47 and that compares with a group that the researchers who developed this test decided to term ‘severely depressed’’.
12. In terms of recruiting counsellors to take part in the research we (Tony Weston and Judy Moore) will meet with and talk to groups of counsellors about the research and answer any questions they may have. We will be clear that we are not attempting to coerce counsellors into consenting to take part in the research and that participation, or otherwise, will have no impact upon their counselling role e.g. in terms of pass/fail decisions about students or making changes to counsellors terms of employment.
13. Individual counsellors will each be mailed the paperwork about this research (Appendix 3) so that they may consider their involvement in a private space, if they wish. To participate they may mail back their completed consent and demographic forms to Tony Weston. Each counsellor will be given a code by him to put upon the psychological questionnaires that their clients complete, so that Tony Weston can match client with counsellor. No one else in the

University Counselling Service/Centre for Counselling Studies will see this information, to ensure counsellor confidentiality.

14. If they wish, prospective counsellor participants may receive a copy of this submission to the Research Ethics Committee so that they are fully informed, if they wish, and as a minimum they will receive a copy of the Research Information Sheet for Counsellors (Appendix 3), as above.

15. Counsellors will be able to participate at two levels, either in doing Exploratories (not diploma students) or as counsellors at client sessions and the Counsellor Consent form enables counsellors to be clear in communicating their involvement, if they choose to participate.

5. Do you want to undertake the research in order to change something, to illuminate, to improve, to satisfy curiosity, to add to knowledge, for sheer enjoyment or for some other reason?

We want to undertake the research in order to:

- Improve the effectiveness of clients' experience of therapy and to
- Add to knowledge about the evidence base for the person centred psychotherapies

6. Where will you be doing your fieldwork?

We will be doing fieldwork in the University Counselling Service.

7. How are you going to gain access to research participants?

We will gain access to research participants as the clients that come to the University Counselling Service. We will give prospective clients the opportunity to take part in the research if they so wish or to do counselling without participation in the research.

We will make it clear that:

- ‘If you decide not to participate this will not affect your counselling/therapy in any way.’
- They may opt out in future at any stage if they do decide to participate, as such consent is an ongoing process.

This research will be limited to the first one hundred prospective clients that come for an exploratory session.

As much as is practically possible we will let prospective clients know in advance that the University Counselling Service is engaged in research and that they maybe invited to participate e.g. through website, handouts in Counselling Service, standard text on e-mails etc (Appendix 4).

Access to counsellor participants is outlined above, section 4, points 12-15.

8. What are the ethical issues, in particular of consent, confidentiality, anonymity and the negotiation of accounts?

In examining the ethical issues of this research I have drawn upon the UEA MA Study Guide (UEA 2004), UEA Research Ethics Committee Pack (UEA 2003) and BACP Ethical Research Guidelines (BACP 2004). Furthermore I attended a meeting of the UEA Education

and Lifelong Learning Ethics Committee on Monday 5th December to discuss some related research (Weston 2005c).

In considering communicating ethical issues to prospective research subjects we have drawn upon the UEA Tape recording: Undertaking of Confidentiality, CORE Client Information Sheet and CORE Client Consent Form.

Drawing all of these sources together we plan to use the Client Research Information and Consent Form (Appendix 1) and Counsellor Research Information and Consent Form (Appendix 2) to address all of the ethical issues of consent, confidentiality, anonymity and negotiation of accounts.

9. What are the power dynamics within the research?

BACP Ethical Research Guidelines (BACP 2004, Sections 2, 3.1 (6), 3.2, 3.5 and 4.2) address the issue of power dynamics within research where the researcher is also the Counsellor; ‘a Client’s refusal to participate in research must not interfere with the Counselling or Psychotherapeutic relationship’. This concept is communicated in the Client and Therapist Research Information and Consent Forms (Appendices 1 and 2).

Some of the power issues may include:

- People giving consent when really they are not happy too (coercion)
- People being unable to give clear consent e.g. because of linguistic, intellectual, emotional or other reasons
- Consent changing the therapeutic relationship and getting in the way of therapy

We will seek to address these power issues by:

- Using an experienced (post qualification) counsellor at the exploratory session to judge whether the client is competent to have given consent

- Encouraging the client's counsellor to continue to monitor whether the client is competent to give consent during ongoing counselling and that this can be the subject of professional supervision as required
- Encouraging counsellors to monitor therapeutic work to determine whether the consent issue is affecting the work
- Encouraging supervisors to monitor therapeutic work to determine whether the consent issue is affecting the work of the counsellor

If counsellors have concerns about these power issues they may raise them with their clients as part of their therapeutic work together. Counsellors and supervisors can discuss power issues as part of their Counselling Supervisor. Monitoring the power dynamics within the research is an issue for ongoing professional (Counselling) and research (Academic) Supervision.

10. How are you going to collect data?

We will collect the data through completed psychological questionnaires and demographic information, limited to: age, gender, marital status, parental status, whether staff or student and ethnic status.

11. How many interviews, observations etc?

Client participants will be asked to complete three psychological questionnaires at their exploratory session plus:

- If there is a gap of more than six weeks between Exploratory and First Session clients will be invited to re-complete the three psychological questionnaires
- At the start of the last session clients will be invited to re-complete the three psychological questionnaires. Additionally clients will be invited to complete a Barrett-Lennard Relationship Inventory (BLRI) questionnaire at this stage

For each client this represents a minimum of 7 questionnaires. It is unlikely that any participant will be asked to complete more than 10 questionnaires.

We will limit this research to the first 100 clients attending for an Exploratory session.

In total we estimate the scale of this research to be 700-1,000 questionnaires.

12. Will you keep a personal journal, make field notes?

I may keep a personal journal or make field notes if this seems useful to me during the research.

13. What documents will you collect?

The documents we will collect from the research are completed consent forms, demographic information and psychological questionnaires, as described above.

14. What is the timetable of the research? Can you set this out in terms of the phases of inquiry i.e. setting up, data collecting, analysis, writing up.

This is the first time that we have done something like this and we plan to gradually implement this at a pace that we feel is appropriate, erring on the side of caution. We plan to use this to build experience of this type of research for subsequent research funding bids and to educate counselling students and staff. At this stage our estimate of the outline timetable of research is as follows:

February 2006 – Research Ethics Committee

February 2006 – Begin ongoing data capture. Ongoing review.

June 2006 – Anticipated end of data collection. Ongoing review.

October 2006 – February 2007 – Ongoing data analysis and compilation of findings.

Spring 2007 – Complete review of research findings and use this for possible research funding bids

15. Are there other people involved in your inquiry, for example a critical friend or co-researchers?

Other people involved in our inquiry include:

- Counselling clients
- Counselling staff
- Counselling students (Diploma, Masters)
- Counselling Supervisors

Given the power dynamics of researching the clinical effectiveness of counselling students and staff it is important that they are able to opt into this research and that they know that the research findings will not be used in a punitive way e.g. To make pass/fail decisions about students or to make decisions about the employment status of employees. We have prepared a Counsellor Consent Form (Appendix 2).

16. What resources will you need?

The resources we will need include:

- Psychological questionnaires
 - Beck Anxiety Inventory (BAI) x 200
 - Beck Depression Inventory II (BDI-II) x200
 - Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) x200
 - Barrett-Lennard Relationship Inventory (BLRI) x200
- Consent forms
 - Clients
 - Counsellors
- PC with SPSS software

17. What is the literature connected with your area of inquiry and on methodology that you intend to look at?

The following literature is relevant:

- Literature connected with research methodologies and psychological testing
- Literature connected with ‘client centred research’/‘outcome management’
- Literature connected with effectiveness more generally
- Literature connected with the theory, practice and underlying philosophy of the person centred psychotherapies

References:

BACP (2002) British Association for Counselling and Psychotherapy Ethical Framework for Good Practice in Counselling and Psychotherapy 2002

BACP (2004) British Association for Counselling and Psychotherapy Ethical Guidelines for Research in Counselling and Psychotherapy July 2004

http://www.bacp.co.uk/research/e_g.pdf

Beck, A T, Steer, R A and Brown, G K (1987) BDI-II Manual. San Antonio: The Psychological Corporation

Beck, A T and Steer, R A (1993) BAI Manual. San Antonio: The Psychological Corporation

Clark, D A and Beck, A T (2002) Clark-Beck Obsessive Compulsive Inventory Manual. San Antonio: The Psychological Corporation

CORE (2004) 'Clinical Outcomes in Routine Evaluation System User Manual', CORE IMS Ltd 47 Windsor Street, Rugby CV21 3NZ <http://www.coreims.co.uk/>

Elliott, R, Greenberg, L S and Lietaer, G (2004) Research on Experiential Psychotherapies. In M J Lambert (Ed) *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change* 5th Edition (pp 139-193) New York: John Wiley & Sons

Lambert, M J (2001) Psychotherapy outcome and quality improvement: Introduction to the special section on patient-focused research. *Journal of Consulting and Clinical Psychology*, 69, 147-49

Lambert, M J, Whipple, J L, Smart, D W, Vermeersch, D A, Nielsen, S L & Hawkins, E L (2001) 'The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced?' *Psychotherapy Research*, 11, 49-68

Lambert, M J, Whipple, J L, Vermeersch, D A, Smart, D W, Hawkins, E J, Nielsen, S L & Goates, M (2002) 'Enhancing psychotherapy outcomes via providing feedback on client progress: A replication' *Clinical Psychology and*

Lambert, M J and Ogles, B M (2004) The Efficacy and Effectiveness of Psychotherapy. In M J Lambert (Ed) *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change* 5th Edition (pp 139-193) New York: John Wiley & Sons

NICE (2004a) National Institute for Clinical Excellence <http://www.nice.org.uk/>

NICE (2004b) National Institute for Clinical Excellence
<http://www.nice.org.uk/page.aspx?o=guidelines.completed>

NICE (2004c) National Institute for Clinical Excellence CG9 Eating Disorders
<http://www.bps.org.uk/eatingdisorders/files/ED.pdf>

NICE (2004d) National Institute for Clinical Excellence CG16 Self Harm
<http://www.nice.org.uk/pdf/CG016NICEguideline.pdf>

NICE (2004e) National Institute for Clinical Excellence
<http://www.nice.org.uk/page.aspx?o=guidelines.inprogress>

NICE (2004f) National Institute for Clinical Excellence Anxiety (Generalised) Second Draft for Consultation http://www.nice.org.uk/pdf/Anxiety_NICE_guideline.pdf

NICE (2004g) National Institute for Clinical Excellence Depression Second Draft for Consultation http://www.nice.org.uk/pdf/Depression_NICEguideline_2ndconsultation.pdf

NICE (2004h) National Institute for Clinical Excellence Post Traumatic Stress Disorder (PTSD) First Draft for Consultation http://www.nice.org.uk/pdf/PTSD_1stcons_full.pdf

Purton, C (2004) 'Person-Centred Therapy: The Focusing-oriented approach' Hampshire: Palgrave Macmillan

Rogers, C R (1985) 'Toward a more human science of the person' *Journal of Humanistic Psychology* 25 (4): 7-24

Thorne, B (2003) 'Carl Rogers' London: Sage 2nd Edition

UEA (2003) University of East Anglia School of Education and Professional Development Research Ethics Committee Information and Discussion Pack UEA February 2003

UEA (2004) University of East Anglia School of Education and Lifelong Learning MA in Counselling Unit 2 Dissertation Study Guide September 2004

Weston, T (2005a) The Clinical Effectiveness of the Person Centred Psychotherapies. Unpublished MA dissertation, University of East Anglia.

Weston, T (2005b) Clinically Effective Counselling. Unpublished e-mail exchange with the National Institute for Clinical Excellence (NICE), 23 August 2005.

Weston, T (2005c) Doctor of Philosophy (PhD) in Counselling Research Proposal for consideration by Ethics Committee

APPENDIX A: 'Clinical Studies of Client Progress through Therapy' a research project that is considered in part-fulfilment for the award of Doctor (PhD) in Counselling at the University of East Anglia (UEA)



UNIVERSITY COUNSELLING SERVICE

CLIENT RESEARCH INFORMATION FORM

1. What is the research about?

This research looks at what changes in well-being happen through counselling and in the process of waiting for the first full counselling appointment. The research is anonymous: we do not use your name nor any information identifiable to you. You will not be identified in anything that is written about the research. The benefits of the research will be to improve the effectiveness of counselling.

2. How will this research affect me?

It is our aim that this research will not have any adverse affect upon you.

If you would like to take part you will be invited to complete 3 standard psychological questionnaires at your exploratory session plus at your first full and final counselling sessions. The questionnaires ask you about any symptoms of depression, anxiety and other forms of distress you may be experiencing. Usually it takes people about 10-15 minutes to fill in all three forms. Additionally at your second and fifth session you will be invited to complete a questionnaire about support you may have received before coming to counselling and your experience of counselling. These will take about 5 minutes each. You may if you wish place the questionnaires at your last session in an envelope so that your counsellor does not see your responses.

With the consent form we would like to collect some standard demographic and prescription medication information from you so that we know how applicable the findings are to the general population.

You are free to withdraw from this research at any point should you wish to do so, as is your right, and this will be respected without judgement. If you choose to withdraw from the research you are free to continue in counselling and this will not affect your counselling in any way, nor will it affect how long you wait to see a counsellor.



3. What do I do if I have any concerns about the research?

If you have any concerns about this research you can in the first instance discuss these with your counsellor, additionally you may contact Judy Moore, Director of the Counselling Service on 01603-592651 Judith.Moore@uea.ac.uk The research project is being conducted by Tony Weston on behalf of the Counselling Service as part of his PhD in Counselling 'Clinical Studies of Client Progress through Therapy'.

4. What else do I need to know?

There are some practical difficulties in conducting a research project like this across a large counselling service. For various technical reasons not all of the counsellors at the University Counselling Service are participating in this research project. It is possible that you may complete the questionnaires at your Exploratory session and then, for practical reasons totally unrelated to the research, you may be allocated to a counsellor who is not participating in the research, in which case you will not need to complete any further questionnaires. If this does happen we want you to know that the data we take from the questionnaires you completed at the Exploratory session is in itself important to this research and you haven't 'wasted' your time. However, if you should wish us to destroy this data we will do so.

5. What do I do if I want to take part?

If you are interested in taking part in this research, please complete the three questionnaires given to you with this form.

To give you time to consider taking part in this research you do not need to sign the consent and demographic form now, instead you may choose to do so at your first counselling session: However, for the purposes of the research, it is important that the questionnaires are completed now. We will not use your data without your signed consent.

6. What do I do if I don't want to take part?

If you do not wish to proceed with inclusion in the research you need do nothing further and your counselling will continue as normal.

**UNIVERSITY COUNSELLING SERVICE****CLIENT RESEARCH CONSENT FORM**

Please complete this form if you would like to take part in this research project.

My consent is informed and freely given:

1. I have read the research information sheet provided and I agree to the fair and lawful processing of personal information for the purposes of analysis and research in line with the Data Protection Act 1998.

2. I understand that the researchers using data collected will not have access to any personal data provided (e.g. name, address, date of birth) that makes the information identifiable to me and that I will not be identified in anything that is written or reported about the research.

Name (block capitals)

Signature

Date



UNIVERSITY COUNSELLING SERVICE

CLIENT DEMOGRAPHIC INFORMATION FORM

Please complete this form if you would like to take part in this research project.

My standard demographic information:

1. My age:

2. My gender:

3. My marital status:

4. My parental status:

(i.e. how many, if any, children you have)

5. I am a student / member of staff (delete as appropriate)

6. How I would describe my ethnic status:

My prescription medication information:

1. I am currently taking the following medication:

APPENDIX B: 'Clinical Studies of Client Progress through Therapy' a research project that is considered in part-fulfilment for the award of Doctor (PhD) in Counselling at the University of East Anglia (UEA)

COUNSELLOR RESEARCH INFORMATION FORM

What is the research about?

This research looks at what changes in client well-being happen through counselling. For clients and counsellors the research is anonymous, we do not use your name nor any information identifiable to you, you will not be identified in anything that is written about the research. The benefits of the research will be to improve the effectiveness of counselling.

How will this research affect me?

It is our aim that this research will not have any adverse affect upon you.

If you would like to take part you will be invited to complete a consent form and to give us some demographic information, this is to help us assess how the findings relate to the general population of counsellors.

We would like you to ask your clients to complete 3 standard psychological questionnaires at their Exploratory and final session. Usually it takes people about 10-15 minutes to fill in all three forms. Additionally, at their last session clients will be invited to complete a questionnaire about their perception of the relationship they have had with you, again this will take a few minutes.

We will analyse the effectiveness of different groups of counsellors (e.g. students, post qualification, experienced and so on) and this analysis will be at aggregate level so that individuals are not identifiable. We will not use this research to make pass/fail decisions about students or to make changes in the employment status of any counsellors.

You are free to withdraw from this research at any point should you wish to do so, as is your right, and this will be respected without judgement. If you choose to withdraw from the research you are free to continue as a counsellor and this will not affect your counselling role in any way.

What do I do if I have any concerns about the research?

If you have any concerns about this research you can in the first instance discuss these with your supervisor/tutor, additionally you may contact any of:

- Tony Weston, Researcher 0870-4051858 Tony.Weston5@BTinternet.com
- Judy Moore, University Counselling Service 01603-592651 Judith.Moore@uea.ac.uk
- Professor Nigel Norris, Chair of the Research Ethics Committee, School of Education & Learning, University of East Anglia, Norwich NR4 7TJ N.Norris@uea.ac.uk

What else do I need to know?

There are some practical difficulties in conducting a research project like this across a large counselling service. There are two circumstances where this research will differ slightly from the description given above:

1. If your client has waited for six weeks or more between their Exploratory session and their first counselling session we would like you to invite them to complete the three questionnaires again, so that we have a better understanding of where they are starting counselling from.
2. For various technical reasons not all of the counsellors at the University Counselling Service are participating in this research project. It is possible that you may conduct an Exploratory session and subsequently your client is allocated to a counsellor who is not participating in this research, in which case they will not need to complete any further questionnaires. If this does happen we want you to know that the data we take from the questionnaires completed at the Exploratory session is in itself important to this research and you haven't 'wasted' your time.

If you choose to participate you will be given instructions on the practicalities of administering and returning the psychological questionnaires.

What do I do if I want to take part?

If you are interested in taking part in this research, please complete the consent form given to you with this information.

To give you time to consider taking part in this research you do not need to sign this consent form now, instead you may consider giving your consent. We will not use data from sessions you have undertaken without your signed consent.

What do I do if I don't want to take part?

If you do not wish to proceed with inclusion in the research you need do nothing further and your counselling role will continue as normal, this will not affect your counselling role in any way.

'Clinical Studies of Client Progress through Therapy' a research project that is considered in part-fulfilment for the award of Doctor (PhD) in Counselling at the University of East Anglia (UEA)

COUNSELLOR RESEARCH CONSENT FORM

Please complete this form if you would like to take part in this research project.

My participation in this research is as follows (delete as appropriate):

I agree/do not agree to participate at Exploratory Sessions

I agree/do not agree to participate at Counselling Sessions

My consent is informed and freely given:

1. I have read the information sheet provided and I agree to the fair and lawful processing of personal information for the purposes of analysis and research in line with the Data Protection Act 1998.

2. I understand that the researchers using data collected will not have access to any personal data provided (e.g. name, address, date of birth) that makes the information identifiable to me and that I will not be identified in anything that is written or reported about the research.

Name (block capitals)

Signature Date

'Clinical Studies of Client Progress through Therapy' a research project that is considered in part-fulfilment for the award of

Doctor (PhD) in Counselling at the University of East Anglia (UEA)

COUNSELLOR DEMOGRAPHIC INFORMATION FORM

My demographic information:

1. My age:

2. My gender:

3. My marital status:

4. My parental status:

(How many, if any, children you have)

5. I am a diploma student / other (delete as appropriate)

6. How I would describe my ethnic status:

7. My experience –

I am a diploma student

I took the diploma last year

OR number of years since I took the diploma....

(complete as appropriate)

APPENDIX C - Research Checklist

Client Number:

Please indicate date for each item completed:

Exploratory Session

1. Clinical Outcomes Routine Evaluation – Outcome Measure (CORE-OM).....
2. Beck Anxiety Inventory (BAI).....
3. Beck Depression Inventory (BDI-II).....

Client Consent Form.....

Start of Final Session (usually Sixth Session)

1. Clinical Outcomes Routine Evaluation – Outcome Measure (CORE-OM).....
2. Beck Anxiety Inventory (BAI).....
3. Beck Depression Inventory (BDI-II).....
4. Barrett-Lennard Relationship Inventory (BLRI).....

Appendix D: Outline of proposed text for promoting awareness of research at University Counselling Service

Words similar to these to appear on (e.g.):

- Website
- Handout to people making exploratory appointments
- Laminated sheets in Counselling Service Receptions
- E-mails sent in response to prospective clients e-mails
- If appropriate, to use with telephone enquiries

The University Counselling Service is conducting a research project that we would like to make you aware of in case you are asked to participate.

This research has been approved by the University Research Ethics Committee.

The research is designed to look at what changes in well-being occur during counselling. The research is anonymous, we do not use your name nor any information identifiable to you, you will not be identified in anything that is written about the research. The benefits of the research will be to improve the effectiveness of counselling.

If you are invited to take part in the research you will be given an information sheet to tell you more about the research so that you can make an informed decision about whether to take part. If you choose to take part in the research you will be invited to complete 3 standard psychological questionnaires at your Exploratory and final session. Usually it takes people about 10-15 minutes to fill in all three forms. Additionally at your last session you will be invited to complete a questionnaire about the relationship you have had with your counsellor, again this will take a few minutes.

You are free to withdraw from this research at any point should you wish to do so, as is your right, and this will be respected without judgement. If you choose to withdraw from the research you are free to continue in therapy and this will not affect your therapy in any way.

Your decision to participate, or not, will not affect your counselling in any way, nor will it affect, how long you wait to see a counsellor.

Appendix 2: Characteristics of sample upon which coefficient alphas were estimated.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication		Number of Sessions			Number of Therapists ^c
		Mean	SD	Range	% female	% single	% now married	% no children	% British	% none at start ^a	% none at end ^b	Mean	SD	Range	
Overall sample	321	31.2	12.0	12-80	66	62	26	67	88	84	88	7.1	10.2	0-102	27+
BDI-II Sample	299	30.9	11.6	12-63	67	63	25	67	88	83	87	6.5	8.4	0-63	27+
BAI Sample	301	30.9	11.9	12-80	67	63	25	67	88	83	87	6.5	8.4	0-63	27+
CORE-OM Sample	251	28.5	10.4	15-63	67	75	18	78	87	85	87	5.9	6.2	0-37	27+
PBQ Sample	53	39.6	10.6	14-59	62	19	53	30	94	76	93	14.4	19.6	1-102	1
BLRI Sample	118	34.4	11.3	19-63	67	49	31	56	86	80	88	12.0	14.1	1-102	19

Note: ^a Number of clients for whom it was not known whether they were taking relevant medication at the start, for each row in order 3, 3, 3, 3, 0 & 2 respectively.

^b Number of clients for whom it was not known whether they were taking relevant medication at the end, for each row in order 26, 25, 26, 26, 0 & 3 respectively.

^c Number of clients for whom it was not known who the allocated therapist was, for each row in descending order 48, 47, 48, 48, 0 & 0 respectively. In addition there were also 38 clients who were not allocated to a therapist because they did not attend a first session at the UCS.

Appendix 3: Participant flow and demographic characteristics for each sample analysed

Naturalistic research has been criticised e.g. Clark, et al. (2007). This research sought to address such criticisms and one such way to address these was to report the research in line with the TREND Guidelines (Des Jarlais, et al., 2004), the naturalistic equivalent to the CONSORT Guidelines. One of the requirements of the TREND Guidelines, described above, was to make clear what the participant flow was and the demographic characteristics for each sample analysed and this section was designed to comply with the TREND reporting requirements.

In this section, for each of the main outcome measures, there is a participant flow diagram and a summary of demographics for each sample reported upon in section ‘6. Results’, but not for the more ‘exploratory’ section ‘7. Further results’, apart from the wait-list control UCS clients, see below.

The main samples analysed were as follows:

- Depression (BDI-II)
- Anxiety (BAI)
- Distress (CORE-OM)
- Clients in wait-control analysis (depression – BDI-II, anxiety – BAI and distress – CORE-OM)
- Clients in hypothesis testing (outcomes A1 to A3 and prediction B1 to B3)

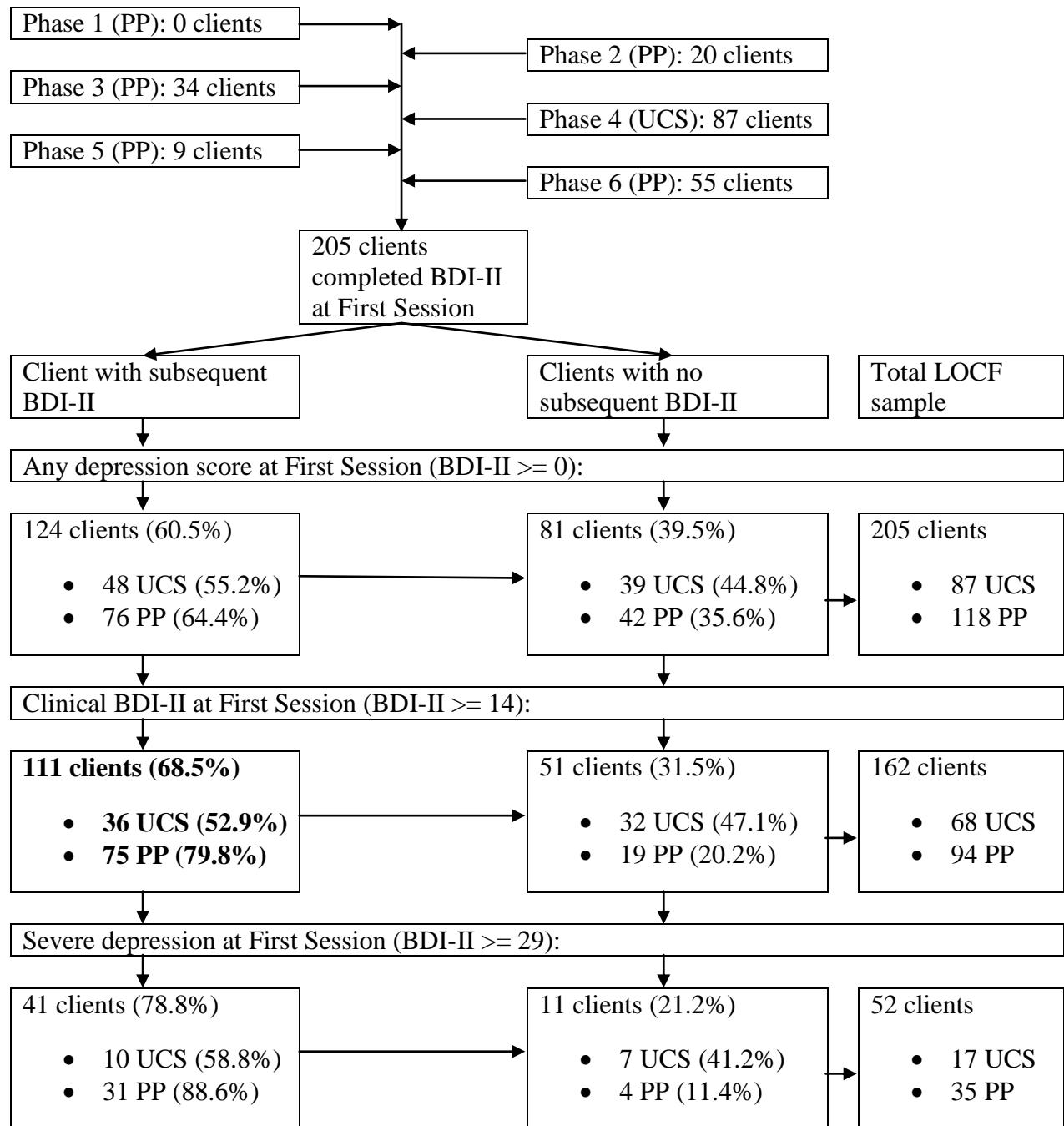
Each of these samples is represented in a participant flow diagram and demographic information table. Table 49 shows the demographics for the overall sample plus subsets of the UCS and PP samples. For depression outcomes (BDI-II) participant flow is shown (Figure 43) and demographic characteristics for each sample analysed (Table 50).

Table 49: Characteristics of overall sample plus subsets of University Counselling Service and Private Practice samples.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication		Number of Sessions			Number of Therapists ^c
		Mean	SD	Range		% female	% now single			% no children	% British	% none at start ^a	% none at end ^b	Mean	SD
Overall sample	321	31.2	12.0	12-80	66	62	26	67	88	84	88	7.1	10.2	0-102	27+
PP Sample	137	38.4	11.7	12-80	61	26	52	35	93	84	91	9.2	13.4	1-102	1
UCS Sample	184	25.9	9.1	18-63	69	89	7	90	85	83	85	5.5	6.4	0-37	27+
- Exploratory only	38	24.0	8.4	18-63	79	90	8	90	95	87	87	0	-	-	-
- Enter therapy	146	26.3	9.2	18-63	66	89	7	90	82	82	84	7.0	6.4	0-37	27+
- No 1 st session paperwork	59	25.6	8.1	18-60	68	90	5	90	88	75	75	5.9	4.7	0-21	5+
- Enter therapy 'in research'	87	26.8	9.9	19-63	66	89	8	91	78	87	91	7.7	7.4	1-37	26
- No last session paperwork	40	25.9	10.6	19-61	63	93	3	93	80	93	93	7.2	7.9	1-31	19
- Do last session paperwork	47	27.6	9.3	19-63	68	85	13	89	77	83	89	8.1	6.9	2-37	18
- Last session paperwork + proper ending ^d	36	28.1	10.1	19-63	69	89	6	89	75	86	92	8.3	7.1	2-37	16

Note: ^a Numbers of clients for whom it was not known whether any relevant medication was taken at the start (in descending order of sample rows) were: 3, 0, 3, 0, 3, 0, 3, 2, 1 and 1 respectively. ^b Numbers of clients for whom it was not known whether any relevant medication was taken at the end (in descending order of sample rows) were: 26, 0, 26, 5, 21, 15, 6, 3, 3 and 1 respectively. ^c There were 38 clients at UCS who did not have a first session and were not allocated to a therapist, additionally numbers of clients for whom it was not known who the therapist was (in descending order of sample rows): 49, 0, 49, 38, 21, 49, 0, 0, 0 and 0. ^d 'Proper ending' defined as that where therapist had marked client record card as 'mutually agreed and satisfactory ending'.

Figure 43: Origin of clients (Phase of Protocol) and questionnaire completion for clients completing BDI-II at First Session.



Note: Percentages refer to percentage of Last Observation Carried Forward (LOCF) sample that did or did not complete a subsequent BDI-II and add across rows to equal 100% e.g. Of 52 clients with severe depression at First Session 41 clients, 78.8%, completed a subsequent BDI-II and 11 clients, 21.2%, did not. PP = Author's Private Practice clients. UCS = University Counselling Service clients. **Bold** indicates clients in the hypotheses testing samples.

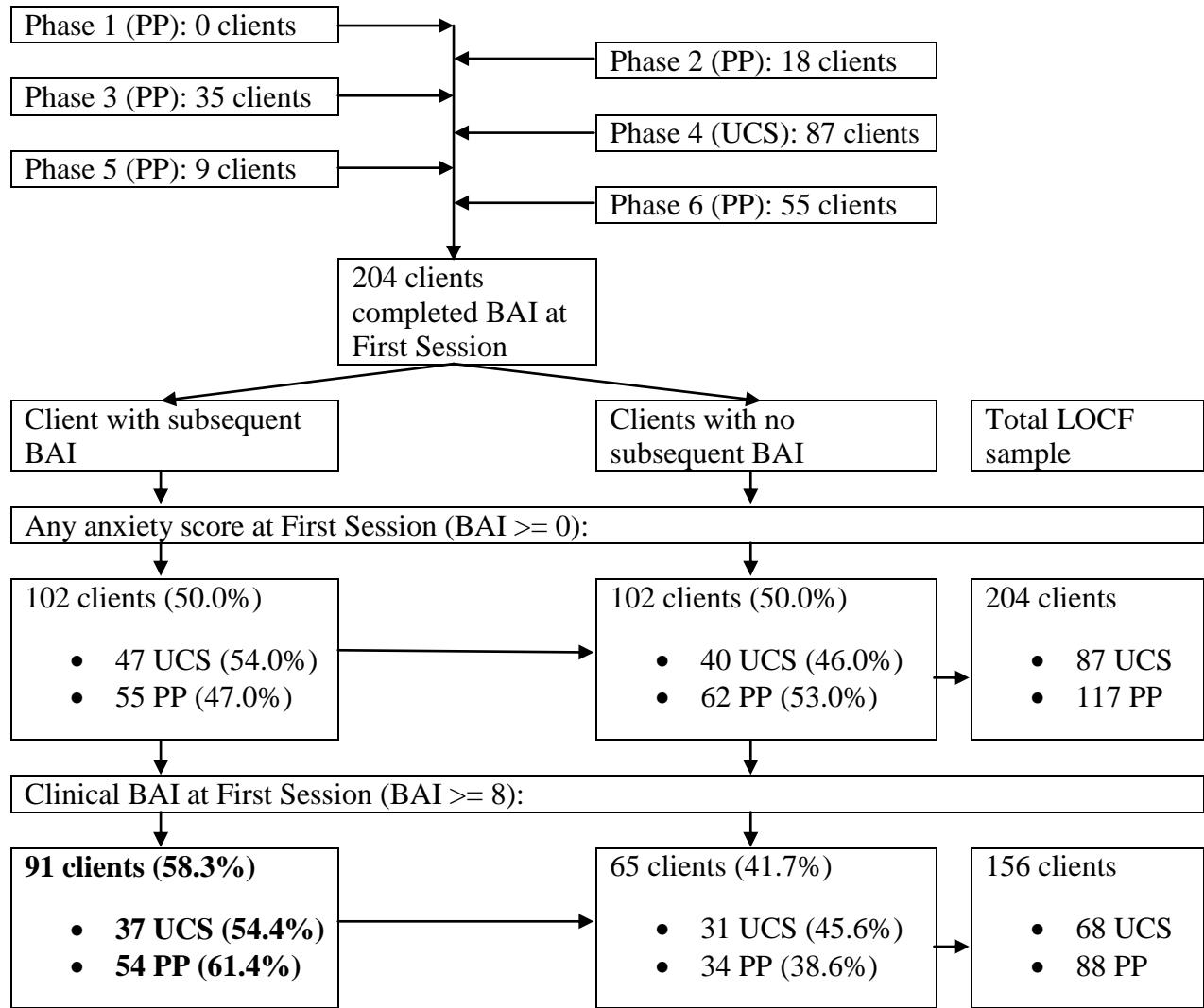
Table 50: Characteristics of sample for clients with depression (BDI-II) outcomes.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication		Number of Sessions ^a			Number of Therapists	
		Mean	SD	Range		% female	% single			% no children	% British	% none at start ^b	% none at end ^c	Mean	SD	Range
Overall sample	321	31.1	12.0	12-80	66	62	26		88	84	88	7.1	10.2	0-102	27 ^d	
BDI >=0 All _{LOCF}	205	33.6	12.0	12-63	64	51	34		57	87	84	90	7.8	9.4	1-63	26
- UCS	87	26.8	9.9	19-63	66	89	8		91	78	87	91	7.7	7.4	1-37	26
- PP	118	38.5	10.9	12-60	64	23	53		32	93	81	90	8.0	10.7	1-63	1
BDI>=0 All _{subseq}	124	35.5	11.0	19-63	66	43	38		50	86	79	90	8.2	7.2	2-48	18
- UCS	48	27.4	9.2	19-63	69	85	13		90	75	83	90	8.2	6.8	2-37	18
- PP	76	40.5	8.8	19-60	65	16	54		25	93	76	90	8.2	7.5	2-48	1
BDI>=14 All _{LOCF}	162	33.8	11.5	14-60	66	49	33		55	87	82	89	8.7	10.0	1-63	25
- UCS	68	26.5	8.9	19-53	69	88	7		91	79	85	88	8.5	7.9	1-37	25
- PP	94	39.1	10.2	14-60	64	20	52		29	93	79	89	8.8	11.3	1-63	1
BDI>=14 All_{subseq}	111	36.3	10.5	19-60	69	39	40		47	87	78	88	8.5	7.5	2-48	16
- UCS	36	27.7	8.3	19-53	75	86	11		92	75	81	86	9.1	7.6	2-37	16
- PP	75	40.4	8.8	19-60	65	16	53		25	93	76	89	8.2	7.5	2-48	1
BDI>=29 All _{LOCF}	52	35.1	11.6	14-60	73	42	33		48	90	73	87	8.9	10.2	1-58	10
- UCS	17	25.2	5.4	19-35	82	94	6		94	82	82	82	8.7	7.4	1-26	10
- PP	35	40.0	10.7	14-60	69	17	46		26	94	69	89	9.0	11.4	1-58	1
BDI>=29 All_{subseq}	41	38.6	10.3	19-60	68	32	37		37	88	68	85	8.8	7.9	2-43	6
- UCS	10	27.5	5.5	20-35	70	90	10		90	70	80	80	10.1	7.7	3-26	5
- PP	31	42.2	8.8	19-60	68	13	45		19	94	65	87	8.4	8.1	2-43	1

Note: ^a ‘Number of sessions’ was overall number of sessions for ‘overall sample’ and ‘LOCF’ samples. For samples where there was a subsequent BDI-II measurement ‘number of sessions’ was number of sessions to final BDI measurement (measurement occurred at start of session). ^b Numbers of clients for whom it was not known whether any relevant medication was taken at the start (in descending order of sample rows) were: 3, 3, 3, 0, 1, 1, 0, 3, 3, 0, 1, 1, 0, 0, 0, 0, 0, 0, and 0 respectively. ^c Numbers of clients for whom it was not known whether any relevant medication was taken at the end (in descending order of sample rows) were: 26, 6, 6, 0, 3, 3, 0, 6, 6, 0, 3, 3, 0, 3, 3, 0, 2, 2, and 0 respectively. ^d Within the ‘overall sample’ were 49 clients for whom it was not known who the therapist was, plus 38 clients who were not allocated to a therapist). **Bold** indicates clients in the hypotheses testing samples.

For anxiety outcomes (BAI) participant flow is shown (Figure 44) and demographic characteristics for each sample analysed (Table 51).

Figure 44: Origin of clients (Phase of Protocol) and questionnaire completion for clients completing BAI at First Session.



Note: Percentages refer to percentage of Last Observation Carried Forward (LOCF) sample that did or did not complete a subsequent BAI and add across rows to equal 100% e.g. Of 156 clients with clinical anxiety at First Session (BAI ≥ 8), 91 clients, 58.3%, completed a subsequent BAI and 65 clients, 41.7%, did not. PP = Author's Private Practice clients. UCS = University Counselling Service clients. **Bold** indicates clients in the hypotheses testing samples.

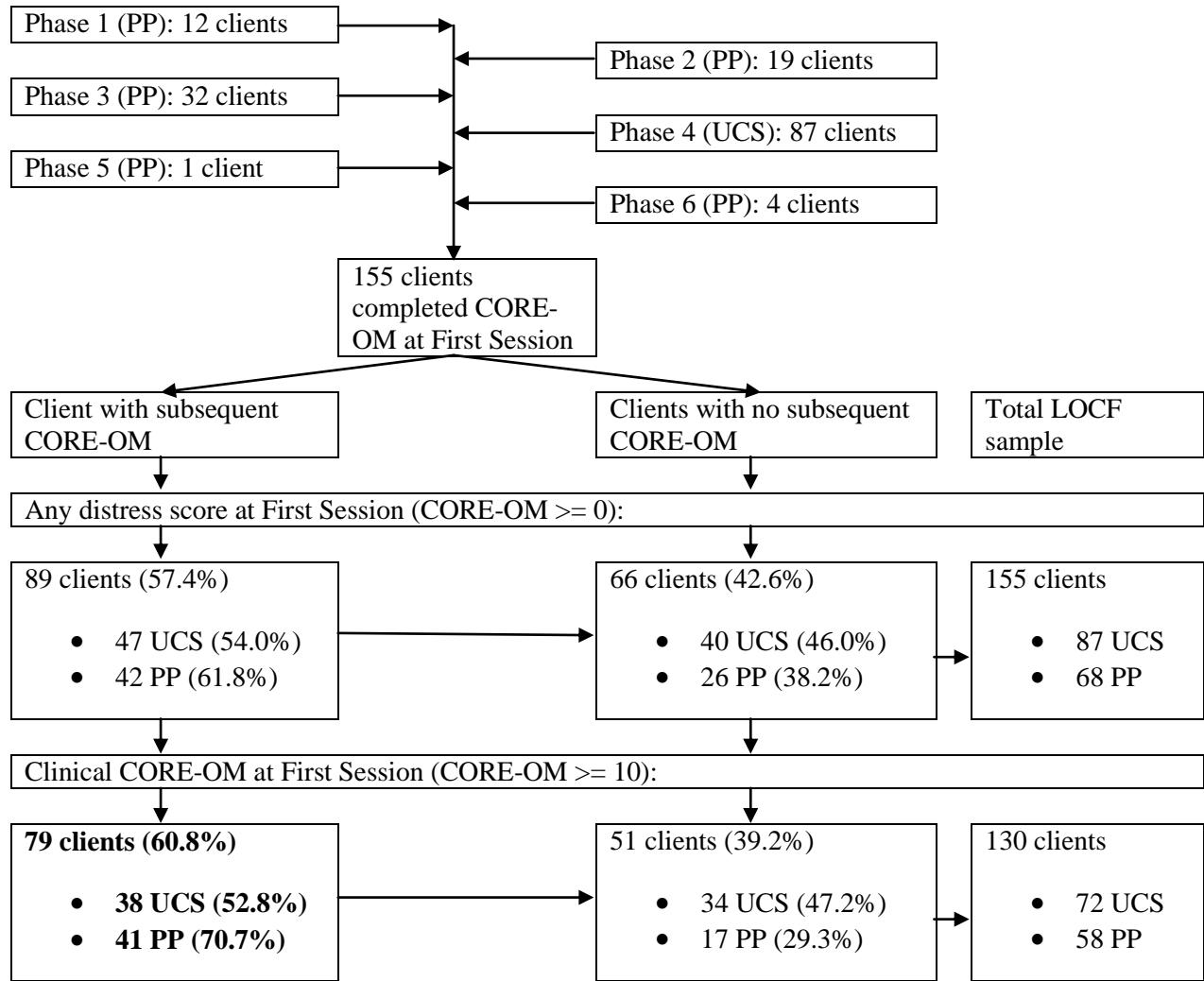
Table 51: Characteristics of sample for clients with anxiety (BAI) outcomes.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication		Number of Sessions ^a			Number of Therapists	
		Mean	SD	Range		% female	% single			% no children	% British	% none at start ^b	% none at end ^c	Mean	SD	Range
Overall sample	321	31.1	12.0	12-80	66	62	26		67	88	84	88	7.1	10.2	0-102	27 ^d
BAI >=0 All _{LOCF}	204	33.8	12.3	12-80	65	51	34		56	87	84	90	7.8	9.4	1-63	26
- UCS	87	26.8	9.9	19-63	66	89	8		91	78	87	91	7.7	7.4	1-37	26
- PP	117	38.9	11.4	12-80	64	22	53		31	93	82	90	7.9	10.7	1-63	1
BAI>=0 All _{subseq}	102	34.7	11.9	19-80	65	48	34		55	85	79	89	7.9	5.8	2-37	17
- UCS	47	27.6	9.3	19-63	68	85	13		89	77	83	89	8.1	6.9	2-37	18
- PP	55	40.8	10.4	20-80	62	16	53		26	93	77	89	7.8	4.6	2-30	1
BAI>=8 All _{LOCF}	156	33.7	12.0	14-80	67	50	32		56	88	81	89	7.9	8.2	1-46	24
- UCS	68	26.9	9.0	19-53	65	88	7		91	82	85	90	8.5	7.8	1-37	24
- PP	88	39.1	11.2	14-80	68	21	51		28	92	77	89	8.5	11.1	1-63	1
BAI>=8 All_{subseq}	91	35.4	11.5	19-80	64	45	35		53	88	78	90	8.1	6.0	2-37	16
- UCS	37	27.7	8.3	19-53	68	87	11		92	81	81	89	8.7	7.5	3-37	16
- PP	54	40.7	10.4	20-80	61	17	52		26	93	76	91	7.7	4.7	2-30	1

Note: ^a ‘Number of sessions’ was overall number of sessions for ‘overall sample’ and ‘LOCF’ samples. For samples where there was a subsequent BAI measurement ‘number of sessions’ was number of sessions to final BAI measurement (measurement occurred at start of session). ^b Numbers of clients for whom it was not known whether any relevant medication was taken at the start (in descending order of sample rows) were: 3, 3, 3, 0, 1, 1, 0, 3, 3, 0, 1, 1 and 0 respectively. ^c Numbers of clients for whom it was not known whether any relevant medication was taken at the end (in descending order of sample rows) were: 26, 6, 6, 0, 3, 3, 6, 6, 6, 0, 3, 3 and 0 respectively. ^d Within the ‘overall sample’ were 48 clients for whom it was not known who the therapist was, plus 38 clients who were not allocated to a therapist. **Bold** indicates clients in the hypotheses testing samples.

For distress outcomes (CORE-OM) participant flow is shown (Figure 45) and demographic characteristics for each sample analysed (Table 52).

Figure 45: Origin of clients (Phase of Protocol) and questionnaire completion for clients completing CORE-OM at First Session.



Note: Percentages refer to percentage of Last Observation Carried Forward (LOCF) sample that did or did not complete a subsequent CORE-OM and add across rows to equal 100% e.g. Of 130 clients with clinical distress at First Session (CORE-OM ≥ 10), 79 clients, 60.8%, completed a subsequent CORE-OM and 51 clients, 39.2%, did not. PP = Author's Private Practice. UCS = University Counselling Service. **Bold** indicates clients in the hypotheses testing samples.

Table 52: Characteristics of sample for clients with distress (CORE-OM) outcomes.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication		Number of Sessions ^a			Number of Therapists		
		Mean	SD	Range		% female	% single			% no children	% British	% none at start ^b	% none at end ^c	Mean	SD	Range	
Overall sample	321	31.1	12.0	12-80	66	62	26		88	67	84	88	7.1	10.2	0-102	27 ^d	
CORE>=0 All _{LOCF}	155	30.6	11.1	15-63	64	67	25		84	70	88	91	7.3	6.5	1-37	26	
- UCS	87	26.8	9.9	19-63	66	89	8		91	78	87	91	7.7	7.4	1-37	26	
- PP	68	35.4	10.7	15-58	62	38	46		91	44	91	88	91	6.8	5.4	1-32	1
CORE>=0 All _{subseq}	89	32.0	11.0	19-63	63	62	28		85	66	87	90	8.4	6.2	2-37	18	
- UCS	47	27.6	9.3	19-63	68	85	13		89	77	83	89	8.1	6.9	2-37	18	
- PP	42	37.0	10.7	19-58	57	36	45		95	41	91	91	8.6	5.4	2-32	1	
CORE>=10 All _{LOCF}	130	30.8	10.9	15-58	66	65	24		85	70	86	90	7.9	6.9	1-37	25	
- UCS	72	26.6	8.9	19-53	68	89	7		92	81	85	89	8.5	7.8	1-37	25	
- PP	58	35.9	11.0	15-58	64	36	45		91	43	88	91	7.2	5.5	1-32	1	
CORE>=10 All_{subseq}	79	32.5	10.7	19-58	63	60	29		89	65	85	89	8.8	6.4	2-37	16	
- UCS	38	27.4	8.2	19-53	71	87	11		92	82	79	87	8.9	7.5	2-37	16	
- PP	41	37.3	10.7	19-58	56	34	46		95	39	90	90	8.8	5.4	2-32	1	

Note: ^a 'Number of sessions' was overall number of sessions for 'overall sample' and 'LOCF' samples. For samples where there was a subsequent CORE-OM measurement 'number of sessions' was number of sessions to final CORE-OM measurement (measurement occurred at start of session). ^b Numbers of clients for whom it was not known whether any relevant medication was taken at the start (in descending order of sample rows) were: 3, 3, 1, 0, 1, 1, 0, 3, 3, 0, 1, 1 and 0 respectively. ^c Numbers of clients for whom it was not known whether any relevant medication was taken at the end (in descending order of sample rows) were: 26, 6, 3, 0, 3, 3, 0, 6, 6, 0, 3, 3 and 0 respectively. ^d Within the 'overall sample' were 48 clients for whom it was not known who the therapist was, plus 38 clients who were not allocated to a therapist. **Bold** indicates clients in the hypotheses testing samples.

The protocol for clients at the UCS was to complete outcome measures at Exploratory, First and Last Session. The intention was that each client would then act as their own wait control; outcome measure at start of wait compared with outcome measure at end of wait, although each client had a session with a qualified therapist as their exploratory so had received some counselling input. Changes during wait could then be compared with changes during 'active treatment'; outcome measure at first session compared with last session. The demographic characteristics of the clients in the 'wait time analysis', comparing change during wait with treatment are shown (Table 53).

Table 53: Characteristics of sample for clients in wait time analysis: Clients with clinical scores at first session, any score at exploratory and last session plus numbers of days in wait and treatment conditions.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication		Number of Sessions ^a			Number of Therapists	
		Mean	SD	Range		% female	% single			% no children	% British	% none at start ^b	% none at end ^c	Mean	SD	Range
Overall sample	321	31.1	12.0	12-80	66	62	26		88	67	84	88	7.1	10.2	0-102	27+ ^d
BDI-II	36	27.7	8.3	19-53	75	86	11		75	92	81	86	9.1	7.6	2-37	16
BAI	36 ^e	27.7	8.3	19-53	69	86	11		81	92	81	89	8.8	7.6	3-37	15
CORE-OM	37 ^e	27.5	8.3	19-53	73	87	11		81	92	78	87	9.0	7.5	2-37	15

Note: ^a ‘Number of sessions’ was overall number of sessions for ‘overall sample’ and ‘LOCF’ samples. For samples where there was a subsequent measurement ‘number of sessions’ was number of sessions to final measurement (measurement occurred at start of session). ^b Numbers of clients for whom it was not known whether any relevant medication was taken at the start (in descending order of sample rows) were: 3, 1, 1 and 1 respectively. ^c Numbers of clients for whom it was not known whether any relevant medication was taken at the end (in descending order of sample rows) were: 26, 3, 3 and 3 respectively. ^d Within the ‘overall sample’ were 48 clients for whom it was not known who the therapist was, plus 38 clients who were not allocated to a therapist. ^e One client (case 303) with clinical BAI and CORE-OM scores, but non-clinical BDI-II at first session, missing date of first session and cannot determine wait or treatment length, although overall process was 68 days.

Participant flow for the outcome hypotheses samples are highlighted (**bold**) in each of the outcome measure flow charts (Figure 43, Figure 44 and figure 45) and the demographic characteristics of the clients who formed the outcome hypotheses testing samples are highlighted (**bold**) in each of the outcome measure demographic tables (Table 51, Table 52 and Table 53) and are also summarised together in one table (Table 54). In addition to the outcome hypotheses are the process-outcome hypotheses and participant flow for the samples to examine putative prediction of depression is shown (Figure 46), for anxiety (Figure 47) and distress (Figure 48). It is important to note a key difference between the outcome and process-outcome samples for the analysis. Analysis of outcomes for clients with clinical scores, by definition, *excludes* clients with sub-clinical scores, whereas the predictor samples *include* clients with sub-clinical scores. This is because a test of person-centred theory (Rogers, 1957, 1959) must include clients who are 'well' (congruent clients in addition to incongruent clients) within the sample (Watson, 1984, p. 37)

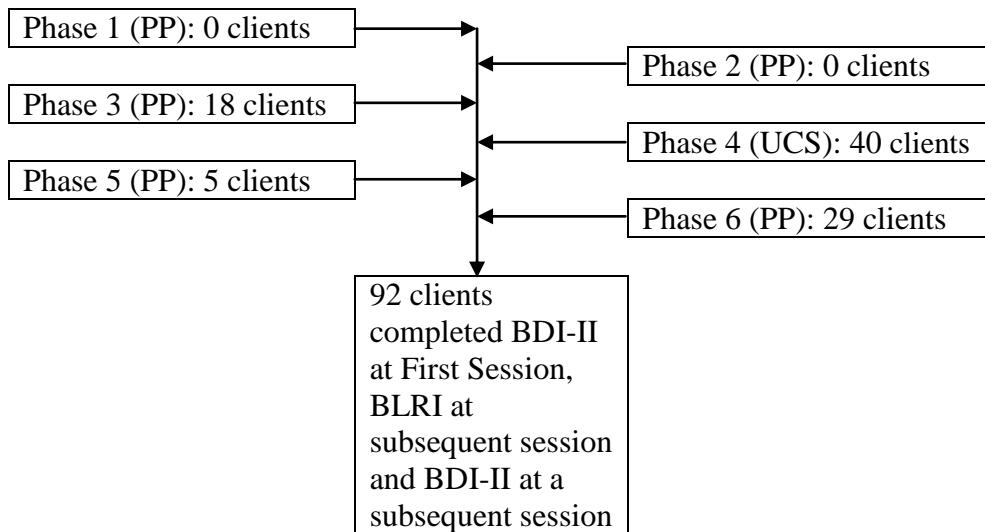
Table 54: Characteristics of clients in hypotheses testing.

Sample	N	Age (years)			Gender	Marital Status		Parental Status	Nationality	Relevant medication ^a		Number of Sessions ^b			Number of Therapists
		Mean	SD	Range		% female	% single			% no children	% British	% none at start	% none at end	Mean	SD
Depression	111	36.3	10.5	19-60	69	39	40	47	87	78	88	8.6	7.6	2-48	16
Anxiety	91	35.4	11.5	19-80	64	45	35	53	88	78	90	8.1	6.0	2-37	16
Distress	79	32.5	10.7	19-58	63	60	29	65	89	85	89	8.8	6.4	2-37	16
Dep Prediction	92	35.4	10.9	19-63	70	44	35	52	83	76	87	9.2	8.0	2-48	16 ^c
- Std Resids < 3	90	35.2	10.9	19-63	69	44	37	53	82	77	88	8.8	7.3	2-48	16 ^c
Anx Prediction	75	34.0	11.2	19-63	69	51	29	60	81	75	87	8.8	6.3	3-37	16 ^c
- Std Resids < 3	73	34.1	11.3	19-63	69	52	27	59	82	75	86	8.8	6.4	3-37	16 ^c
Distress Prediction	54	31.3	10.8	19-63	69	65	22	72	80	82	85	8.6	6.5	3-37	16 ^c
- Std Resids < 3	as above														

Note: ^a Relevant medication at start or end, this percentage includes only those clients confirmed as 'none', there were some further clients who were are 'not known'

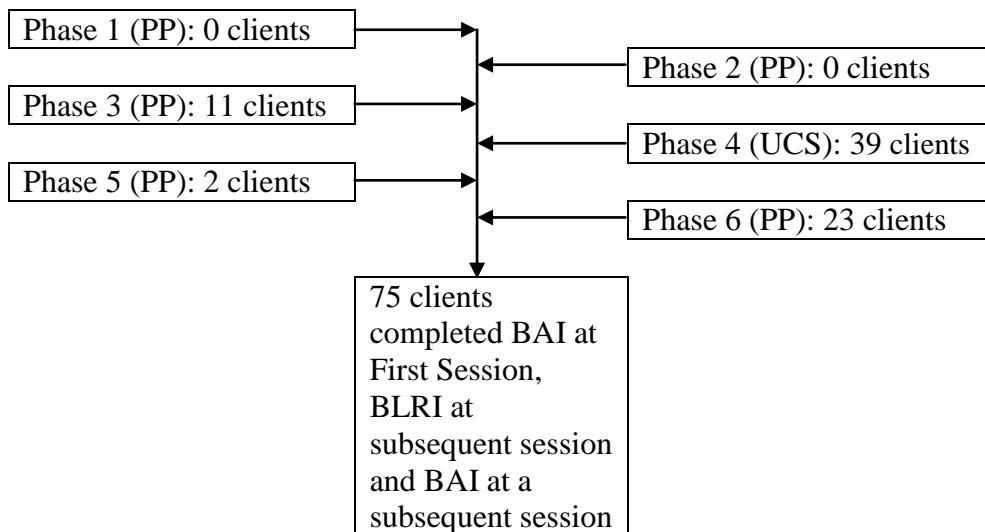
whether they were taking relevant medication. ^b Number of sessions to final measurement on this questionnaire. Some clients had further sessions. ^c In the predictor models, one therapist (the author) saw the following percentages of clients in the mediation models (in row descending order): 64.1%, 63.3%, 56.0%, 56.2% and 38.9% respectively.

Figure 46: Origin of clients (Phase of Protocol) in depression predictor sample: Clients who completed BDI-II at First Session, BLRI at a subsequent session and BDI-II at a subsequent session.



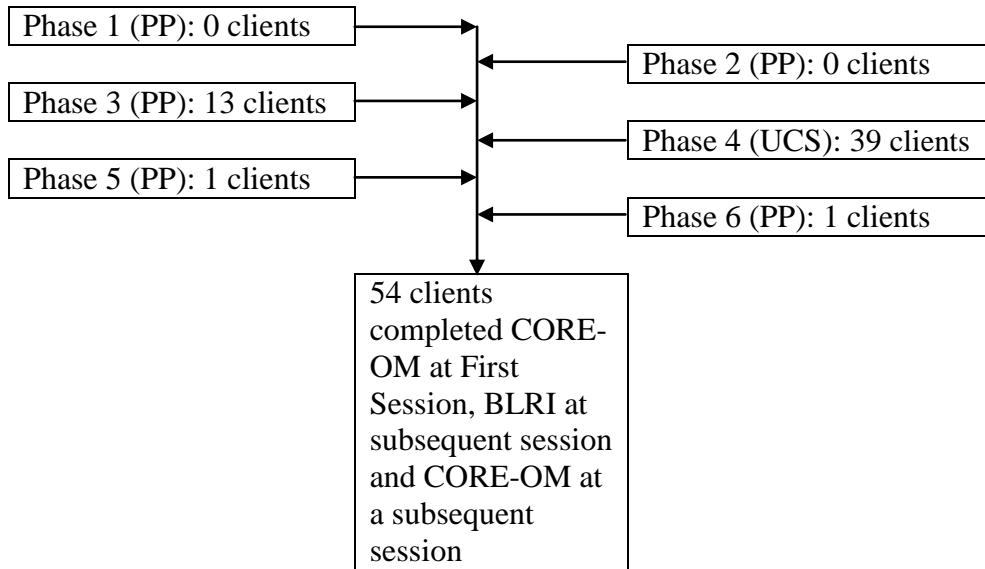
Note: Clients subsequently excluded from prediction model (clients with standardized residuals > 3) came from phase 3 (1 client) and phase 5 (1 client) to leave a total of 90 clients in the revised predictor model.

Figure 47: Origin of clients (Phase of Protocol) in anxiety predictor sample: Clients who completed BAI at First Session, BLRI at a subsequent session and BAI at a subsequent session.



Note: Clients subsequently excluded from predictor model (clients with standardized residuals > 3) came from phase 4 (2 clients) to leave a total of 73 clients in the revised predictor model.

Figure 48: Origin of clients (Phase of Protocol) in distress predictor sample: Clients who completed CORE-OM at First Session, BLRI at a subsequent session and CORE-OM at a subsequent session.



Note: No clients were excluded from predictor model (no clients with standardized residuals > 3).