



The
University
Of
Sheffield.

Access to Electronic Thesis

Author: Germaine Thorpe
Thesis title: Identifying and Predicting Deterioration During Psychotherapeutic Interventions
Qualification: DClinPsy

This electronic thesis is protected by the Copyright, Designs and Patents Act 1988. No reproduction is permitted without consent of the author. It is also protected by the Creative Commons Licence allowing Attributions-Non-commercial-No derivatives.

This thesis was embargoed until February 2017.

If this electronic thesis has been edited by the author it will be indicated as such on the title page and in the text.

Identifying and Predicting Deterioration During Psychotherapeutic Interventions

By G. Leigh Thorpe

Thesis submitted to the University of Sheffield for the degree of Doctor of
Clinical Psychology

February 2012

The results, discussions and conclusions presented herein are identical to those in the printed version. This electronic version of the thesis has been edited solely to ensure conformance with copyright legislation and all excisions are noted in the text. The final, awarded and examined version is available for consultation via the University Library”.

DECLARATION

This work has not been submitted to any other institution for any other
qualification

Thesis Abstract

The literature review critically evaluated research articles focusing on deterioration in psychotherapy published since a watershed review by Mohr (1995). This review adopted the recommendations made by Mohr (1995) as a framework for the literature. A total of 28 studies were identified and reviewed using a quality rating system derived from Mohr's recommendations according to the extent to which these recommendations were implemented in the identified studies. The review yielded a higher average rate of deterioration (9-17%) in comparison with Mohr's review (5-10%). It was concluded that research into deterioration generally has continued to suffer from methodological limitations.

The intention of the research report was to investigate the phenomena of *overall deterioration* and *sudden deterioration* in a routinely collected data set collected from the Improving Access to Psychological Therapies (IAPT) initiative. Sudden deterioration was explored to determine whether it existed and how it may be defined. The rates of deterioration within the IAPT data were identified, and predictors of these were assessed.

It was determined that an appropriate definition for sudden deterioration was a reliable between-session change using the Patient Health Questionnaire–9; PHQ-9), that was not allied to sudden gains. Rates of sudden deterioration and overall deterioration were found to be 3.4% and 3.1% respectively. It was concluded that sudden deterioration exists as a phenomenon, is closely related to overall deterioration and that rates of deterioration in the IAPT dataset were relatively low.

Acknowledgements

I would like to thank my supervisors, Professors Michael Barkham and Gillian Hardy for their support, advice and patience throughout the whole research process.

I would also like to thank Dave Saxon, Amy Ashworth and Emma Briddon for their assistance at various points of the research.

It is also necessary to thank my family and friends, especially Dave, Alex, Kate, Emma, Lindsay and Mum for all their support, encouragement and tolerance over the past year.

Structure and Word Counts

Literature review

Prepared according to the guidance for the journal: Clinical Psychology: Science and Practice (Appendix B (i))

Empirical Report

Prepared according to the guidance for the journal: British Journal of Clinical Psychology (Appendix B (ii))

Section 1	Word Count
Literature Review	7998
References	1567
Total	9565
Section 2	
Empirical Report	11717
References	1226
Total	12943
Section 3	
Appendices	5314
Total Word Count	19715
Total with References and Appendices	27822

List of Contents

Section 1: Literature Review

Deterioration During Psychotherapy: A review of the literature since Mohr (1995)

Abstract	2
Introduction	3
Method	9
Section 1: Evaluation of the impact of the Recommendations from Mohr.	25
Section 2: Developments in Understanding of Deterioration in Psychotherapy since 1995.	37
Discussion and Implications	49
References	51

Section 2: Empirical Report

Identifying and Predicting Overall Deterioration and Sudden Deterioration among IAPT clients in routine Primary Care Settings

Abstract	61
Introduction	62
Method	72
Results	82
Discussion	115
References	125

Section 3: Appendices

Appendix A: Journal approval letter	133
Appendix B: Journal notes for contributors	134
Appendix C: Ethical approval letter	138
Appendix D: Substantial amendment ethical approval letter	142
Appendix E: PHQ-9 (excluded due to copyright)	144
Appendix F: GAD-7 (excluded due to copyright)	145
Appendix G: WSAS (excluded due to copyright)	146
Appendix H: IAPT Phobia questions	147
Appendix I: Breakdown of literature search	148
Appendix J: Account of literature quality rating	152
Appendix K: Distribution histograms for continuous variables	161
Appendix L: Results of interactions of significant variables	163

LITERATURE REVIEW:

**Deterioration During Psychotherapy: A review of the literature since
Mohr (1995)**

Abstract

This literature review critically evaluated research articles focusing on the phenomenon of deterioration in psychotherapy published since a watershed review by Mohr (1995). The current review adopted the recommendations made by Mohr (1995) as an organisational framework for the literature. A total of 28 studies were identified and reviewed using a quality rating system derived from Mohr's recommendations according to the extent to which these recommendations were implemented in the identified studies. The higher quality rated articles only were then considered in summarising the current state of the literature. The review yielded a higher average rate of deterioration (9-17%) in comparison with Mohr's review (5-10%). It was concluded that research into deterioration generally has continued to suffer from methodological limitations.

Introduction

Psychological therapists aim to alleviate psychological distress via the delivery of a range of therapeutic interventions. Paralleling this professional activity, there is a substantial body of evidence spanning 60 years to suggest that many of these interventions are effective (Crits-Christoph, 1992; Lambert & Ogles, 2004; Orlinsky, Ronnestad, & Willutski, 2004). Extensive efforts have been made to improve the research evidence through endeavours such as treatment manuals, clinical trials, and effectiveness research (e.g., Addis, 1997; Safer, Robinson & Jo, 2010).

However, there is also evidence that some clients deteriorate during therapy (Bergin, 1971; Shapiro & Shapiro, 1982; Strupp, Hadley, & Gomes-Schwartz, 1977). Bergin (1971) reviewed the research evidence and observed a consistent finding in 28 of 32 identified articles that some proportion of participants deteriorated significantly more than in their respective control groups. In response, Bergin advanced a number of recommendations including that research effort should focus on the processes of change, employ better outcome measures, and aim to determine the characteristics of clients who deteriorate. He also recommended that the use of specific rather than global outcome measures was preferable and that increasing the number of assessments was beneficial in uncovering deterioration and gaining a wider understanding. Bergin also lamented that several studies used the 'undifferentiated category of not improved or worse¹.' He advocated further studies into the characteristics of both clients and therapists who

¹ Bergin (1971) In Handbook of Psychotherapy and Behavior Change (Eds Bergin & Garfield) p.248.

experience deterioration as well as process research investigating those therapeutic dyads that resulted in deterioration.

Mohr (1995) conducted a critical review of both general deterioration effects and negative outcome in psychotherapy that echoed and updated a number of recommendations made by Bergin (1971). Mohr identified 42 articles published during the period 1950 to 1994 that mentioned negative outcomes and noted a number of methodological limitations in these papers as well as significant gaps in the knowledge base relating to deterioration. Mohr's review has since become a landmark/watershed study due to the cohesive analysis of the state of evidence at that point and his vehement request for this area of research to become more central to effectiveness literature.

In light of the importance of Mohr's review, the primary aim of the current review was to evaluate the impact of Mohr's (1995) article and to consider the extent to which his recommendations, with particular regard to methodological improvements, have been adopted and implemented in the field of the psychological therapies. This is considered important as the majority of these limitations were highlighted almost 25 years before by Bergin in 1971, and evidently were not rigorously put in practice in the intervening years in order to remain recommendations from Mohr's review. The secondary aim was to update and re-evaluate the current theoretical understandings and knowledge base of those clients who deteriorate during receipt of a psychotherapeutic intervention.

Methodological recommendations by Mohr (1995)

In order to provide a framework for the review, this section sets out a brief summary of the methodological recommendations made in Mohr's (1995) review under 5 headings: (1) identification of negative responders, (2) assessment strategy – outcome measures, (3) assessment strategy – source/perspective; (4) timing of assessments; and (5) recognition of negative responders as separate group.

1. Identification of negative responders

Mohr (1995) recommended ensuring consistency across research studies by ensuring that the same concept was being investigated. Accordingly, it is important that a common standard is implemented. This requires adopting definitions of deterioration that account for both statistical change (i.e., greater than random fluctuation) and clinical significance (i.e., indicative of a different psychological state). Therefore Mohr advocated methods initially proposed by Jacobson and Revenstorf (1988) and subsequently superseded by more specific recommendations by Jacobson and Truax (1991) regarding two components: reliable change index, which relates to the *extent* of change, and clinically significant change, which relates to the *end point*. Both these procedures are briefly outlined below.

The reliable change index (RCI) refers to a pre- to post-treatment change in outcome measure scores that is statistically larger than the standard error of measurement for that outcome measure and is a function primarily determined by the reliability of the outcome measure. Thus, the higher the

reliability of the measure, the smaller the measurement error. Also the change in score that would be required for it to be considered reliably improved or deteriorated is also smaller. This procedure therefore yields differing RCIs for different outcome measures. Overall, this procedure ensures that the change is unlikely to be a product of random fluctuations in test scores.

The concept of clinically significant change (CSC) refers to the notion that the post-treatment score is required to move to membership of a different population: from clinical to non-clinical for clinical improvement and vice versa for clinical deterioration. Combining both the concept of reliable change and clinically significant change yields a definition of deterioration in which a client's score reliably moves from a non-clinical to a clinical population by more than the RCI. An alternative definition requires only that the score changes by the RCI and would therefore encompass all incidents of reliable deterioration regardless of the population.

2. Assessment Strategy – appropriate outcome measures

Mohr (1995) argued that the difference between specific and global outcome measures was of particular importance. Use of global measures can result in the likelihood of finding change being diminished by distinct changes in different areas of functioning. Covariance is also important and should be considered given that improvement achieved in one area may also be accompanied by deterioration in another area. Accordingly, outcome measures should cover a number of areas of client experience. Two key

points arise from these considerations: 1) specific measures are more likely to tap any negative effects than are global measures, and 2) use of a variety of measures is more appropriate than using a single measure.

3. Assessment Strategy - Deterioration recognised by whom?

Mohr (1995) placed a central importance on the source of the evaluation of deterioration. Deterioration can be measured and understood from a number of different sources including the following: society, client, mental health professionals, family members, or in group therapy by other group members. There is no reason to assume that all of these sources will correlate highly with one another due to each of them having different priorities. In order to fully explore and understand the phenomenon of deterioration, it is important to consider the views of a number of different stakeholders of therapy and not just those directly involved (i.e., client and therapist).

4. Timing of Assessments

A further issue considered by Mohr (1995) related to the timing of the assessments. This is important as there are potentially confounding factors that limit conclusions about deterioration that can be drawn from results. For example:

- If assessments are taken during treatment, deterioration effects can be more likely to be found than when taken overall (e.g., relaxation training and increase in anxiety symptoms). There is a need for a

distinction to be made between negative outcomes and negative influences that diminish the positive effects of therapy.

- Should assessments focus on the follow-up period rather than post-treatment, there is the potential for the data to capture the phenomenon of relapse rather than deterioration.

Therefore, Mohr advocated that assessments should ideally be completed at key stages in therapy as follows: pre-treatment, during treatment, post-treatment, and follow-up stages.

5. Recognition of negative responders as a separate group

Mohr (1995) also acknowledged the need to differentiate clients who deteriorate from those who do not respond to therapy. Hence, it is important to consider non-responders and negative responders as separate client groups. It cannot be assumed that positive and negative responders are at opposite ends of continuum as they have been shown, in some instances, to have a curvilinear relationship (i.e., negative responders can be more closely related to positive responders rather than non responders, Mohr et al., 1990).

The purpose of this literature review is to evaluate the research that has reported deterioration published since Mohr's 1995 review, with regards to the methodological recommendations provided. In addition, to update the evidence provided by these studies regarding deterioration.

Method

Databases and search terms

The electronic databases PsychInfo and Web of Knowledge (including Medline) were utilised in searching for literature focusing on deterioration.

The search was carried out using four key terms:

1. Deterioration (including deteriorat*, negative outcome, negative respon*)
2. Psychotherapy (including psychotherap*, therap*, counselling, counseling)
3. Predictors (including predict*, client, treatment)
4. Harm (including harmful treatment, iatrogenic, treatment failure)

These terms were chosen to ensure a broad scope of articles and to negate the impact of different researchers using different definitions of terms. Each term was then combined in the two databases (See Appendix I for a breakdown for these searches).

Inclusion and Exclusion Criteria

The search was refined by adoption of inclusion and exclusion criteria.

Inclusion criteria

There were three inclusion criteria: (1) studies published between January 1995 and April 2011; (2) the published articles were available in English language; and (3) the study sample comprised adult participants.

Exclusion criteria

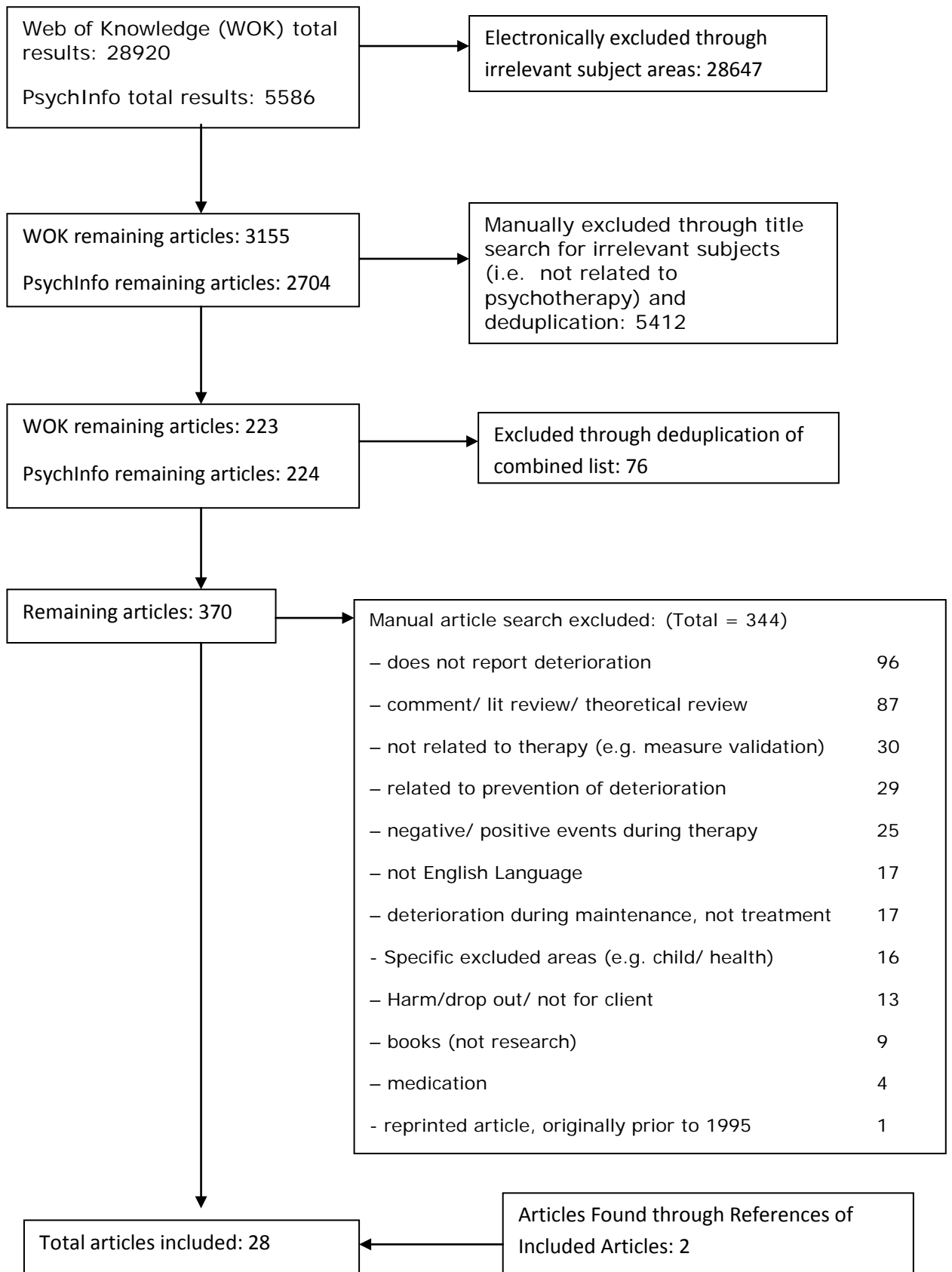
A total of 7 exclusion criteria were applied to studies that met one or more of the following: a focus on (1) physical rather than psychological/mental health; (2) deterioration during a follow-up period; (3) negative events during therapy; (4) interventions that cause harm²; (5) deterioration for those other than the client; (6) drop-out from therapy rather than deterioration; and (7) prevention of deterioration as opposed to reporting deterioration.

Output

The search yielded a total of 34,506 articles from PsychInfo and Web of Science databases combined. Successive stages of excluding articles on specified criteria led to a pool of 26 articles. For further identification of papers, reference lists of the 26 identified articles were searched manually and a further 2 articles were identified for inclusion, yielding a final datapool of 28 articles. A detailed account of the stages involved in the sifting of articles is presented in Figure 1.

² Described further in Appendix I.

Figure 1: Literature Search Flow Diagram



Quality Rating

In order to give greater weight in the review to better designed studies, a quality rating system was adopted. The 28 identified articles were rated using a quality rating system developed for the specific purpose of this review by LT on the basis of Mohr's (1995) methodological suggestions. The rating scale comprised six methodological areas as previously noted: (1) Identification of negative responders; (2) Assessment strategy – outcome measurement (global or specific); (3) Assessment strategy – outcome measurement (number of measures); (4) Assessment strategy – measured by more than one perspective; (5) Timing of assessment – pre/post, follow up, during treatment; and (6) Recognition of negative responders as a separate group

Each of the 6 areas comprised a 3-point scale as to the extent to which a study met each of these criteria: 1 = No efforts made by investigators to follow recommendation; 2 = Partial efforts made but not implemented methodically; and 3 = Investigators methodically followed recommendation³. The obtained scores for each area (i.e., 1, 2, or 3) were then squared (i.e., 1, 4, 9) in order to yield greater differentiation between the ratings. A detailed account of the rating system is presented in Appendix J. The 28 articles were ordered chronologically and then articles assigned alternately into two groups. The order of articles in the second group was then reversed, enabling determination of the effect of rating fatigue on chronological order.

³Note: In relation to the 6th criterion - *Recognition of negative responders as a separate group*, a score of 2 was not possible.

The author (LT) rated all 28 articles. In addition, a representative sample of 9 articles was also rated by an independent rater (IR). The IR was a clinical psychologist who received 2 hours training by LT in the use of the rating scale. The sample of 9 studies was chosen by ordering the 28 articles according to the rating they achieved from LT, then split into three groups comprising approximately equal numbers (allowing for studies with the same score to be in the same group) from which the highest-, mid-, and lowest-rated article from each group were selected. This ensured a test of the range in quality of the studies. The sample articles were arranged alphabetically and rated by the IR. The agreement level between the two raters for the 9 articles as determined by intraclass correlation (ICC; Shout & Fleiss, 1979) was .89.

A mid-point on the quality rating scale was taken as a cut-point for quality. Hence, articles that did not achieve at least 50% of the potential quality score were considered of lower methodological quality. This somewhat arbitrary cut-off was chosen as it placed those articles with the weakest methodology in the lower group, and kept those articles with more robust methods in the higher group. In the development of this cut-point a number of others were also tested (these were 25%, 33%, 40%, 60% and 66%) however 50% offered the most balance between the stronger and weaker methodologies. To give less weight to the lower rated articles, they have only been utilised in relation to the first aim of this review.

Review of Studies

This review first presents a summary of the 28 studies, reporting on the lower- and higher-quality studies separately. Then the main section of the review comprises two sections based on the two aims described above. The first section evaluates the impact of Mohr's (1995) review and considers the extent to which his methodological recommendations have been adopted and implemented. The second section re-evaluates and updates the current theoretical understandings and knowledge base of those clients who deteriorate during the course of a psychotherapeutic intervention. The review draws on some studies to a greater extent than others as a function of quality, amount of information presented (specifically regarding deterioration), and relevance to the topic considered.

Summary of review

Table 1 presents a summary of the 11 lower-quality studies ordered from lowest-rated study to the highest-rated study in this grouping. Studies are numbered from #1 to #11 for ease of future collective referencing in the text. Table 1 details the authors, client population sampled and country in which the study was carried out, therapy employed, deterioration rate reported, criteria used for determining deterioration, and study findings.

Of the 11 studies in this group, 7 studies were undertaken in the USA, two in Germany, one in Denmark, and one reported studies in Canada and the USA. A total of 9 of the 11 studies were published since 2000. The average

quality rating for the 11 studies was 39.5% and ranged from 27.8% to 48.1%. Scrutiny of the publication year and quality ratings showed no significant correlation between year of publication and quality ($r(9) = -0.073, p < 0.415$). These 11 studies are reviewed only in relation to evaluation of the impact of Mohr's review (Aim 1) as their conclusions regarding deterioration are less reliable as a result of methodological limitations.

Table 1 – Literature Summary of 11 Lower Quality Studies (ordered from lowest to highest)

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#1. Fournier, DeRubeis, Shelton, Hollon, Amsterdam, & Gallop (2009)	180 participants with moderate to severe depression as primary diagnosis USA	16 week Cognitive Therapy (N=60) or antidepressant medication (N=120)	Not reported	Higher than predicted score based on expected curve.	Chronic depression, older age, and lower IQ predicted poorer response.	27.8
#2. Samstag, Batchelder, Muran, Safran, & Winston (1998)	73 participants USA	Brief psychotherapy	27.4% = 'poor outcome' deterioration during therapy.	Poor outcome defined as not RCI improvement.	Therapists ratings not predictive of poor outcome	33.3
#3. Kanter, Landes, Busch, Rusch, Brown, & Baruch (2006)	2 participants diagnosed with depression USA	Functional Analytic Psychotherapy	50% (1 participant)	Self report diary of target problems.	Questions raised regarding clients for whom social interactions are not reinforcing.	33.3
#4. Pike, Walsh, Vitousek, Wilson, & Bauer (2003).	33 post hospitalisation participants with Anorexia Nervosa USA	Randomised assignment to CBT or Nutritional Counselling (NC)	22% - 53% relapse for CBT and NC respectively.	Relapse reported as weight loss, increased suicidality, or increased depression and referred to alternative care.	CBT offers both lower levels of deterioration and better improvement than NC.	37.0

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#5. Rufer, Fricke, Moritz, Kloss & Hand (2006)	104 participants with OCD Germany	CBT inpatient treatment	42.2% non-responders.	Non-responders not separate from negative responders. 35% decrease on outcome measure = positive response	Those participants with hoarding behaviour were more likely to deteriorate during treatment.	38.9
#6. Scogin, Floyd, Jamison, Ackerson, Landreville & Bissonette (1996)	188 participants from 5 studies USA & Canada	Self administered treatments	1% (clinician rated measure) - 9% (self report measures)	1 standard error of measurement on Beck Depression Inventory (BDI).	Within normal deterioration range for therapist provided therapy, and so they argued not a higher risk for self-administered treatments.	42.6
#7. Samstag, Muran, Wachtel, Slade, Safran & Winston (2008)	48 client and therapist dyads with participants with personality disorders. USA	30-session manualized treatment with psychodynamic, CBT, supportive therapy and a relational therapy.	33.3% defined as 'poor outcome' planned percentage	Poor outcome defined as not RCI improvement.	'Poor outcome' dyads demonstrated the highest degree of 'hostile complementarity.'	42.6
#8. Moos, Nichol, & Moos (2002)	8427 participants treated for substance use disorders (2809 deteriorated during treatment, matched with 5618 stable or improved participants) drawn from data of 21036 participants USA	Varied, most common being twelve step self help groups, CBT, psychodynamic psychotherapy and eclectic approaches.	33.3% (prescribed/ chosen percentage) (13.4% deterioration from full data)	Deteriorated by one standard deviation or more (defined as <i>two</i> more problems on ASI)	Risk factors for deterioration were identified as younger age, non-married status, residential instability, long term drug use, prior arrests, prior alcohol treatment, combined drug and alcohol abuse, cocaine abuse and psychiatric problems.	42.6

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#9. Moos, Moos, & Finney (2001)	Data of 2616 participants chosen from 21036 and three matched groups (872 in each – improved, nonresponsive, deteriorated) USA	Varied, most common being twelve step self help groups, Cognitive Behaviour Therapy (CBT), psychodynamic psychotherapy and eclectic approaches.	33.3% (prescribed/ chosen percentage) (4.1% deterioration from full data)	Deteriorated by one standard deviation or more (defined as <i>three</i> more problems on Addiction Severity Index; ASI)	Deterioration predicted by younger age, African-American race, psychiatric symptoms, arrests, prior drug treatment, recent inpatient or residential care. Also alcohol and drug use combined, personality disorder, shorter episode of care, fewer outpatient visits.	42.6
#10. Jensen, Mortensen, & Lotz (2010)	236 participants Denmark	39 session (3 month) short term psychodynamic group therapy.	2.4% - 12.3% total Symptom Check List-90-90-R Global Severity Index score (SCL-90-R GSI; 1.0% - 21.7% among different disorders)	Reliable change index using cultural (Dutch) norms	Many clients do not improve with short-term therapy and may need longer treatment.	46.2
#11 ⁴ . Moritz, Fricke, Jacobsen, Kloss, Wein, Rufer et al. (2004)	53 participants with Obsessive Compulsive Disorder Germany	Exposure response management with either social skills training, problem solving or stress-coping.	47.2% non responders/ deterioration.	Non-responders not separate from negative responders. 35% decrease on outcome measure = response.	Higher levels of positive schizotypal features predict non-response.	48.1

⁴ This number refers to the article's quality ranking within the 28 articles.

Table 2 presents a summary of the 17 higher-quality studies ordered from lowest-rated study to the highest-rated study in this grouping. Studies are numbered from #12 to #28 for ease of future collective referencing in the text. Table 2 presents information on the authors, client population sampled and country in which the study was carried out, therapy employed, deterioration rate reported, criteria used for determining deterioration, and study findings. Of the 17 studies in this group, 7 studies were undertaken in the USA, 4 studies in Germany, two studies in the UK, and one each in Italy, Norway, and Spain.

A total of 12 of the 17 studies were published since 2000. The average quality rating for the 17 studies was 69.2% and ranged from 51.9% to 90.7%. The top three rated studies^[#26-#28] all predated 2000. There was no significant correlation between year of publication and quality ($r(15) = -.229, p < .189$).

These studies are compared with the lower quality studies in the evaluation of the impact of Mohr's review (Aim 1) section but are the only source for the review when considering developments in understanding of deterioration in psychotherapy (Aim 2).

Table 2: Literature Summary of 17 higher Quality Studies (ordered from lowest to highest)

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#12. TARRIER, Pilgrim, Sommerfield, Faragher, Reynolds & Graham et al. (1999)	72 participants with chronic PTSD UK	Cognitive therapy (CT) vs. Imaginal Exposure (IE) (randomised)	16.7% total (12.5%IE, 4.2%CT) at post, 11.1% total follow up (6.9%IE, 4.2%CT)	Worsening = increase of CAPS score	Significantly more participants 'worsened' over treatment for IE that CT (not sig at follow up.) Those tended to miss more sessions, thought therapy less credible, and therapist rated as less motivated.	51.9
#13. Lopez-Goni, Fernandez-Montalvo, Menendez, Yudego, Garcia, Esarte (2010)	112 participants Spain	Therapeutic Community program including group and individual therapy, and occupational therapy.	0.0% -25.6% had deterioration on 9 life domains. (including follow up period)	Reliable change index	Life domains with highest levels of deterioration were alcohol, medical, job satisfaction and economic.	51.9
#14. Ilgen & Moos (2006)	3322 male participants with substance use disorders (15 site) USA	Residential treatment for substance use disorders – individual and group therapy incorporating behavioural skills training, relapse prevention and peer support	13% deteriorated during treatment, 2% non response	Any increase in score considered deterioration – non responders = no change	Those who deteriorated were more likely to also suffer psychosis, lower self-efficacy, reliance on coping by emotional expression and have negative view of treatment.	51.9

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#15. Fricke, Moritz, Andresen, Jacobsen, Kloss, & Rufer et al. (2006)	55 participants with OCD as primary diagnosis Germany	CBT	44% non-responders	Non responders not separate from negative responders. 35% decrease on outcome measure = response	Those with Personality Disorder responded statistically similar to those without (less but not significant)	57.4
#16. Lunnen, Ogles, & Pappas (2008)	66 out patient participants USA	Individual therapy.	24.5% deterioration	Reliable change index	Satisfaction surveys should also include questions related to deterioration, as limited prediction of satisfaction to end point functioning.	59.3
#17. Pekarik & Wolff (1996)	152 participants from 3 community mental health centres with mild to moderate disorders, primarily adjustment related, dysthymia and personality disorder. USA	22 therapists using family systems, CBT, eclectic, gestalt, Alderian and reality therapy.	Between 28% - 59% (on different measures from different perspectives)	Clinically significant change including reliable change index	Very low levels of correlation between satisfaction and treatment outcome.	63.0
#18. Callahan, Almstrom, & Swift (2009)	76 participants (archival data) treated in a training clinic. USA	CBT informed work	6.6% deterioration on BDI-II	Reliable change index (RCI)	16% of variance in outcome (including deterioration) associated with supervisors.	66.7

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#19. Beutal, Hoflich, Kurth, & Reimer (2005)	83 inpatient participants. Germany	Short term psychotherapy inpatient treatment including individual psychodynamic and group interventions	15.7% deteriorated to follow up on GSI; 4% - 5% within delayed outpatient further psychotherapy group for somatic, psychological, and social relationships.	Deteriorated from 'healthy' category to 'dysfunctional.' No statistical criteria employed.	Predictors of deterioration included infantile object relationship pattern, social avoidance, negative vocational changes and lack of a confidant.	66.7
#20. Lindgren, Werbart, & Philips (2010)	131 participants aged 18 – 25. Sweden	Individual or group psychoanalytic psychotherapy	3.1% - 6.1%	Reliable change index for each outcome measure	3.1% (on IIP) - 6.1% (on GSI) from intake to termination. From intake to follow up – 3.8% (on GSI), 4.6% (on IIP.) No further analysis on these individuals.	72.2
#21. Lincoln, Rief, Hahlweg, Frank, Witzleben, & Schroeder (2005)	287 participants diagnosed with social phobia seeking treatment. Germany	Short intensive treatment including in vivo exposure and cognitive restructuring	13% deteriorated at follow up.	Reliable change Index.	Only predictor of deterioration at follow up was number of feared situations, but tended also to have younger age, larger number of feared situations and higher levels of anxiety in these situations.	72.2

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#22. Prestano, Lo Coco, Gullo, & Lo Verso (2008)	8 female participants with either Anorexia Nervosa or Bulimia Nervosa. Italy	Group Analytic Therapy	12.5% deterioration (1 participant)	Reliable change index and clinical significance	The participant who suffered deterioration has less relatively psychopathology, and missed a greater number of sessions.	75.9
#23. Kellett, Clarke & Matthews (2007)	176 participants Primary Care clients with symptoms of poor mental health. UK	Group based psycho-educational CBT, individual CBT and psychodynamic interpersonal psychotherapy	0% - 5% deterioration on all measures	Reliable change Index	Similar levels of deterioration for all treatments, although trend towards group CBT as higher deterioration.	75.9
#24. Hartmann, Orlinsky, Weber, Sandholtz, & Zeeck (2010)	40 participants with Bulimia Nervosa. Germany	Inpatient treatment vs. day hospital treatment	45% non responders	Treatment failure regarded as continuing to meet all diagnostic criteria (non response and deterioration not separated)	Intersession experiences, particularly recreating therapeutic dialogue with negative emotions significantly predict failure/ non-response.	75.9
#25. Von der Lippe, Monsen, Ronnestad, & Eilertsen (2008)	373 out patients participants from which 28 cases chosen Norway	Dynamic Psychotherapy	50% (chosen percentage) 46.4% deteriorated rather than no change. (3.5% of total sample)	Lower than reliable positive change considered no or negative change, although they identified the participant who experienced no change	For those who experienced negative change, less 'match' in later therapy sessions between therapist and patient, hostile interplay bet therapist and patient predicted negative outcome, and clients rejected helpful efforts predicts most strongly.	81.5

Study number and reference	Population & Country	Therapy	Deterioration	Criterion	Findings	Quality Rating (%)
#26. Ogles, Lambert & Sawyer (1995)	162 participants suffering from depression who completed therapy. USA	Random assignment to CBT, IPT, medication with clinical management or placebo with clinical management.	Approx. 0% - 5% deterioration on all measures	Reliable and clinically significant change.	No further assessment carried out on this group – recommend further research in this area.	81.5
#27. Ford, Fisher & Larson, (1997).	74 male war veteran participants suffering from PTSD after traumatic service related trauma exposure and history of alcohol or substance abuse. USA	PTSD Residential Rehabilitation Program including practical support and planning, psycho-educational groups weekly individual psychotherapy incorporating exposure and relapse prevention and group psychotherapy	0% - 3 % of participants deteriorated on any of numerous measures.	Reliable change index.	Those with very low object relations more likely drop out or deteriorate, however still very small percentage.	81.5
#28. Pekarik & Guidry (1999)	93 adult private practice clients in New England. USA	Private practice therapists over 10 clinics treated with orientations of CBT, eclectic, psychodynamic and family systems.	Between 9% - 37% (on different measures)	Clinically significant change including reliable change index	Satisfaction with treatment does not correlate with treatment success or 'failure.'	90.7

Section 1: Evaluation of the Impact of the Recommendations from Mohr (Aim 1)

This section focuses on an evaluation of the impact of Mohr's (1995) review and considers the extent to which his methodological recommendations have been adopted and implemented. The review is structured around the recommendations made by Mohr. When referring to a group of reviewed studies, they are identified by their assigned number from Tables 1 and 2. However, when a specific study is referred to in detail, it is identified using standard referencing procedures as well as the assigned number for ease of referencing to the Tables 1 and 2.

Recommendation 1 - Identification of Negative Responders

The majority of articles (24 of 28) included a level of statistical difference, as recommended by Mohr (1995), in the identification of clients who improved, deteriorated, or did not change. Only 4 of the 28 studies^[#3, #12, #14, & #19] did not meet this standard. Fifteen of the articles reviewed^[#2, #7, #10, #13, #16-#18, #20-#23, & #25- #28] contained the RCI to ensure that reliable change could be determined. By contrast, studies that focused on Obsessive Compulsive Disorder^[#5, #11, #25] and used the Yale-Brown Obsessive-Compulsive Scale symptom checklist employed the strategy of considering a decrease in score of 35% as reliable improvement, a procedure that is standard in this area (Goodman et al., 1989).

Of the 4 articles^[#3, #12, #14, #19] that failed to use any statistical method of evaluating change, TARRIER et al. (1999:^{#12}) stated only that there was a ‘worsening⁵’ of the clinician-administered PTSD Scale score when comparing cognitive therapy with imaginal exposure. Kanter et al. (2006:^{#3}) discussed one ‘successful’ and one ‘unsuccessful’ treatment case of participants diagnosed with depression who were treated with Functional Analytic Therapy. They used self report diaries of identified target problems to evaluate outcome, however, no attempt was made to statistically assess the ‘unsuccessful treatment.’ Ilgen and Moos (2006:^{#14}) described the deteriorated⁵ group as those for whom there was an increase in ‘psychiatric symptoms’ during substance misuse treatment across 15 different clinical sites but did not provide any further information as to how this status was assessed. Beutal et al. (2005:^{#19}) investigated short-term inpatient psychotherapy from routine practice, based on psychodynamic formulation including individual and group treatments as a part of a therapeutic community. They employed the method of identifying deterioration through a change in classification from ‘healthy’ to ‘dysfunctional’ category on the Symptom Checklist-90-Revised (SCL-90-R; Global Severity Index), with no account of the change being reliable statistically. Kanter et al. (2006:^{#3}) chose to present one ‘successful’ and one ‘unsuccessful’ case, therefore the percentage of clients who deteriorated was chosen, while the rates for the other three studies were 16.7% (TARRIER et al., 1999:^{#12}), 15.7% (Beutal et al., 2005:^{#19}), and 15% (Ilgen & Moos, 2006:^{#14}).

⁵ TARRIER et al. (1999), page 16.

The criterion of clinically significant change (CSC) was less often utilised or considered as a method for defining change with 13 of the 28 studies^[#7, #10, #13, #16-#18, #20, #22, #23, #25-28] adopting this approach, at least for classifying those who improved. However, only three studies^[#16, #23, #27] used these criteria for classifying clinically significant deterioration (CSD). Ford et al. (1997:^{#27}) reported a rate of 0-3% while Kellett et al. (2007:^{#23}) found 0-5%, both of these rates being significantly lower than the average range of between 5 and 10% (Mohr, 1995). Lunnen et al. (2008:^{#16}) found a much higher level of 24.5% in assessing the relationship between therapeutic change and satisfaction.

Consideration of CSD as opposed to improvement requires participants to have pre-treatment scores below the clinical cut-off point. It is arguable that in such cases treatment may not be appropriate or required and so there may be much smaller numbers to analyse, potentially resulting in unrepresentative results. However this is likely to be less relevant in research in routine practice studies where clients entering services may have pre-treatment scores within the non-clinical range of an outcome measure. Another interpretation could again be the lack of systematic interest or focus in researching deterioration effects that led Mohr to make such similar recommendations as Bergin (1971).

Recommendations 2 and 3 - Assessment Strategy

The assessment strategy (related to quality criteria 2 and 3 as described earlier) employed in research can have a significant impact on the validity of

the results and conclusions. Mohr described the importance of considering certain issues when planning research. The type of outcome measure (specific or global) employed can impact on whether deterioration is found. Of the 28 articles, 4^[#2, #7, #10, #16] used only global measures and reported deterioration rates of 27.4% (Lunnen et al., 2008:^{#16}), 1% (clinician-rated) and 9% (self-rated; Jensen et al., 2010:^{#10}), 24.5% (Samstag et al., 2008:^{#7}), and between 2.4% and 12.3% (Samstag et al., 1998:^{#2}). Mohr's concern that global measures reduce the likelihood of finding change suggests that these results may be overly conservative and that the actual rates of deterioration may be higher.

The number of measures employed is important as change in human beings is multi-factorial and covariant. Accordingly some clients may improve in one area of assessment (e.g., functioning) and yet deteriorate in another (e.g., well-being). Six studies^[#1, #4, #8-#10, #13] used a single outcome measure and so are only able to consider change in one area of a client's range of experience.

It is also important to consider whose perspective is being captured, whether only the client's or therapist's, or from a wider range such as family members or other health professionals involved who can also be impacted upon as a result of treatment success or failure. Seven studies^[#1, #4, #10, #14, #18, #22, #23] assessed from a single perspective only whereas two studies^[#16 & #24] utilised outcome measures targeting the client, therapist, and other sources including significant others or independent observers.

Jensen et al. (2010:^{#10}) was the only article that scored the lowest ranking for each of the three areas related to assessment strategy. The authors evaluated short-term psychodynamic group therapy in order to assess the effectiveness and effect size of this routinely used model in the Netherlands and found deterioration rates of between 2.4% and 12.3% (using different culturally based cut-offs, Dutch and American respectively). The only outcome measure they used was the Symptom CheckList-90-Revised (SCL-90-R; Global Severity Index). The SCL-90-R was specifically identified by Mohr as the global measure commonly used in psychotherapy research which can lead to missing change effects where they may be present. Therefore this may represent an underestimate of the rate of deterioration found. However, it should be noted that in all other areas of the quality rating based on Mohr's recommendations this article scored well, as the negative responders were appropriately identified using the RCI and recognised as a separate group from clients who did not change. Hence, although potentially an underestimate of deterioration, the appropriate population were assessed suggesting the conclusion is likely to be defensible despite the limitations outlined above.

Only one article (Hartman et al., 2010:^{#24}) reached relatively high standards, using a variety of outcome measures, both specific and global, and from multiple perspectives. The authors used outcome measures assessing interpersonal problems, social adjustment, specific eating disorder measures, and a general measure of psychopathology. They also ensured different perspectives were used by using client, therapist, and trained independent raters (for the Structured Inventory of Anorexic and Bulimic

Syndromes; Fichter & Quadflieg, 2001). Unfortunately, they failed to recognise negative responders as a separate group and so limited the reliability of the results found, despite adopting such a strong assessment strategy.

In conclusion, researchers in this field have struggled to meet Mohr's recommendations in relation to assessment strategy. It is recognised that within research there can be potential constraints on the outcome measures employed and the number of sources of measurement used such as availability, routine practice, and funding limitations.

Recommendation 4 - Timing of Assessments

A consideration of the timing of assessments is important when conducting research into deterioration during psychotherapy. The vast majority of the articles reviewed did not conduct assessments at the full range of potentially relevant time periods: pre-therapy, during therapy, post-therapy, and at follow-up. Only two articles^[#24 & #28] achieved this level of data collection. This allows an intricate level of understanding about an individual's experience through therapy.

Three articles^[#8, #9, #13] gained the lowest possible score on the quality rating, meaning that they were unable to appropriately report on deterioration because the assessments were performed pre-therapy and at follow-up but not at post-therapy. Therefore their findings are potentially confounded by relapse rather than deterioration. This can be seen by the fact that significant

changes occurred within the group of participants who deteriorated in a comparison of imaginal exposure and cognitive therapy for clients with Post Traumatic Stress Disorder (PTSD; Tarrier et al., 1999:^{#12}). This article is included here as an illustration of the changes that may occur between post treatment and follow up periods. They found that significantly more participants 'worsened' over treatment for imaginal exposure than for cognitive therapy. However, this difference was not significant at follow-up. Hence significant differences can occur within this timeframe, which could be misinterpreted if the post-treatment assessments had not been taken.

In 23 of the 28 studies^[#1-#7, #10-#12, #14-#23, #25-#27] the timings of assessments were appropriate. However they were not as thorough as those articles that achieved the highest ratings by administering assessments at pre-, during, post- and follow-up of treatment. In certain circumstances, particularly when using data from routine practice, researchers only have access to limited assessments used by the service for clinical purposes. In addition obtaining follow up data especially can be difficult and expensive. This can still provide useful information regarding deterioration and those who respond negatively to psychotherapy.

Recommendation 5 - Recognition of Negative Responders as a Separate Group

Mohr (1995) and Bergin (1971) both lamented researchers who did not recognise negative responders as a separate group because this gives no reliable information regarding those clients who deteriorate. There is

evidence that non-responders can be from a different population than those who deteriorate (Mohr et al., 1990). They reported that as opposed to intuitive understanding, psychic distress would be related sequentially from positive effect, non-response, and then to negative effect. In fact it was found that the relationship was non-response, positive effect, then negative outcome. This demonstrated that potentially negative and positive responders might be more closely related on certain variables than negative responders and non-responders. This finding suggests there can be significant differences between non-responders and those who deteriorate during therapy. However, many researchers continue to combine these two groups and treat them as one population, including 10 articles in this review^[#1-5, #7, #11, #12, #15, #24]. They have focussed instead on separating those who reliably improved from those who did not and going no further.

Failing to separate those who deteriorate and those who do not change limits any conclusions that are proposed regarding these two groups. Samstag et al. (1998:^{#2}) concluded that therapist ratings were not predictive of poor outcomes. However, had they separated non-responders from those who deteriorated it is possible they might have secured significant results as it is plausible that therapists may detect deterioration more easily than those who do not respond to treatment. At least, it cannot be concluded from this evidence that therapist ratings were not predictive of deterioration.

Figure 2: Undifferentiated rates of deterioration and non-response ($N = 8$ review studies)

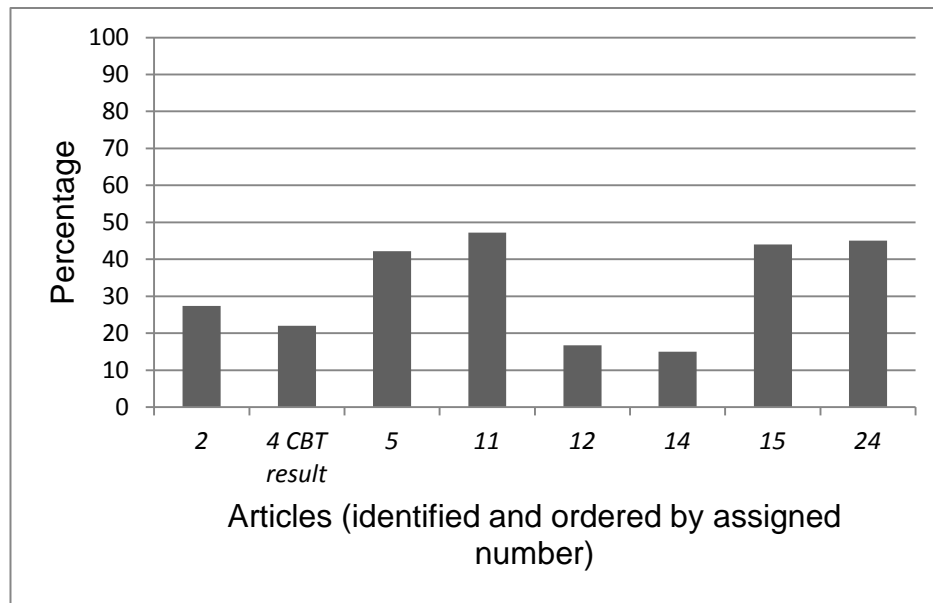
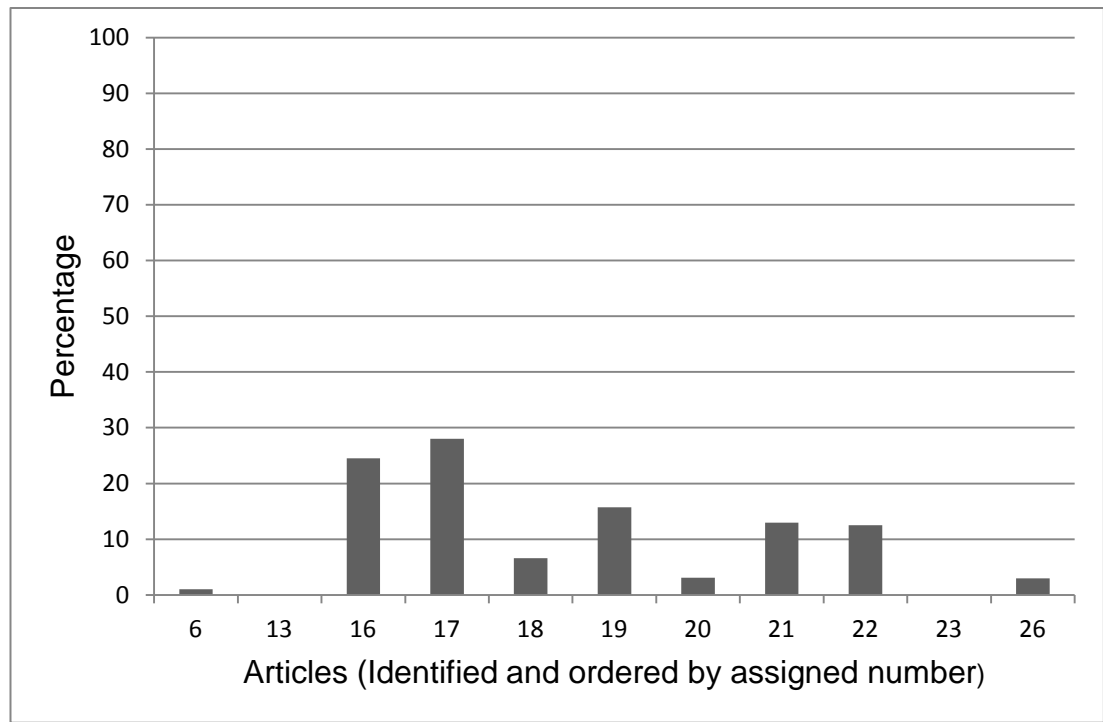


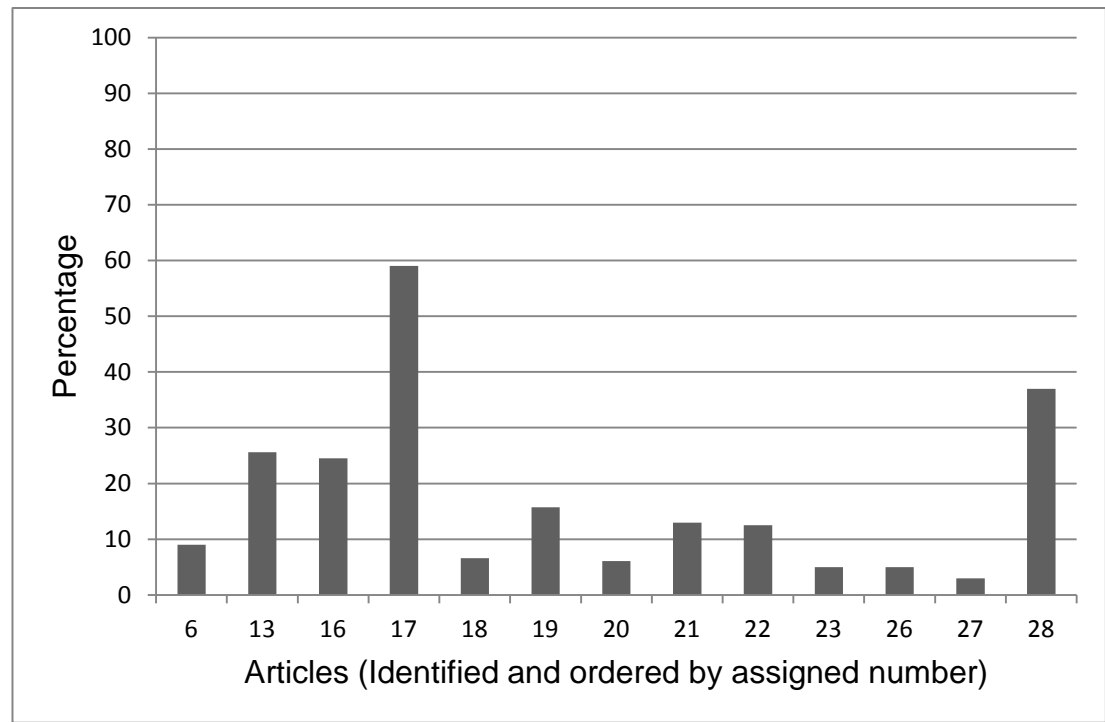
Figure 2 presents the levels of deterioration reported by those studies that did not separate clients who deteriorated from those clients who did not respond to treatment (arranged and ordered by assigned number). These 8 articles yielded a mean deterioration rate of 32.4% (SD = 13.6%). This percentage is considerably higher than the 5-10% range reported by both Mohr (1995) and by Lambert and Ogles (2004).

Figure 3: Percentage of deterioration reported (conservative range) (N = 13 higher quality studies)



Figures 3 and 4 represent the conservative (i.e., minimum) and liberal (i.e., maximum) values respectively of the ranges of deterioration provided by the 13 high methodological quality studies that recognised clients who deteriorated as a separate group and did not use a prescribed percentage in order to match with other groups. The respective means for these two groups were 9.0% (SD = 9.4) and 17.1% (SD = 16.1). The high values for the standard deviations show that there can be dramatic differences in the rates of deterioration reported by studies, even when focussing only on deterioration. Therefore it is important to separate the groups of people who deteriorate and those who achieve minimal change, as a usual or standard proportion cannot be presumed.

Figure 4: Percentage of deterioration reported (maximum range) (N = 13 higher quality studies)



There is the potential argument to be made that non-response may in fact represent deterioration due to the tendency for, especially extreme, outcome measure scores to regress towards to the mean (e.g., Hsu, 1995). However, as Mohr stated, this does not represent a view that is likely to resonate with either researchers or clinicians, especially when a control group has been employed. In addition, the fact that there has been demonstrated differences between non-response and negative response groups described above, suggests this argument lacks substance.

It is possible that this fairly common interpretation may stem from an understandable focus on treatment success rather than a broader desire to explore the full range of therapeutic change possible. This may be due to treatment success being generally the desired outcome, not just for the client

but also the therapist, the service involved, the profession as a whole, and the wider population. However, as scientists, researchers must also examine the different undesirable outcomes, and not dismiss them by bracketing them together and assuming them to be the same group.

Summary

In the research area of deterioration in psychotherapy, investigators continue to overlook the recommendations made 40 years ago by Bergin (1971) and echoed and updated by Mohr (1995). The lack of consistency in approach, assessment strategies, identification and even definitions of negative responders have led to an increasingly confusing state in this literature. It is vital that certain recommendations are followed including the recognition of negative responders as a separate group, and the identification of clients who deteriorate through at least statistical and if possible clinical significance. Assessment strategy is also relevant and can represent methodological limitations. However, this can be understood in relation to best practise standards in this area to be adhered to when constraints allow.

Section 2: Developments in Understanding of Deterioration in Psychotherapy since 1995 (Aim 2)

The second aim of this review is to report on the findings from the higher quality articles. A summary of these articles was presented in Table 2 above. Mohr's (1995) review was organised into the following areas: patient variables, therapist variables, process variables, and modalities. This review is structured similarly. However, in line with the evidence available from the articles reviewed here, therapist variables and modalities have been encompassed into a 'service level variables' section. A brief summary of Mohr's conclusions in each section is included in order to provide a context for the evidence published since his 1995 review.

Client Variables

Diagnosis

Mohr (1995) evaluated the deterioration literature available relating to the diagnoses of OCD and Borderline Personality Disorders. These revealed a complex picture and he concluded that 'diagnosis may not be the most useful indication of risk for negative outcome- at least not by itself⁶.' In this section articles that have focussed on what can be learnt about negative therapeutic change for individuals with specific disorders are considered, to examine if there is any further clarity in this area. Fricke et al. (2006:^{#15}) investigated the impact of Personality Disorder on the treatment of OCD with CBT. They

⁶ Page 11.

found that those with Personality Disorder responded statistically similarly to those without the diagnosis. However they noted that clients with Personality Disorders tended to respond 'less well', but this was not a statistically significant finding. In particular those with passive aggressive or schizotypal traits tended to do 'less well,' although they reported that there was still benefit from receiving treatment. The authors performed principal component analysis (PCA) whereas a factor analysis may have been more appropriate as PCA can limit how substantive the results are (Field, 2009).

Two articles^[#12 & #27] focussed on Post Traumatic Stress Disorder (PTSD) and found a broad range of deterioration rates within this group (0-16.7%). Ford et al. (1997:^{#27}) investigated residential treatment for male war veterans with PTSD and a history of alcohol or substance misuse. Although not clearly stated, it can be inferred that their aim was to investigate whether object relations may be a predictor of treatment outcome, as they argued that this might be more consistent with a dimensional approach to personality rather than categorical Axis II DSM diagnosis. They reported a very low rate of deterioration ranging from 0% to 3%, and those with very low object relations were more likely to either drop out from therapy or deteriorate.

Tarrier et al. (1999:^{#12}) compared cognitive therapy (CT) with imaginal exposure (IE) for participants diagnosed with PTSD and reported on differences between improvement and deterioration rates throughout therapy. They stated the following levels of 'worsening' for clients: 16.7% total (12.5% for IE; 4.2% for CT) at post-therapy, and 11.1% total at follow-up (6.9% for IE, 4.2% for CT). Significantly more participants 'worsened' over treatment for IE than CT although this difference was not significant at follow

up. These participants tended to miss more sessions, thought therapy less credible, and the therapist was rated as less motivated.

However, the Tarrier article has significant limitations including that the assessment of credibility and motivation from the therapist utilised non-validated questions. Also the term 'worsening' was not defined and Devilly and Foa (2001) note that Tarrier (the lead author) had, in a personal communication, described 'worsening' as a lack of significant improvement which, as described earlier, does not separate negative responders from non-responders. Devilly and Foa (2001) also argued that the description of IE was not how this intervention is generally performed, which further weakened the results of this study.

Mohr's review contained only limited mention of substance misuse, suggesting only that there was some evidence indicating that this may contribute to negative outcomes for those with Borderline Personality Disorders. In the present review, articles were more focussed on residential treatments and found a wide range of deterioration rates. Ilgen and Moos (2006:^{#14}) worked with residential participants with substance use disorders and found a 13% deterioration rate. Deterioration was defined as occurring either in substance misuse or psychiatric symptoms when baseline psychiatric symptoms were controlled. Those who deteriorated were more likely to also suffer psychosis, lower self-efficacy, have reliance on coping by emotional expression, and have a negative view and poorer experience of treatment. Additionally they tended to use substances whilst in residential treatment, or drop out of therapy.

In more recent research treating 112 participants with substance misuse disorders within a therapeutic community, Lopez-Goni et al. (2010:^{#13}) provided group, individual, and occupational therapy. The level of deterioration they found was between 0% and 25.6%, and the life domains as assessed on the Addiction Severity Index (ASI) with highest levels of deterioration were alcohol, medical, job satisfaction and economic domains and domains that showed no deterioration were social and family. The use of the ASI as an outcome measure, however, has become controversial since Makela presented a review of the psychometric properties of the ASI based on 37 research reports and he concluded that 'ASI severity ratings should not be used in research or clinical decision making' (Makela, 2004, p 408) due to highly inconsistent reliability results.

In consideration of treatment of eating disorders, Mohr reported that behaviour therapy had been suggested as presenting an increased risk of deterioration. He noted though that this was based on only three case studies and so this conclusion was pre-emptive. Two studies^[#22] in this review focussed on psychodynamic models. Prestano et al. (2008:^{#22}) provided group analytic therapy for eight participants with either anorexia nervosa or bulimia nervosa and found that one participant, or 12.5% of clients, experienced deterioration. In consideration of this the authors reported that the participant who suffered deterioration had relatively less psychopathology and missed a greater number of sessions. Hartman et al. (2010:^{#24}) researched treatment for 40 participants with bulimia nervosa in day hospital and inpatient settings. They found that intersession experiences, particularly recreating therapeutic dialogue with negative emotions

significantly predicted the 45% of cases that experienced what they describe as failure. However they defined failure as non-response and deterioration and so limited the reliability of these findings for those clients who deteriorate.

Process variables/ Attitude to therapy

Mohr's (1995) review briefly covered 'process variables' within the literature he reviewed and reported that 'therapist competence and process are more strongly related to outcome than any other therapist or client variable⁷.' Within this review, studies focussed on the therapeutic relationship as a source of information regarding the phenomenon of deterioration during psychotherapy rather than specific effects of the therapist. Von der Lippe et al. (2008:^{#25}) examined out-patient data which comprised 14 participants whose dynamic psychotherapy treatment had resulted in positive change as well as 14 who experienced negative effects or no change. They then explored the process of therapy for these clients discovering that for those who experienced negative change there was less 'match' in later therapy sessions between therapist and client, hostile interplay between therapist and client predicted negative outcome, and clients rejecting helpful efforts predicted deterioration most strongly.

Hartman et al. (2010:^{#24}) researched treatment for anorexia nervosa in day hospital and inpatient settings. They found that intersession experiences,

⁷⁷ Page16.

particularly recreating therapeutic dialogue with negative emotions, significantly predicted what they described as failure. However they defined failure as non-response and deterioration. It can be seen that there has been more focus on the client and the therapeutic relationship rather than on the therapist, despite the suggestion from Mohr that this may be the most powerful influence in deterioration. However, negative influences within the therapeutic relationship appears to have potential as a burgeoning area of deterioration research.

Service levels variables

Comparisons of Models

Regarding a comparison between different models, Mohr noted that it would not be of benefit for researchers or proponents of different therapeutic models to use deterioration research to dismiss other models. He noted that expressive-experiential therapies, particularly gestalt therapy, tend to have higher levels of deterioration. However there is the potential to discover if certain models are better suited to certain disorders or situations, or could present a lower risk of deterioration. In the comparison studies presented below, however, there were no statistically significant differences between psychotherapeutic treatment types employed.

In a comparison between group-based psycho-educational cognitive-behavioural therapy (CBT), individual CBT, and psychodynamic interpersonal psychotherapy (IPT) for participants with poor mental health in

a primary care setting, Kellett et al. (2007:^{#23}) found 0-5% deterioration rates on a variety of measures. They found that there were similarly low levels of deterioration for all treatments, although there was a trend towards group CBT ending in higher rates of deterioration. They argued that this represents an increased need for mental health support alongside this group, which is similar to Mohr's recommendation regarding psycho-educational groups and self-administered treatment. Ogles et al. (1995:^{#26}) found a similarly low rate of deterioration from 0 to 5%. In this study clients were randomly assigned to CBT, IPT, medication (Imipramine) with clinical management, or placebo with clinical management for participants with depression. Although there was no statistical difference between the psychotherapy treatment types, there was a trend for a higher rate of deterioration to occur within the CBT group. The authors recommended further research into this area although they did not perform any further analyses on this specific treatment group.

Cognitive Behavioural Therapy (CBT)

Two studies^[#15 & #21] focused solely on CBT. Fricke et al. (2006:^{#15}), described more fully above, investigated the impact of personality disorder on the treatment of obsessive-compulsive disorder when treated with CBT. They found that those with personality disorder responded statistically similarly to those clients without the disorder. Lincoln et al. (2005:^{#21}) found some predictors of deterioration when they analysed individuals who were diagnosed with social phobia treated with short intensive treatment including in vivo exposure and cognitive restructuring. The authors asserted that the

treatment provided was similar to CBT. However there were some stated differences in the delivery, such as the short duration and using specifically tailored real-life situations for habituation. Accordingly, there are limits to the generalisability of the findings. A total of 13% of participants experienced deterioration and the only predictor of deterioration at follow-up was the number of feared situations. However, they also tended to be younger, have a greater number of feared situations and experience higher levels of anxiety in these situations.

Psychodynamic and Psychoanalytic Psychotherapies

Mohr reported that there were a greater number of studies regarding personality type and predictors of negative response to therapy in these therapies than in other research areas. He concluded that psychodynamic style treatment might be less appropriate for more disturbed individuals, especially those who required additional structure such as borderline, psychopathic, narcissistic, or masochistic features.

Beutal et al. (2005:^{#19}) investigated short-term inpatient psychotherapy from routine practice based on psychodynamic formulation including individual and group treatments as a part of a therapeutic community. They found that predictors of the 15.7% deterioration included infantile object relationship pattern, social avoidance, negative vocational changes, and lack of a confidant. This is an elegantly recounted study with significant comments on deterioration and its predictors. Unfortunately they focused solely on deterioration after treatment and did not report on deterioration (or even lack

thereof) effects within treatment. This may represent a limitation due the possibility of capturing relapse in addition to deterioration. Despite this, they did use a wide variety of outcome measures, both specific and more general, and from both client and therapist perspectives.

Lindgren et al. (2010:^{#20}) researched participants who were young adults treated with either group or individual psychoanalytic psychotherapy. They found deterioration rates of between 3.1% and 6.1%. However, they performed no further analyses on this group. As stated above, Prestano et al. (2008:^{#22}) reported a deterioration rate for group analytic therapy for eating disorders of 12.5%.

Supervision

Mohr's review did not encompass clinical supervision or, in particular, the impact of supervisors on clients' response to treatment. Indeed, there remains scant research in this area. Callahan et al. (2009:^{#18}) explored the impact that supervisors have on the outcome of clients in a training clinic. They found a deterioration rate of 6.6% and discovered that a significant proportion of 16% of variance in outcome (including deterioration) was associated with supervisors. Hence, this represents an important potential predictor of negative response to therapy that is rarely investigated and may be a beneficial area for future study.

Satisfaction

There has been a recent increase in consumer satisfaction studies in psychotherapy as a method of service evaluation and clinical justification (e.g., Lambert & Ogles, 2004). Mohr did not consider this area in his review. However, these studies appear to find unusually high levels of deterioration and, most significantly, tend to consider deterioration in a broader sense from a wider range of perspectives. For this reason, this area is included in this section. All studies presented here report limited, if any, correlation between client satisfaction levels and treatment outcome.

Lunnen et al. (2008:^{#16}) reported a high level of deterioration of 24.5% in out-patient psychotherapy services while examining the relationship among satisfaction, symptomatic improvement, perceived change, and end point functioning from multiple perspectives. The authors also found that satisfaction was not related to change in symptoms and, therefore, is insufficient to be used as the sole evaluation of a service. Importantly they not only employed clients and their therapists, but also significant others living with the participant. They discovered only limited relationships between satisfaction and the three areas investigated. There were several limitations of this study, partly due to the outcome measures used. They were very specific in choosing clusters of larger outcome measures (e.g., Patient Questionnaire (Q-P) cluster 2 (amount of change) and 3 (Current status); Strupp, Fox & Lessler, 1969). However, after finding very close correlations between these Q-P cluster scores and the Client Satisfaction Questionnaire (CSQ-8; Larsen et al., 1979) they found a number of them to be very similar. Hence the assessment strategy they employed was overly complex yet not

systematically planned. According to the recommendations by Mohr, using such broad assessments can lead to underestimating change, either positive or negative. They recommended that satisfaction surveys should also include questions related to client deterioration as these are generally not included.

Pekarik and Wolff (1996:^{#17}) sampled 152 participants from 3 community mental health centres with mild to moderate disorders, primarily categorised as adjustment-related disorders, dysthymia and personality disorder. Therapists treated clients with a variety of therapeutic models including family systems, CBT, eclectic, gestalt, Alderian, and reality therapy. The level of deterioration they found was between 28% and 59% on different measures from different perspectives. They also found very low levels of correlation between satisfaction and treatment outcome. Pekarik and Guidry (1999:^{#28}) again found satisfaction with treatment did not correlate with treatment success or 'failure.' Deterioration rates were found to be between 9% and 37% on different measures. Hence, it can be seen that higher rates of deterioration are uncovered in research into satisfaction as opposed to in clinical trials or other research using routine practice data.

Deterioration Rates

Mohr reported an average percentage of deterioration during psychotherapy of between 5 and 10%. Within the reviewed articles, deterioration levels were only included provided that they recognised negative responders as a separate group, and those who chose the percentage of negative responders to analyse as these may artificially inflate the result. The mean of the

percentage of deterioration from this group of 13 articles^[#6, #13, #16-#23, #26-#28] was found to be between 9% and 17% (See Figures 3 and 4 above). This is slightly higher than that found by Mohr. However, this may be understood considering both the smaller number of studies and that some of the researchers chose to investigate a wider understanding of deterioration from different perspectives, particularly Pekarik and Guidry (1999:^{#28}) and Pekarik and Wolff (1996:^{#17}) who found very high levels of deterioration in this way when examining satisfaction surveys.

Discussion and Implications

In conclusion, it can be seen that this is a highly complex area of research and literature. There are many different areas to be considered such as different types of treatment intervention, numerous and multiple presenting difficulties along with client and therapist variables. In the time since Mohr's (1995) review, there has been some increase in the reporting and recognition of the importance of deterioration. However, this is still not standard and when it does occur it often lacks sufficient rigour or reflects partial understanding of the issues. It can be seen that deterioration remains an under researched issue in psychotherapy. It is also an area without much agreement regarding the most appropriate manner of assessing, analysing and reporting of results. The different strands of research are disparate with widely different objectives and even definitions of deterioration.

There continues to be evidence that a usually small yet significant number of individuals experience deterioration during therapy, with studies from the present review reporting a range from 9-17%. There continues to be few indications currently that there is a specific population that are at increased risk of deterioration, and the risk of this cannot yet be eliminated or reduced pre-emptively. There has been some work into predictors of deterioration, although only in certain areas, particularly in substance use disorders. However, this work has produced a very confusing picture with few, if any, clinically usable guidelines regarding who may be most likely to deteriorate. Potential predictors identified within the articles in this review are low object relations (Ford et al., 1997^{#27}; Beutal et al., 2005^{#19}), missing sessions (Prestano et al, 2008^{#22}), negative views towards therapy or the therapist or

rejection of helpful efforts of the therapist (Ilgen & Moos, 2006^{#14}; Von der Lippe et al., 2008^{#25}; Tarrier et al., 1999^{#12}), lower levels of initial psychopathology in eating disorders (Prestano et al., 2008^{#22}) or higher levels of initial psychopathology for clients with anxiety disorders (Fricke et al., 2006^{#15}).

There also have been certain improvements in the methodological research and reporting of deterioration. However, there are still many ways in which deterioration could be reported more significantly. It is strongly suggested that researchers do not report on both non-responders and negative responders combined as there is significant evidence that there is not always a linear relationship and they are separate groups. It is also recommended that deterioration should be reported much more widely as a general and expected part of all efficacy and effectiveness research, and in a systematic and cohesive manner. Any new treatment approach or type as a matter of course reports on efficacy or effectiveness, and it is important that deterioration is incorporated in reporting procedures to ensure continued safe and appropriate treatment of clients. In this way, more cohesion can be brought to this disparate and highly complex area of literature.

References

(Articles included in the review are marked = (lower quality) and ‡ (higher quality)

- Addis, M. (1997). Evaluating the treatment manual as a means of disseminating empirically validated psychotherapies. *Clinical Psychology: Science and Practice*, 4, 1-11. [doi:10.1111/j.1468-2850.1997.tb00094.x](https://doi.org/10.1111/j.1468-2850.1997.tb00094.x)
- Benjamin, L. (1974). Structural analysis of social behavior. *Psychological Review*, 81, 392-425. [doi:10.1037/h0037024](https://doi.org/10.1037/h0037024)
- Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change*, 217–270. New York: John Wiley & Son
- ‡Beutel, M., Hoflich, A., Kurth, R., & Reimer, Ch. (2005). Who benefits from inpatient short-term psychotherapy in the long run? Patients' evaluations, outpatient after-care and determinants of outcome. *Psychology and Psychotherapy: Theory, Research and Practice*, 78, 219-234. [doi:10.1348/147608305X27665](https://doi.org/10.1348/147608305X27665)
- ‡Callahan, J., Almstrom, C., Swift, J., Borja, S., & Heath, C. (2009). Exploring the contribution of supervisors to intervention outcomes. *Training and Education in Professional Psychology*, 3, 72-77. [doi:10.1348/147608305X27665](https://doi.org/10.1348/147608305X27665)
- Crits-Christoph, P. (1992). The efficacy of brief dynamic psychotherapy: A meta-analysis. *American Journal of Psychiatry*, 149, 151-158.
- Deville, G., & Foa, E., (2001). The investigation of exposure and cognitive therapy: Comment on Tarrier et al. (1999). *Journal of Consulting and Clinical Psychology*, 69, 114-116. [doi:10.1037/0022-006X.69.1.114](https://doi.org/10.1037/0022-006X.69.1.114)

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-based Mental Health*, 1, 70-72.

[doi:10.1136/ebmh.1.3.70](https://doi.org/10.1136/ebmh.1.3.70)

Fairburn, C., & Cooper, Z. (1993). The Eating Disorder Examination. In C. Fairburn & G. Wilson (Eds.), *Binge eating: Nature, assessment and treatment*. 12th edition. New York: Guilford, 317-360.

Fichter, M., & Quadflieg, N. (2001). The structured interview for anorexic and bulimic disorders for DSM-IV and ICD-10 (SIAB-EX): Reliability and validity. *European Psychiatry*, 16, 38-48. [doi:10.1016/S0924-](https://doi.org/10.1016/S0924-9338(00)00534-4)

[9338\(00\)00534-4](https://doi.org/10.1016/S0924-9338(00)00534-4)

Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications Ltd.

‡Ford, J., Fisher, P., & Larson, L. (1997). Object relations as a predictor of treatment outcome with chronic Posttraumatic Stress Disorder. *Journal of Consulting & Clinical Psychology*, 65, 547-559. [doi:10.1037/0022-](https://doi.org/10.1037/0022-006X.65.4.547)

[006X.65.4.547](https://doi.org/10.1037/0022-006X.65.4.547)

=Fournier, J., DeRubeis, R., Shelton, R., Hollon, S., Amsterdam, J., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting & Clinical Psychology*, 77, 775-787. [doi:10.1037/a0015401](https://doi.org/10.1037/a0015401)

[doi:10.1037/a0015401](https://doi.org/10.1037/a0015401)

‡Fricke, S; Moritz, S; Andresen, B; Jacobsen, D., Kloss, M., & Rufer, M., et al. (2006). Do personality disorders predict negative outcome in obsessive-compulsive disorders? A prospective 6-month follow-up study.

European Psychiatry, 21, 319-324. [doi:10.1016/j.eurpsy.2005.03.010](https://doi.org/10.1016/j.eurpsy.2005.03.010)

- Goodman, W., Price, L., Rasmussen, S., Mazure, C., Fleisschman, R., Hill, C., et al. (1989). The Yale-Brown Obsessive-Compulsive Scale: Part I. Development, use and reliability. *Archives of General Psychiatry*, *46*, 1006-1011.
- Harmon, S., Lambert, M., Smart, D. Hawkins, E., Nielsen, S., Slade, K., & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist-client feedback and clinical support tools. *Psychotherapy Research*, *17*, 379-392. [doi:10.1080/10503300600702331](https://doi.org/10.1080/10503300600702331)
- ‡Hartmann, A., Orlinsky, D., Weber, S., Sandholtz, A., & Zeeck, A. (2010). Session and intersession experience related to treatment outcome in Bulimia Nervosa. *Psychotherapy*, *47*, 355-370. [doi:10.1037/a0021166](https://doi.org/10.1037/a0021166)
- Hatfield, D., McCullough, L., Frantz, S., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy*, *17*, 25 – 32.
- Hsu, L. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 141-144. [doi:10.1037/0022-006X.63.1.141](https://doi.org/10.1037/0022-006X.63.1.141)
- ‡Ilgen, M., & Moos, R. (2006). Exacerbation of psychiatric symptoms during substance use disorder treatment. *Psychiatric Services*, *57*, 1758-1764. [doi:10.1176/appi.ps.57.12.1758](https://doi.org/10.1176/appi.ps.57.12.1758)
- Jacobson, N., & Revenstorf, D. (1988). Statistics for assessing the clinical significance of psychotherapy techniques: Issues, problems, and new developments. *Behavioral Assessment*, *10*, 133-146.

- Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19. [doi:10.1037/0022-006X.59.1.12](https://doi.org/10.1037/0022-006X.59.1.12)
- =Jensen, H., Mortensen, E., & Lotz, M. (2010). Effectiveness of short-term psychodynamic group therapy in a public outpatient psychotherapy unit. *Nordic Journal of Psychiatry, 64*, 106-114. [doi:10.3109/08039480903443874](https://doi.org/10.3109/08039480903443874)
- =Kanter, J., Landes, S., Busch, A., Rusch, L., Brown, K., & Baruch, D.E., & Holman, G. (2006). The effect of contingent reinforcement on target variables in outpatient psychotherapy for depression: A successful and unsuccessful case using functional analytic psychotherapy. *Journal of Applied Behavior Analysis, 39*, 463-467. [doi:10.1901/jaba.2006.21-06](https://doi.org/10.1901/jaba.2006.21-06)
- ‡Kellett, S., Clarke, S., & Matthews, L. (2007). Delivering group psychoeducational CBT in Primary Care: Comparing outcomes with individual CBT and individual psychodynamic-interpersonal psychotherapy. *British Journal of Clinical Psychology, 46*, 211-222. [doi:10.1348/014466506X146188](https://doi.org/10.1348/014466506X146188)
- Lambert, M., & Ogles, B. (2004). The efficacy and effectiveness of psychotherapy. In M. Lambert (Ed.) *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. (5th ed., 139-193) New York: Wiley.
- Larsen, D., Attkisson, C., Hargreaves, W., & Nguyen, T. (1979). Assessment of client/ patient satisfaction: Development of a general scale. *Evaluation and Program Planning, 2*, 197-207.

- ‡Lincoln, T., Rief, W., Hahlweg, K., Frank, M., Witzleben, I., Schroeder, & Fiegenbaum, W. (2005). Who comes, who stays, who profits? Predicting refusal, dropout, success, and relapse in a short intervention for social phobia. *Psychotherapy Research*, 15, 210-225.
[doi:10.1348/147608309X464422](https://doi.org/10.1348/147608309X464422)
- ‡Lindgren, A., Werbart, A., & Philips, B. (2010). Long-term outcome and post-treatment effects of psychoanalytic psychotherapy with young adults. *Psychotherapy and Psychotherapy: Theory, Research and Practice*, 83, 27-43. [doi:10.1348/147608309X464422](https://doi.org/10.1348/147608309X464422)
- ‡Lopez-Goni, J., Fernandez-Montalvo, J., Menendez, J., Yudego, F., Garcia, A., & Esarte, S. (2010). Group and individual change in the treatment of drug addictions: A follow-up study in therapeutic communities. *Spanish Journal of Psychology*, 13, 906-913.
- ‡Lunnen, K., Ogles, B., & Pappas, L. (2008). A multiperspective comparison of satisfaction, symptomatic change, perceived change, and end-point functioning. *Professional Psychology: Research & Practice*, 39, 145-152. [doi:10.1037/0735-7028.39.2.145](https://doi.org/10.1037/0735-7028.39.2.145)
- Makela, K. (2004). Studies of the reliability and validity of the Addiction Severity Index. *Addiction*, 99, 398-410. [doi:10.1111/j.1360-0443.2003.00665.x](https://doi.org/10.1111/j.1360-0443.2003.00665.x)
- Mohr, D. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice*, 2, 1-27. [doi:10.1111/j.1468-2850.1995.tb00022.x](https://doi.org/10.1111/j.1468-2850.1995.tb00022.x)
- Mohr, D., Beutler, L., Engle, D., Shoham-Salomon, V., Bergan, J., & Kaszniak, A. et al. (1990). Identification of patients at risk for nonresponse

- and negative outcome in psychotherapy *Journal of Consulting and Clinical Psychology*, 58, 622-628. [doi:10.1037/0022-006X.58.5.622](https://doi.org/10.1037/0022-006X.58.5.622)
- =Moos, R., Moos, B., & Finney, J. (2001). Predictors of deterioration among patients with substance-use disorders. *Journal of Clinical Psychology*, 57, 1403-1419. [doi:10.1002/jclp.1105](https://doi.org/10.1002/jclp.1105)
- =Moos, R., Nichol, A., & Moos, B. (2002). Risk factors for symptom exacerbation among treated patients with substance use disorders. *Addiction*, 97, 75-83. [doi:10.1046/j.1360-0443.2002.00063.x](https://doi.org/10.1046/j.1360-0443.2002.00063.x)
- =Moritz, S., Fricke, S., Jacobsen, D., Kloss, M., Wein, C., Rufer, M., et al. (2004). Positive schizotypal symptoms predict treatment outcome in Obsessive-Compulsive Disorder. *Behaviour Research and Therapy*, 42, 217-227. [doi:10.1016/S0005-7967\(03\)00120-7](https://doi.org/10.1016/S0005-7967(03)00120-7)
- National IAPT Programme Team, (2011). The IAPT Data Handbook: Guidance on recording and monitoring outcomes to support local evidence-based practice. Version 2.0. <http://www.iapt.nhs.uk/silo/files/the-iapt-data-handbook.pdf>
- ‡Ogles, B., Lambert, M., & Sawyer, J. (1995). Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research program data, *Journal of Consulting & Clinical Psychology*, 63, 321-326. [doi:10.1037/0022-006X.63.2.321](https://doi.org/10.1037/0022-006X.63.2.321)
- Orlinsky, D., Ronnestad, M., & Willutski, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.) *Handbook of psychotherapy and behaviour change* (5th Ed.). New York: John Wiley & Sons

- ‡Pekarik, G., & Guidry, L. (1999). Relationship of satisfaction to symptoms change, follow-up adjustment, and clinical significance in private practice. *Professional Psychology: Research and Practice*, 30, 474-478.
[doi:10.1037/0735-7028.30.5.474](https://doi.org/10.1037/0735-7028.30.5.474)
- ‡Pekarik, G., & Wolff, C. (1996). Relationship of satisfaction to symptom change, follow-up adjustment, and clinical significance. *Professional Psychology: Research and Practice*, 27, 202-208. [doi:10.1037/0735-7028.27.2.202](https://doi.org/10.1037/0735-7028.27.2.202)
- =Pike, K., Walsh, B., Vitousek, K., Wilson, G., & Bauer, J. (2003). Cognitive behavior therapy in the posthospitalization treatment of anorexia nervosa. *The American Journal of Psychiatry*, 160, 2046-2049.
[doi:10.1176/appi.ajp.160.11.2046](https://doi.org/10.1176/appi.ajp.160.11.2046)
- ‡Prestano, C., Lo Coco, G., Gullo, S., Lo Verso, G. (2008). Group analytic therapy for eating disorders: Preliminary results in a single-group study. *European Eating Disorders Review*, 16, 302-310. [doi:10.1002/erv.871](https://doi.org/10.1002/erv.871)
- =Rufer, M., Fricke, S., Moritz, S., Kloss, M., & Hand I. (2006). Symptom dimensions in obsessive-compulsive disorder: Prediction of cognitive-behavior therapy outcome. *Acta Psychiatrica Scandinavia*, 113, 440-446.
[doi:10.1111/j.1600-0447.2005.00682.x](https://doi.org/10.1111/j.1600-0447.2005.00682.x)
- Safer, D., Robinson, A., & Jo, B. (2010). Outcome from a randomised controlled trial of group therapy for binge eating disorder: Comparing Dialectical Behaviour Therapy adapted for binge eating to an active comparison group therapy. *Behavior Therapy*, 41, 106-120.
[doi:10.1016/j.beth.2009.01.006](https://doi.org/10.1016/j.beth.2009.01.006)

- =Samstag, L., Muran, J., Wachtel, A., Slade, A., Safran, J. & Winston, A. (2008). Evaluating negative process: A comparison of working alliance, interpersonal behavior, and narrative coherency among three psychotherapy outcome conditions. *American Journal of Psychotherapy*, 62, 165-194.
- =Samstag, L., Batchelder, S., Muran, J., Safran, J., & Winston, A. (1998). Early identification of treatment failures in short-term psychotherapy: An assessment of therapeutic alliance interpersonal behaviour. *Journal of Psychotherapy: Practice and Research*, 7, 126- 143.
- =Scogin, F., Floyd, M., Jamison, C., Ackerson, J., Landreville, P. & Bissonnette, L. (1996). Negative outcomes: What is the evidence on self-administered treatments? *Journal of Consulting & Clinical Psychology*, 64, 1086-1089. [doi:10.1037/0022-006X.64.5.1086](https://doi.org/10.1037/0022-006X.64.5.1086)
- Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin*, 92, 581-604. [doi:10.1037/0033-2909.92.3.581](https://doi.org/10.1037/0033-2909.92.3.581)
- Shout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Strupp, H., Fox, R., & Lessler, K. (1969). *Patients view their psychotherapy*. Baltimore: Johns Hopkins University Press.
- Strupp, H., Hadley, S., & Gomes-Schwartz, B. (1977). *Psychotherapy for better or worse: The problem of negative effects*. New York: Aronson.
- ‡Tarrier, N., Pilgrim, H., Sommerfield, C., Faragher, B., Reynolds, M., & Graham, E. et al. (1999). A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress

disorder. *Journal of Consulting and Clinical Psychology*, 67, 13-18.

[doi:10.1037/0022-006X.67.1.13](https://doi.org/10.1037/0022-006X.67.1.13)

‡Von Der Lippe, A.L., Monsen, J.T., Ronnestad, M.H., & Eilertsen, D.E.

(2008). Treatment failure in psychotherapy: The pull of hostility.

Psychotherapy Research, 18, 420-432. [doi:10.1080/10503300701810793](https://doi.org/10.1080/10503300701810793)

EMPIRICAL REPORT:

**Identifying and Predicting Deterioration and Sudden Deterioration
among IAPT clients in routine primary care settings**

Abstract

The intention of this research was to investigate the phenomena of *overall deterioration* and *sudden deterioration* in a routinely collected data set collected over 18 months within the Improving Access to Psychological Therapies (IAPT) initiative. The phenomenon of sudden deterioration was explored to determine whether it existed and how it may be defined. The rates of sudden deterioration and overall deterioration within the IAPT data for low intensity interventions, high intensity CBT, and high intensity Counselling were identified, and predictors of these were assessed.

It was determined that an appropriate definition for sudden deterioration was a reliable between session change of 6 points using the Patient Health Questionnaire–9; PHQ-9), that was not allied to sudden gains. Rates of sudden deterioration and overall deterioration were found to be 3.4% and 3.1% respectively. There were no significant differences between the three treatment levels. Predictors found for overall deterioration were initial employment status, drop out, initial PHQ-9 score below the severe range, and if sudden deterioration was present. Predictors for sudden deterioration were the presence of overall deterioration and initial PHQ-9 score below the severe range. The results are considered in relation to the existing literature. It was concluded that sudden deterioration does exist as a phenomenon and that it is closely related to overall deterioration. Rates of deterioration in the IAPT dataset were relatively low, and there were no differences in the rates of deterioration between low-intensity interventions, high-intensity CBT, and counselling.

Introduction

One of the central aims of any psychotherapeutic intervention is to reduce an individual's psychological distress. Unfortunately this outcome is not always achieved and some clients will not improve despite being in receipt of an approved and appropriate psychological intervention (Lambert & Ogles, 2004). Further, a proportion of clients may in fact deteriorate, as measured by a standard outcome measure, and there is evidence to suggest that between 5-10% of clients will deteriorate while in receipt of a psychological intervention (Mohr, 1995). In some cases deterioration has been shown to be related to the treatment provided, and as such there is a small body of research focusing on 'psychological treatments that cause harm' (Lilienfeld, 2007).

Research into the area of deterioration is notoriously difficult for a number of reasons. These reasons include locating the direct cause of deterioration, understanding the impact of deterioration on family systems, and ethical issues of continuing interventions that may be causing deterioration (Mohr, 1995). A further difficulty has been that some researchers have combined clients who deteriorate with clients who do not respond to a psychological intervention – that is, non-responders (e.g., Fricke et al., 2006; Hartman et al., 2010). However, there is not always a linear relationship between clients who improve, those who remain unchanged, or those who deteriorate. Indeed, Mohr (1990) reported that in some circumstances the relationship can be curvilinear. He provided an example of this relationship when reporting that levels of interpersonal distress were high in both negative and positive responders, but low in those clients who did not respond to therapy.

Hence, the three client groups – improved, unchanged, or deteriorated – need to be conceptualised and assessed separately.

The body of evidence relating to negative effects from therapy is still relatively sparse, although there has recently been an increased awareness and interest in this area (e.g., Barlow, 2010; Boisvert, 2010; Swift et al., 2010). For example, methods have been put forward to detect harm experienced by clients that can be attributable to specific psychotherapies (Dimidjian & Hollon, 2010; Lambert, 2007, Lilienfeld, 2007). However, it has been demonstrated that a significant number of practitioners can underestimate the occurrence of negative effects or not to realise that such effects exist (Boisvert & Faust, 2006). Therefore it is important that both clinicians and researchers acknowledge the importance and implications that negative effects of psychological treatments may have on clients. Accordingly, the present study aimed to investigate the phenomenon of deterioration within a specific service delivery model of psychological therapies.

Overall Deterioration, Sudden Deterioration (and Sudden Gains)

In considering the phenomenon of *deterioration*, two distinct classes of events, and hence definitions, can be identified. The first definition considers the *overall deterioration* of a client as an outcome determined at the end of the course of therapy and operationalised by the change in outcome score from pre- to post-therapy. Hence, this definition relates to the end-state for a client and equates with the traditional view of deterioration in which a client is

worse off after therapy as compared with their initial status. Methodologically, this class of deterioration requires only two data points and therefore places minimal administrative demands on the client and service. However, collection of Time 2 (i.e., end of treatment) data is often problematic and without it, a change score— and hence improvement or deterioration – cannot be determined. It is more likely that it is clients who experience positive outcomes of therapy who will complete the treatment course. Hence, the traditional pre-post measurement approach is likely to mask detection of deterioration rates. However, the collection of session-by-session measures would remedy this issue by ensuring there was always a measure from the most recent attended session.

The second definition of deterioration builds on this advancement of session-by-session measures and takes account of a *sudden deterioration* that occurs during the course of therapy from one session to the next. The concept of sudden deterioration is the reverse of *sudden gains*, a phenomenon first reported by Tang and DeRubeis (1999), which refers to a significant and measurable improvement of symptoms for a client in between session intervals. Subsequent work has been carried out to investigate the phenomenon of sudden gains and to determine how likely such gains are to be maintained (e.g., Tang et al., 2007). There is currently no evidence as to whether the opposite phenomenon, sudden deterioration, exists other than as a reversal of sudden gains (Manning, Hardy, & Kellett, 2010). The lack of investigation into sudden deterioration could be part of the lack of attention paid to deterioration generally and unconnected to the importance of the

issue. Hence exploration of this potential phenomenon could be beneficial to psychotherapy process literature.

Accordingly, although the phenomenon of sudden deterioration has not been directly investigated previously, the two related areas of research into deterioration in psychotherapy, and sudden gains in therapy provide a base for this work. Literature regarding deterioration generally within psychotherapy has developed a broadly standard agreement regarding the need for at least a reliably valid change in pre- to post-therapy score such as indicated by the reliable change index (RCI) or the concept of clinically significant change as applied to deterioration (Ford, Fisher & Larson, 1997; Kellett, Clarke & Matthews, 2007; Lunnen, Ogles & Pappas, 2008). These concepts are described below.

The RCI represents the change needed to occur to ensure change is not due to imprecise measurement of the outcome measure through a calculation proposed by Jacobson and Truax (1991), which is based on the standard error of difference between pre-treatment and post-treatment scores. For change to also be clinically significant, the score must also move into the functional distribution by crossing a threshold or 'cut off' score. In relation to the PHQ-9, the RCI has been established as a score of 6 points and the threshold for moving to normal functioning being a score of 9 or below (see Parry et al., 2011).

Adoption of criteria for sudden deterioration is informed by the research into sudden gains in therapy, which has utilised the classification described by Tang and DeRubeis (1999). These criteria are (a) a difference between the

pre-gain and after-gain scores of at least the RCI for the outcome measure used (b) a nominally significant t -value comparing the three pre-gain scores with the three after-gain scores and (c) a nominally significant t -value if only two pre-gain or two after-gain scores were available. Therefore a minimum of 5 sessions are required in order to apply these criteria. Reversals of sudden gains were defined as a 50% reduction in symptom improvement, following the sudden gain. Therefore the location of the sudden change within treatment, and the length of treatment are central considerations in applying these approaches to the investigation of sudden deterioration.

Predictors of deterioration

The current literature suggests a number of factors that might be expected to have a predictive role in deterioration including age, chronicity (e.g., Fournier et al., 2009), presence of a personality disorder (Reich, 2003), and initial symptom severity (Sotsky et al., 1991). However, symptom severity has also been found to predict positive outcome (Woody et al., 1984). A high level of instability in a person's life context (e.g., being single, or unemployed) has also been shown to potentially predict negative outcome (Moos, 2005). This is particularly important for any intervention or programme aimed at returning clients to work. Mohr's (1995) review suggested that people diagnosed with bi-polar affective disorder, severe interpersonal difficulties, poor motivation, or who expected therapy to be a painless process, were more likely to experience deterioration. He also suggested that self-administered treatments may result in higher rates of deterioration, citing a rate of 19%

rather than the average of between 5-10%. However, this view has been challenged by Scogin et al. (1996), who reported five studies with a combined deterioration rate of 9%.

Lambert has focussed on exploring methods of prevention of deterioration (e.g., Lambert, 2010; Lambert et al., 2005). He has advocated and implemented a system of feedback to therapists based on session-by-session outcome measures. If a client's progress is identified as 'not on track' according to their expected treatment response as derived from the data of similar clients, the therapist is alerted via a feedback report and intervention guidance provided. Research evaluating this form of feedback has demonstrated significant reductions in average deterioration rates as well as in comparison with treatment-as-usual conditions. For example, Lambert et al. (2001) reported that clients who were identified as 'not on track' had a deterioration rate of 23% in the 'treatment-as-usual' condition compared to 6% when feedback was provided.

Data set requirements for research into deterioration

The three broad areas of work outlined above –classes of deterioration (including sudden deterioration), rates of deterioration, and predictors of deterioration – provide the foci for the present study. However, in order to investigate these areas there are specific requirements regarding a suitable data set. First, the data set needs to be sufficiently large in order to yield sufficient data relating to deterioration due to the relatively low baseline level of deterioration. Second, the data set needs to comprise session-by-session

measurement in order to ensure a measurement score subsequent to the pre-treatment score. Thirdly, the data set requires sufficient additional variables in order to test for predictors of deterioration. These three requirements are met in the context of data derived from the Improving Access to Psychological Therapies (IAPT) initiative.

Improving Access to Psychological Therapies Initiative

IAPT is a UK government initiative aimed at increasing the availability of evidence-based treatments for common mental health problems to the general public. Layard (2004) recognised that a considerable social and economic burden was being carried by people experiencing a range of common mental health problems for which effective psychological treatments existed but which were not accessible to those in need. He argued that this issue could, and should, be addressed by the additional training and provision of greater resources for mental health provision (Layard, 2006). This led to the IAPT initiative that initially targeted depression and anxiety disorders and that primarily focused on extending the provision of CBT.

The IAPT initiative defined two levels of psychological therapists with each delivering defined interventions, which allowed them to be accommodated by the existing NHS Stepped Care model for mental health services (Bower & Gilbody, 2005.) These two levels are: low intensity workers, subsequently relabelled Psychological Wellbeing Practitioners (PWPs), and high intensity workers who delivered cognitive-behavioural interventions. The majority of clients receive low intensity treatment provided by PWPs who facilitate large-

scale interventions such as psycho-educational groups, computerised CBT (cCBT), and guided self-help. In high intensity treatment, intended for those with moderate difficulties or posttraumatic stress disorder (PTSD), clients are allocated to a therapist who will provide a National Institute for Clinical Excellence (NICE) guideline-approved psychological treatment, such as CBT for depression, CBT for PTSD, couples therapy, or interpersonal therapy. Counselling is also provided through IAPT in some areas as a high intensity treatment and is recommended as an appropriate treatment for depression but not for anxiety disorders. A client can also be 'stepped up' to a higher level of treatment, such as specialist psychotherapy or community mental health teams, if the initial intervention is not yielding progress.

The IAPT initiative has received some criticism including concerns about the cost of such a wide spread change to mental health provision. It has been referred to by Cromby et al. (2008) as an 'expensive experiment' to test the 'Layard hypothesis that large scale expansion of evidence-based psychological therapies will increase both the happiness and productivity of the population.' Cromby et al. (2008) argued that this is not as cohesive an initiative as suggested by initial reports, citing disagreement over the diagnosis of 'depression', the type of treatments to be offered, and the efficacy of these treatments (Richards & Suckling, 2008). Despite these concerns, initial reports from IAPT initiatives around the UK show overall improvement in outcome measure scores as well as reductions in the number of clients who were unable to work due to mental health problems (Clark et al., 2009; Glover, Webb, & Evison, 2010). However deterioration rates were not included in these reports.

The Present Study

It is important that with any new form of psychological intervention or initiative, the effects of deterioration should be explored and evaluated alongside studying effectiveness. Accordingly, the present study used data that had been routinely collected from the Sheffield IAPT programme during an 18-month period of the service in order to investigate the phenomenon of deterioration. This involved a large number of clients being treated, and so provided an extensive dataset meeting the three data set requirements set out earlier.

A specific feature of the IAPT service within Sheffield was that it included a defined counselling service located within the high-intensity service. Hence, the existence of counselling as well as low- and high-intensity CBT, enabled comparison of deterioration rates between intervention levels (i.e., low versus high) as well as between differing therapeutic modalities (i.e., CBT versus counselling). As stated earlier, the IAPT data set required session-by-session outcome measures, thereby enabling determination of rates of *sudden deterioration* as well as *overall deterioration*. It also aimed to identify potential predictors of both sudden and overall deterioration. Accordingly, the specific aims of the study were three-fold:

1. To identify and describe methods of identifying *sudden deterioration*.
2. To determine the rates of *overall deterioration* and *sudden deterioration* and assess whether there are differential rates of deterioration within the three treatment levels (low intensity, high intensity and counselling)

3. To test for predictors of *overall deterioration* – and *sudden deterioration* if present – in any of the treatment levels, at either individual or service levels

Method

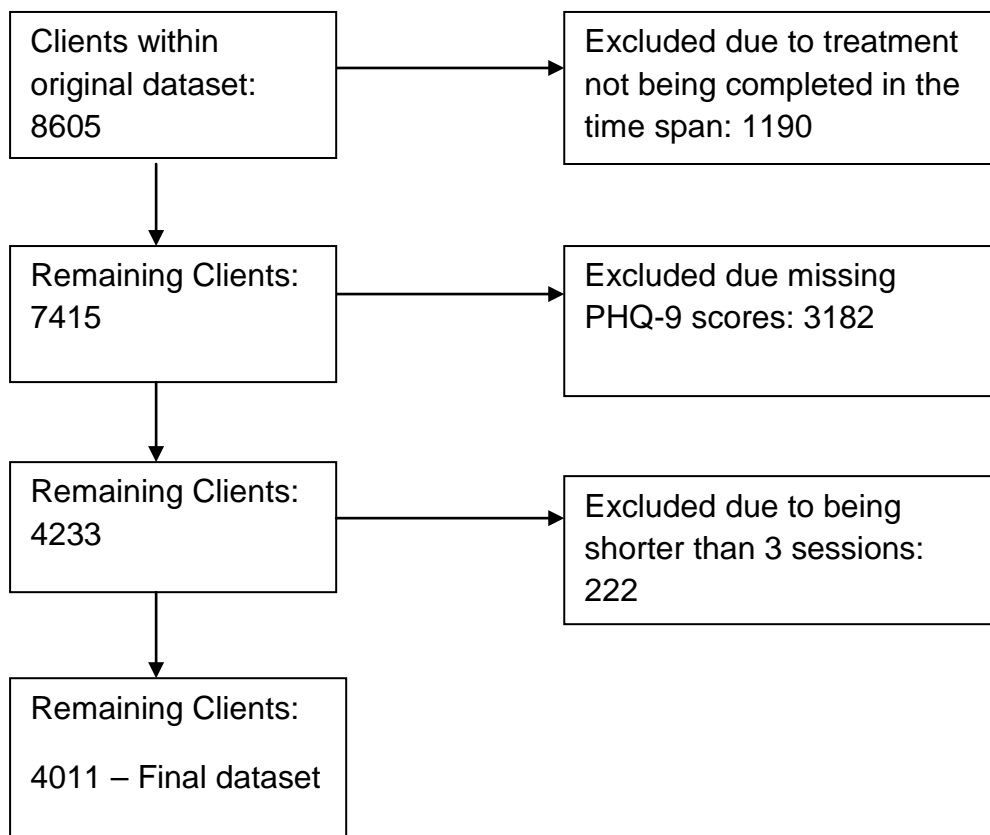
Total datapool, selected sample, and design

The overarching data set comprised an anonymised and routinely collected dataset administered within the Sheffield IAPT service and collected between March 2009 and September 2010. Five work-streams were planned for this dataset comprising investigation into sudden gains, therapist effects, dose-effect relations, benchmarking and stepping up and stepped care, all with the central aim of investigating differences between the three levels of treatment. The author (LT) was invited to join this research team to investigate deterioration effects as this was not present in the original design, but seen to be related and beneficial.

Demographic data was collected in line with routine practice within the NHS, along with the session-by-session outcome measures described below. The original sample of cases comprised 8605 clients. However, 1190 of these clients had either not completed their treatment within the time span of the dataset (i.e. between 1st March 2009 and 30th September 2010) or began their treatment beforehand, and so their data were immediately discarded. Therefore the initial dataset comprised 7415 clients with a gender split of 4775 (64.4%) female and 2640 (35.6%) males and an overall mean age of 34.94 (SD = 14.58) with a range from 16 years to 89 years old. The rates of diagnosis for mental disorder, not otherwise specified, was 3777 (50.9%), mixed anxiety and depressive disorder was 1589 (21.4%), depression was 711 (9.6%) and anxiety was 398 (5.4%).

From the original data pool, a subset was selected for the present study according to specific criteria. First, data were selected in which clients had completed their treatment. Applying this criterion reduced the original data pool from 8605 clients to 7415. However, PHQ-9 scores for each session were available for only 4233 clients. Clients who received less than three sessions (including assessment) were also excluded in order to ensure the potential for a change score between treatment sessions. This resulted in a total of 4011 clients in the dataset. Figure 1 presents a CONSORT diagram of the client data included in the study. All subsequent reporting refers to this data set (N = 4011).

Figure 1 – Diagram of Client data included in this study



Within this client cohort (N = 4011), a between-subjects design was employed wherein clients meeting specified criteria of deterioration (overall or sudden) were contrasted with clients not meeting these criteria. These two categories were utilised as the independent variables, with treatment levels and all other potential predictors as dependent variables.

Study Participants

The mean age of participants was 41 years (SD = 14.4 yrs) with a range from 16 years to 89 years old. The gender split was 35.1% male (n = 1415) and 64.3% female (n = 2596). Table 1 contains the demographic information for the 4011 participants, including details relating to ethnicity, employment, and diagnosis.

Table 1: Demographic Information for Participants

Main grouping	Sub-grouping	N	%
Ethnicity			
	White British	3075	76.7
	Unknown	528	13.2
	Asian/Asian British	162	4.0
	Black/Black British	59	1.5
	White other	62	1.5
	Dual heritage	52	1.3
	Other	70	1.7
Employment			
	Employed full-time	1410	34.9
	Unemployed	982	24.3
	Employed part-time	667	16.5
	Retired	360	8.9
	Student	294	7.3
	Homemaker/carer	292	7.2
Diagnosis			
	Mental disorder not otherwise specified	1978	49.0
	Mixed anxiety and depressive disorder	892	22.1
	Depression	397	10.3
	Anxiety	262	6.8
	Adjustment disorders	106	2.6
	PTSD	55	1.4
	Other	180	4.7

Outcome Measures

A battery of outcome measures was collected by the service at pre- and post-therapy and also on a session-by-session basis. The following two outcome measures were administered prior to every session.

The Patient Health Questionnaire-9 (PHQ-9; see Appendix E) is a 9-item self-report outcome measure for screening for depression (Arroll et al., 2010; Kroenke, Spitzer & Williams, 2001). Validity has been assessed against an

independent structured mental health professional (MHP) interview. Test-retest reliability for the PHQ-9 has been reported as 0.89. A PHQ-9 score of 10 or above had a sensitivity of 88% and a specificity of 88% for major depression (Kroenke, Spitzer & Williams, 2001). Scores of 5, 10, 15, and 20 represent thresholds for mild, moderate, moderately severe and severe depression, respectively. The PHQ-9 was employed as the primary outcome measure.

The Generalised Anxiety Disorder-7 (GAD-7; see Appendix F) is a 7-item self-report outcome measure that screens for anxiety disorders (Lowe et al., 2008). GAD-7 total score for the seven items ranges from 0 to 21. Scores of 5, 10, and 15 represent thresholds for mild, moderate, and severe anxiety, respectively.

The following outcome measures were administered at pre-treatment and post-treatment only.

The Work and Social Adjustment Scale (WSAS; see Appendix G) is a 5-item self-report scale of functional impairment due to an identified problem. Internal scale consistency has been found to be between 0.70 and 0.94 with a test—retest correlation of 0.73 (Mundt et al., 2002). Scores of 10 and 20 represent thresholds for moderate and severe functional impairment.

Phobia Questions (see Appendix H.) These questions have no standardised reliability or validity, but were administered as part of the IAPT initiative. There were three phobia questions, relating to social phobia, agoraphobia

and specific phobias. As such these scores have been included as a potential variable within the analyses. However the generalisability is significantly limited.

Service provision

The service provision descriptors for the 4011 clients included in the sample are presented in Table 2. The majority were in receipt of low-intensity interventions (N = 2565; 63.6%) and broadly equal numbers received high-intensity CBT (N = 694; 17.2%) or counselling (N = 752; 18.6%). The mean PHQ-9 score at intake for the sample was 14.1 (SD = 6.4).

Table 2: Service parameters

Main grouping	Sub-grouping	N	%
Referral source			
	GP	3885	96.3
	Voluntary sector	41	1.0
	Self	13	0.3
	Other	72	1.8
Treatment level			
	Low-intensity	2565	63.6
	High-intensity CBT	694	17.2
	High-intensity counselling	752	18.6
Intervention provided			
	Guided self-help	2182	54.1
	Counselling	735	18.2
	CBT	686	17.0
	Behavioural activation	109	2.7
	cCBT	70	1.7
	Other	229	5.7
Reason for ending intervention			
	Completed intervention	2406	59.6
	Dropped out	874	21.7
	Declined treatment	303	7.5
	Not suitable for service	168	4.2
	Stepped up (level 4)	156	3.9
	Unknown	102	2.5
	Deceased	2	0.0

Data cleaning

In a process that is common when working with data collected through routine practice, the data were cleaned prior to analyses. This involved a line-by-line checking process of all data points, searching for additional variables, and correcting any errors or duplications. This process was

completed in conjunction with a colleague (Amy Ashworth) who performed this procedure on the dataset for all work-streams. Data cleaning that focused on the specific requirements for the present research, such as the required criterion of all session-by-session PHQ-9 scores, was undertaken by the author (LT).

Procedures

Ethics and governance approval

Ethical approval was granted by Sheffield Research Ethics Committee (Reference number 10/H1310/69; see Appendix C) in relation to analyses on the Sheffield cohort data and the specific focus for research on deterioration was provided by a substantial amendment (see Appendix D). Research Governance was gained from the Research Development Unit of the Sheffield Health & Social Care NHS Foundation Trust.

Service User Involvement

Discussions were held with a service user employed within the Centre of Psychological Services Research (CPSR) and associated with the wider project.

Statistical analyses in relation to specific study aims

Aim 1: This study aimed to determine appropriate ways of calculating sudden deterioration. The literature was searched to determine different ways of calculating deterioration, and by inverting methods used to determine how sudden improvement is measured.

Aim 2: Analysis examined whether there were any significant differences in the proportion of clients who deteriorate or suddenly deteriorate on the PHQ-9 in the different treatment levels (low intensity, high intensity, and counselling) using the chi-squared statistic. The three treatment levels will be used as independent variables, with incidence of deterioration or sudden deterioration on the PHQ-9 as the dependent variables. Rates were also examined within several contexts of sudden deterioration (at the end of treatment, and treatment length).

Aim 3: Backwards logistical regressions (Field, 2009) were performed to determine whether there are demographic predictors that indicate which clients are more likely to deteriorate during IAPT interventions, or service level issues that may be iatrogenic within the three treatment levels. The dependent variable used will be whether reliable deterioration was present or not. From reviewing the literature, there are a large number of independent variables that have the potential to be associated with deterioration. However using a large number of variables can lead to an increased possibility of

Type I errors (Forsythe, May & Engelman, 1971). Therefore preliminary analyses appropriate for each variable were performed (that is chi-squared for categorical data, *t*-tests for continuous variables with a normal distribution, and Mann-Whitney test for continuous variables without a normal distribution) to allow the regression to focus on those most likely to be significant. Variables that were anticipated from the literature to be potentially significant as predictors were as follows: the Index of Multiple Deprivation (IMD), age, gender, ethnicity, employment status, initial GAD-7 score, initial PHQ-9 score, and initial Work and Social Adjustment Scale (WSAS) score.

Results

The results are presented in two sections. The first section addresses the development of methods for calculating sudden gains (Aim 1) together with the rates of sudden and overall deterioration (Aim 2). These two aims are presented together as identification of any method, by definition, yields a rate of deterioration. The second section addresses the results relating to predictors of deterioration.

Sudden and overall deterioration: Methods and rates of occurrence

There is no standard method of defining sudden deterioration, therefore this section of results will present initial rates of sudden deterioration yielded through the use of three different potential definitions derived from the overall deterioration and sudden gains literatures.

Methods

The reliable change index was calculated using the standard formula as follows:

$$[(\text{Pre SD}) * \text{sqrt } 2 * (\text{sqrt } (1-\text{rel}))] * 1.96.$$

The Pre-therapy standard deviation (SD) refers to the standard deviation of initial PHQ-9 scores for this dataset which was 6.43. 'rel' refers to the reliability of the measure, which has a reported test-retest reliability of 0.89.

Therefore:

$$[6.43 * 1.41 * 0.33] * 1.96 = 5.86.$$

Hence, calculation of the RCI for the PHQ-9 using this sample yielded a session-by-session increase in PHQ-9 score of 6 points or more.

Establishing the occurrence of sudden deterioration

The criterion of a rise of 6 points or more (the RCI) between sessions on the PHQ-9 was applied to the data to determine the rate of occurrence of *sudden deterioration*. These rates are reported at two levels: (a) events level in which the rate refers to the number of occurrences where a deterioration of 6 points or more occurred on the PHQ-9; and (b) between-session intervals in which the occurrence of a deterioration of 6 points or more was calculated as a proportion of the total number of between-session intervals. The most basic level of *sudden deterioration* with no other criteria applied (an increase of 6 points) yielded rates of 16.5% and 16.6% as shown in Table 3 (row 1).

Table 3 – Rates of sudden deterioration using three potential definitions.

	Rates of sudden deterioration	
	Event level (N = 4110)	Between session interval level (N = 5358)
Reliable & Sudden Deterioration (i.e., meeting an increase of RCI =6 criterion only)	678 (16.5%)	891 (16.6%)
Reliable & Sudden Deterioration not allied to sudden gains (i.e., meeting criteria of RCI =6 and not allied with sudden gains)	136 (3.3%)	141 (2.6%)
Clinically Significant, Reliable & Sudden Deterioration not allied to sudden gains (i.e., meeting criteria of RCI=6, crossing threshold into clinical range=10+ and not allied with sudden gains)	72 (1.8%)	75 (1.4%)

This procedure does not, however, control for the internal variability within therapeutic change intended in the other sudden gains criteria stated above. Hence another method for controlling for variability that allowed for the inclusion of short treatments was necessary. It was considered that if throughout a treatment there was a sudden deterioration of a session-by-session increase in score of at least 6, but there was not also a sudden gain of a session-by-session decrease in score of at least 6, this may represent a

statistically significant sudden deterioration that is not likely to be as a result of variability. The rates of sudden deterioration for this definition were 3.3% and 2.6% and are presented in Table 3 (row 2).

A related concept that was considered to potentially impact on defining sudden deterioration was that of clinically significant deterioration. This definition requires the change score to be both reliable (i.e., 6 or more for PHQ-9) and to cross the threshold from the functional range to the clinical range (i.e., from a score of 9 or below to 10 or above⁸). This was considered useful in addition to the reliable deterioration not being allied to sudden gains definition in order to provide a very stringent definition of sudden deterioration. The rates for this definition were 1.8% and 1.4% (see row 3, Table 3).

Treatment context

It was considered that sudden deterioration may alter as a function of the context regarding both the length of treatment and *when* within the treatment the sudden deterioration occurred. This requires that the whole episode of treatment is taken into account, and so the focus is the event level occurrences of sudden deterioration (i.e., the episodes of treatment where sudden deterioration took place). The data set yielded 678 sudden deterioration events. A focus on the treatment episodes means that as some individuals had more than one episode of treatment, they may be included

⁸ Based on IAPT clinical caseness.

more than once. Table 4 presents rates of sudden deterioration using three different criteria (reliable change only, reliable change not allied to sudden gains, and reliable and clinical significant change) and the contexts within which the sudden deterioration occurred (sudden deterioration within short treatment (4 or fewer sessions), long treatment⁹ (5 or more sessions) and sudden deterioration occurring at the 'end of treatment' (i.e., in the between session interval prior to the final session)).

Table 4—Rates of Sudden Deterioration in Contexts of Treatment (N= 678)

	End of Treatment	4 or fewer sessions	5 or more sessions
Reliable & Sudden Deterioration (i.e., meeting RCI =6 criterion only)	121 (17.8%)	132 (19.5%)	546 (80.5%)
Reliable & Sudden Deterioration not allied to sudden gains (i.e., meeting criteria of RCI =6 and not allied with sudden gains)	45 (6.6%)	56 (8.2%)	80 (11.8%)
Clinically Significant, Reliable & Sudden Deterioration not allied to sudden gains (i.e., meeting criteria of RCI=6, crossing threshold into clinical range=10+ and not allied with sudden gains)	22 (3.2%)	26 (3.8%)	49 (7.2%)

⁹ 5 sessions was chosen to represent longer treatment as this is the minimal number of sessions required in *sudden gains* research.

The percentage deterioration rates were high (80.5%) for longer (5 or more sessions) compared to shorter (19.5%) treatments when considering the first definition (reliable and sudden deterioration). The difference was less pronounced for reliable sudden deterioration not allied to sudden gains, which may suggest that this definition accounts for the increased potential of deterioration that increased numbers of sessions can afford. Hence the latter definition in which sudden deterioration is not allied to sudden gains would be a more appropriate definition for this phenomenon. In addition, this definition ensures that the change is reliable, and accounts for the potential variability within treatment that is equitable with the sudden gains definition. It also allows for the inclusion of shorter (4 or less sessions) treatments. The third, most stringent definition (clinically significant, reliable and sudden deterioration not allied to sudden gains) yielded very low rates in all areas as shown Tables 3 and 4. Although yielding low rates, this may still represent a concept useful in further classification of sudden deterioration.

In conclusion, the definition of sudden deterioration that is considered the most balanced and equitable within related research areas is that of reliable and sudden deterioration that is not allied to sudden gains.

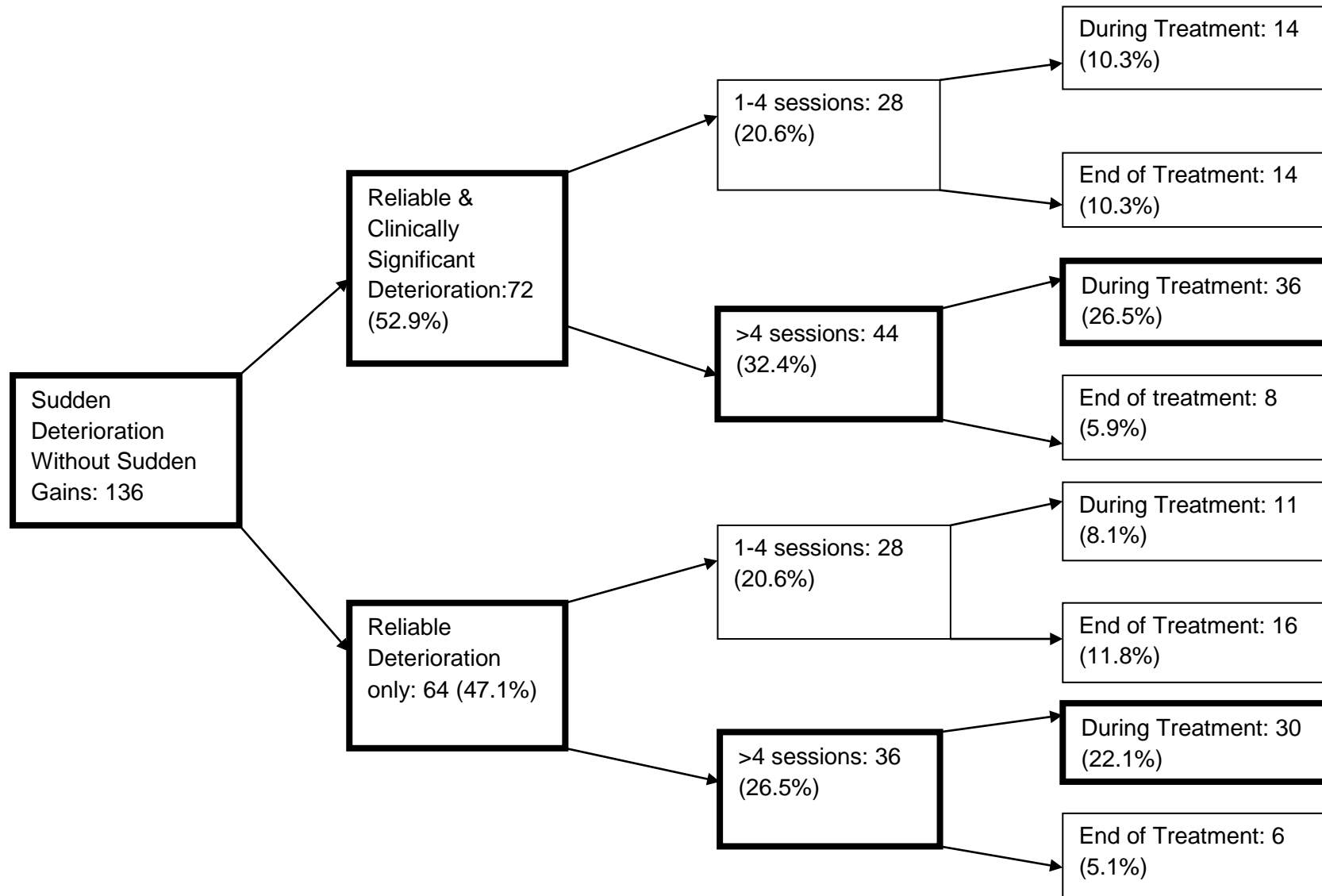
Exploration of Sudden Deterioration

It was considered to be beneficial to explore beyond the rates and definitions of *sudden deterioration* in order to provide some understanding of the event at the level of individual clients. Figure 2 presents a further exploration within the chosen definition of sudden deterioration, (i.e., a reliable negative change

that it is not allied to sudden gains). The flow diagram presents an expression of the flow of individual clients, separated into the classification of clinically significant deterioration or not clinically significant deterioration, and then also the contexts of treatment or when the event of sudden deterioration occurred. It is important to note that those clients whose sudden deterioration was not clinically significant comprise 2 different subgroups. The first are those whose pre-deterioration score was already above the clinical threshold (i.e. 10 for PHQ-9) and so clinically significant change was not possible (N= 18 out of 136, 13.2%). The other subgroup are clients whose pre-deterioration score was low, and even an increase of the reliable change index (i.e. 6 for PHQ-9) did not raise the post-deterioration score above the clinical threshold (N=2 out of 136, 1.5%).

Figure 2 shows that in longer treatments (>4 sessions) sudden deterioration occurred more often during treatment for both reliable and clinically significant and reliable deterioration (26.5% compared to 5.9%, and 22.1% compared to 5.1% respectively). Whereas in short treatments (1-4 sessions) there was less difference between *when* in the treatment the sudden deterioration event occurred. In the subgroup of reliable and clinically significant sudden deterioration the rate was 10.3% for both during and at the end of treatment, and in the reliable sudden deterioration only subgroup the rates were 8.1% and 11.8% respectively.

Figure 2 – Flow Diagram demonstrating classifications and treatment contexts within the sub group of sudden deterioration not allied to sudden gains



Maintenance of sudden deterioration

In addition to the occurrence of sudden deterioration, a further issue concerns whether any such deterioration is maintained. To determine the extent of maintenance of sudden deterioration, a subgroup of clients was constructed comprising only those clients whose treatment was 5 sessions or more, and where the sudden deterioration occurred during treatment. This subgroup comprised reliable and clinically significant sudden deterioration (N= 36), and reliable sudden deterioration only (N= 30), yielding a total subsample comprising 66 clients. The route to identification of this subgroup is highlighted in bold in Figure 2. In order to determine maintenance of deterioration, the concept of overall deterioration was also employed and applied to the final session score for clients. Specifically, the client was deemed to have deteriorated at the end of therapy if their final PHQ-9 score was reliably worse than their initial score –that is, if their end-point PHQ-9 score had deteriorated by 6 points or more when compared with their initial score.

Within the subgroup of 66 clients, 13 (19.7%) clients also met the criterion of overall reliable deterioration, and of these 4 (6.1%) met the criteria for reliable and clinically significant overall deterioration. For these 13 clients, there were 16 occurrences of sudden deterioration with the 3 additional occurrences located in each of the three interventions. The individual therapeutic process and outcomes for these 13 clients are presented in figures 3, 4, 5, 6 and 7, with sudden deterioration events highlighted.

In the low-intensity intervention group, 8 clients experienced sudden deterioration (see Figure 3). For half of these clients (N=4), the sudden deterioration occurred at session 2 (see Figure 4) while for the other 4 clients, it occurred at sessions 3 (3 clients) and 6 (1 client) (see Figure 5)

Figure 3 – Low-intensity (all): Session-by-session PHQ-9 scores for clients who received Low Intensity Treatment within the subgroup of clients who maintained deterioration. Drawn from the sample comprising sudden reliable deterioration, not allied to sudden gains, not at end of treatment, 5+ sessions).

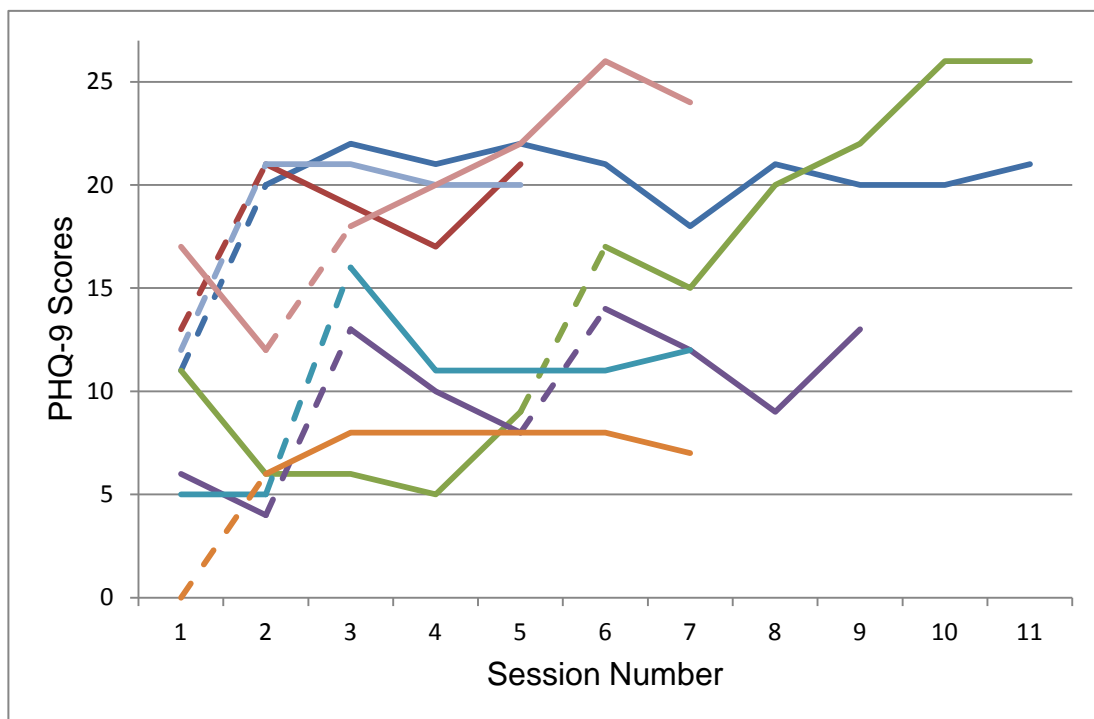


Figure 4 – Low-intensity (sudden deterioration at session 2): Session-by-session PHQ-9 scores for clients who received Low Intensity Treatment, experienced sudden deterioration at session 2, within the subgroup of clients who maintained deterioration (sudden reliable deterioration, not allied to sudden gains, not at end of treatment, 5+ sessions).

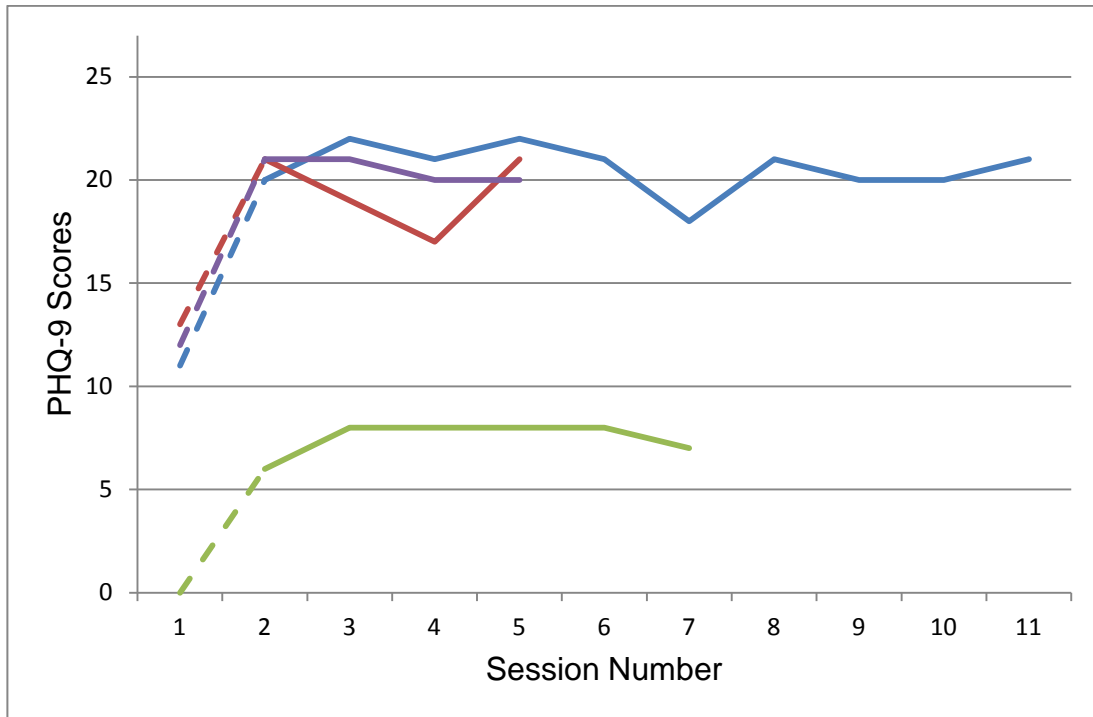


Figure 5 – Low-intensity (sudden deterioration after session 2): Session-by-session PHQ-9 scores for clients who received Low Intensity Treatment, did not experience sudden deterioration at session 2, within the subgroup of clients who maintained deterioration (sudden reliable deterioration, not allied to sudden gains, not at end of treatment, 5+ sessions).

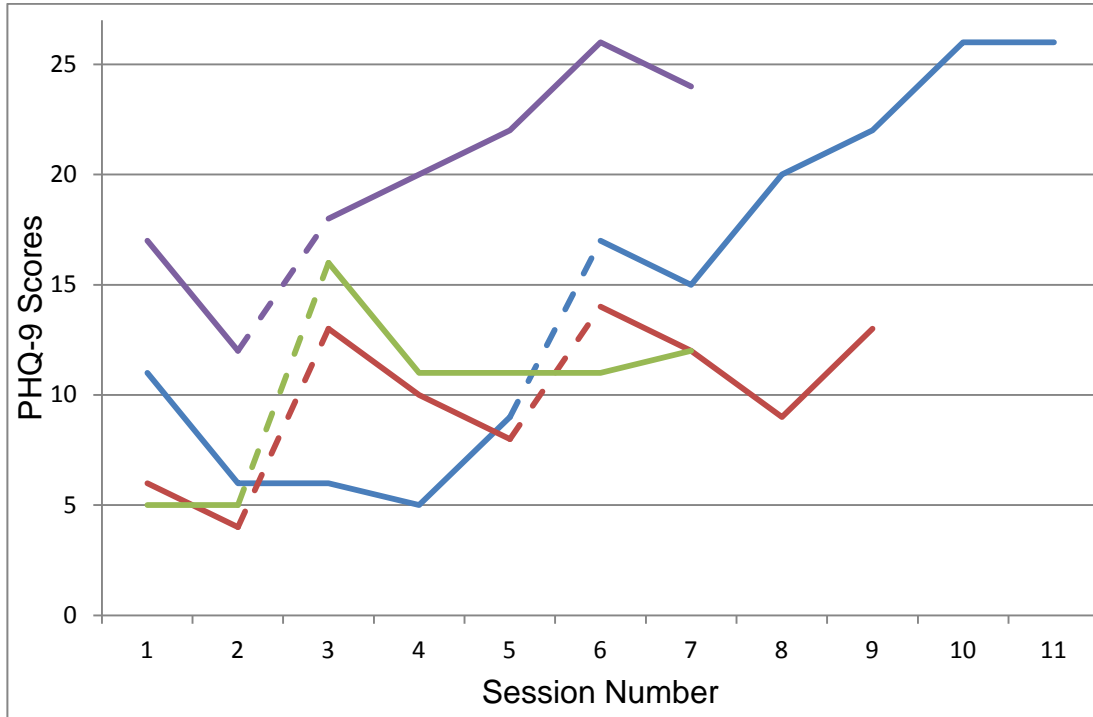


Figure 6 – High-intensity CBT: Session-by-session PHQ-9 scores for clients who received CBT Treatment within the subgroup of clients who maintained deterioration. Drawn from the (sudden reliable deterioration, not allied to sudden gains, not at end of treatment, 5+ sessions).

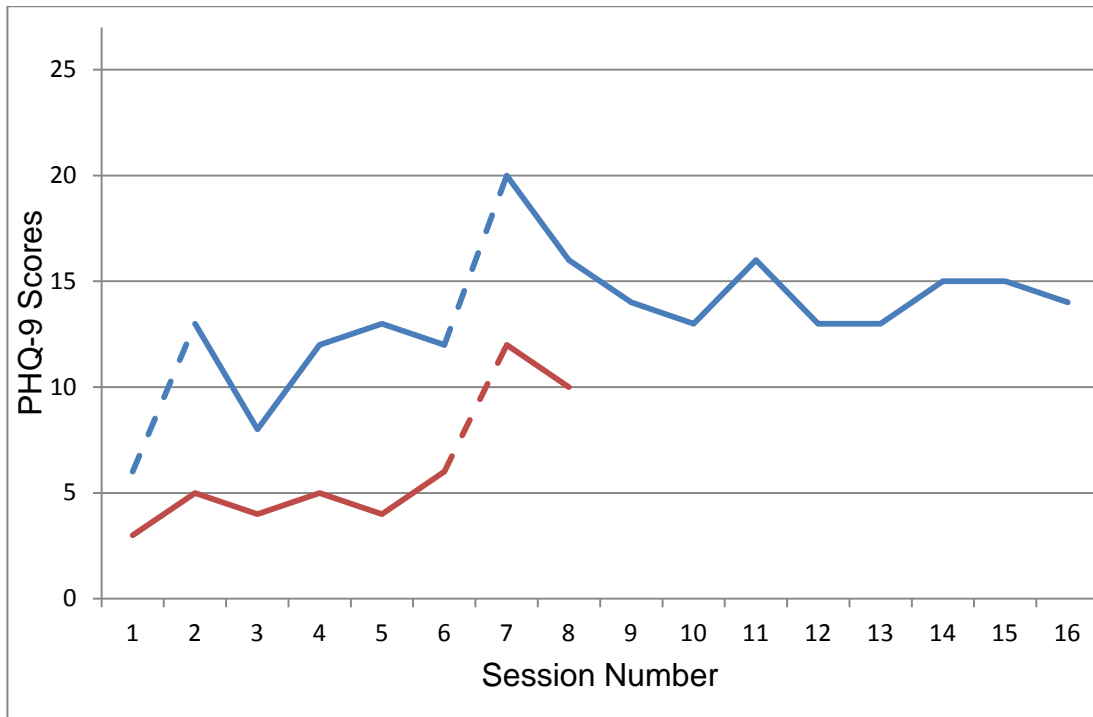
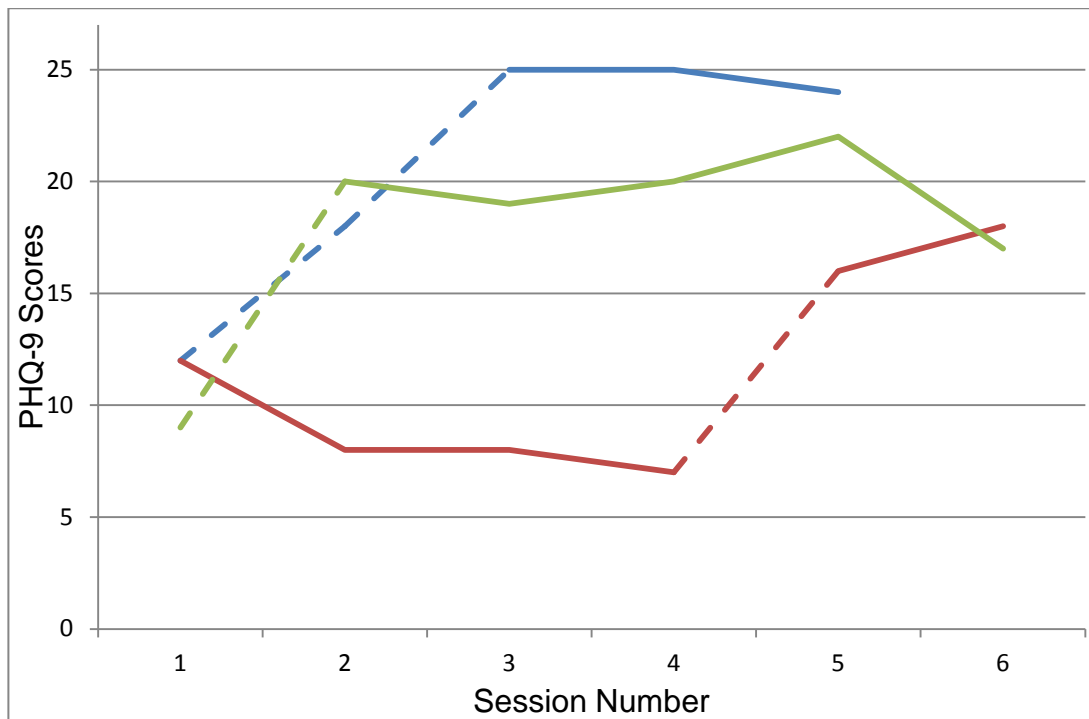


Figure 7 – Counselling: Session-by-session PHQ-9 scores for clients who received Counselling within the subgroup of clients who maintained deterioration. Drawn from the (sudden reliable deterioration, not allied to sudden gains, not at end of treatment, 5+ sessions).



Figures 6 and 7 present the plots for high-intensity CBT and counselling respectively. The occurrence of a sudden deterioration at session 2 occurred for one of the two clients in high-intensity CBT and for 2 of the 3 clients in the counselling group. Hence, for the subgroup of 13 clients, 7 (53.8%) experienced a sudden deterioration at the second session.

Summary

In summary, the phenomenon of sudden deterioration exists within this sample of clients, although similar to overall deterioration, it has a low baseline level and is a rare event. There are a number of different ways in which sudden deterioration could be defined, which have been explored above. It was determined that the most useful was to ensure that the change was reliable according to the reliable change index for the outcome measure employed, and sudden gains were not also present within the treatment episode. In all further analyses in this study, it is this definition of sudden deterioration that has been utilised.

General rates of Sudden Deterioration

In order to determine the rates of sudden deterioration, the definition of sudden deterioration described above was employed, (i.e. reliable change that is not allied to sudden gains). Further to this, within these clients who experienced sudden deterioration, the classifications of overall *reliable improvement*, *no reliable change* and *overall deterioration* were identified along with those incidences where the change was also *clinically significant*, and this information is presented in Table 5. This information is provided for illustration of the dataset, although improvement is not the focus of the study. The comparison yielded rates of 29.0% for overall deterioration and 6.9% for overall improvement.

Table 5 – Rates of overall change throughout treatment within the sub group of sudden deterioration (N=136)

Classification	Number of occurrence (%)	Number of occurrences reaching Clinical Significance (%)
Reliable Deterioration	38 (29.0%)	18 (13.7%)
No change	89 (67.9%)	-
Reliable Improvement	9 (6.9%)	6 (4.6%)

Comparison of the overall rates of change (Table 7) with those above (Table 5) shows there is a much higher rate of overall deterioration within the subgroup of those clients who experienced sudden deterioration, (3.1% and 29% respectively). This suggested that those clients who experience sudden deterioration are more likely to end treatment with deterioration, and less likely to experience overall improvement.

A further aspect of the aim to determine rates of sudden deterioration was to determine if there were any differences in rates of sudden deterioration within the three different treatment levels. These figures are presented in Table 6 below.

Table 6 –Rates of sudden deterioration within three treatment levels

	Low Intensity (N=2565)	High Intensity CBT (N=694)	High Intensity Counselling (N=752)
Sudden Deterioration	88 (3.4%)	27 (3.9%)	21 (2.8%)

The rates of sudden deterioration among the three treatment levels were very similar: 3.4% in Low Intensity, 3.9% in CBT treatment and 2.8% in Counselling. A chi- squared analysis demonstrated that the level of treatment did not significantly impact on the presence of sudden deterioration within treatment ($\chi^2(2) = 1.377, p=0.502$).

The rates of overall deterioration

In order to determine the rates of deterioration, initially the classifications of reliable improvement, no reliable change and deterioration were identified along with those incidences where the change was also clinically significant. *Improvement* and *no change* classifications are also presented. The rate for overall deterioration was 3.1%, with 1.2% reaching clinical significance deterioration. The rate for reliable improvement yielded was 43.1%, and 31.2% for clinically significant improvement. Table 7 presents the rates for all 3 classifications together with the threshold of reliable and clinically significant deterioration.

Table 7 – Classifications of pre-treatment to post-treatment change scores (N= 4011)

Classification	Number of occurrence (%)	Number of occurrences reaching Clinical Significance (%)
Reliable Deterioration	125 (3.1%)	47 (1.2%)
No change	2159 (53.8%)	-
Reliable Improvement	1727 (43.1%)	1253 (31.2%)

The data were further analysed to determine differences between the three treatment levels: low intensity, CBT, and counselling. The occurrences of overall deterioration, represented as a percentage of the total N within each treatment, are shown in Table 8. The data for the *no change* and *reliable improvement* are included for comparative purposes.

Table 8– Classifications of pre-treatment to post-treatment change scores (N= 4011) within the three treatment levels.

	Low Intensity (N=2565)	High Intensity CBT (N=694)	High Intensity Counselling (N=752)
Reliable Deterioration	77 (3.0%)	25 (3.6%)	23(3.1%)
No change	1416 (55.2%)	363(52.3%)	380(50.5%)
Reliable Improvement	1072 (41.8%)	306(44.1%)	349(46.4%)

It can be seen again that the rates of overall deterioration were very similar across the treatment levels, with low intensity at 3.0%, CBT at 3.6% and Counselling at 3.1%. There were no significant differences between the treatment levels ($\chi^2 (2) = 0.662, p=0.718$). This indicates that there were no significant differences in the rates of either sudden deterioration or overall deterioration for the three treatment levels of *low intensity*, *high intensity CBT* and *high intensity Counselling*.

Summary

The rates of *overall deterioration* and *sudden deterioration* yielded a lower rate of deterioration than average (3.1%) with the rate for sudden deterioration being a very similar rate (3.4%). However there is some evidence that those clients who experience *sudden deterioration* are more likely to experience *overall deterioration*, and so the deterioration can be maintained. There was also no evidence that there was an impact of treatment on either *overall deterioration* or *sudden deterioration*.

Predictors of deterioration

The final aim of this study was to determine if any predictors of either *overall deterioration* or *sudden deterioration* could be identified within this dataset. In order to test for predictors of overall and sudden deterioration, data was analysed using SPSS Statistics 19 (SPSS Inc, 2010). Preliminary analyses were performed initially for the potential variables, in order to identify those variables most likely to have a significant effect within the tests for predictors. This was accomplished via chi-squared, *t*-tests or Mann-Whitney analyses. Backwards logistical regressions (Likelihood Ratio) were then performed in order to test for predictors or a predictive model of *overall deterioration* and *sudden deterioration*, using only the potentially predictive variables identified as likely to be significant in preliminary analyses. A further regression was then performed for both overall and sudden deterioration using only the variables identified in the initial regression.

Predictors of overall deterioration

Preliminary Analyses

Variables were separated into continuous and categorical and appropriate analyses were conducted to discover if they were likely to have an impact in the logistical regression.

Categorical Variables

The majority of the potential variables were categorical. These were 1) gender, 2) sudden deterioration present or not, 3) Indices of Multiple Deprivation (separated into high and low categories based on the mean), 4) Ethnicity (white or not white), 5) Treatment format (group or individual), 6) Drop out of treatment or not, 7) treatment level (low intensity, CBT, Counselling), 8) Initial PHQ-9 score in the severe range or not, 9) Initial GAD-7 score in (at least) the moderate range or not, and 10) Initial employment status (employed: full or part time employment, or student; unemployed: unemployed, homemaker, or retired).

In order to provide the potential for clinically useful descriptions regarding initial outcome measure (PHQ-9 and GAD-7) scores (that is- for an individual being assessed for treatment, it may be beneficial to have an indication of the severity range most predictive of overall or sudden deterioration, rather than simply higher or lower scores), these were converted into categorical variables of whether or not the initial score was in the moderate severity range or not, and in the severe clinical range or not. The category (either severe or moderate ranges) with the highest level of significance in the model was kept, and the other discarded.

Table 9 presents the results of chi-squared testing for potential categorical variables. Those variables that were found to be significant are described below.

Table 9 – Preliminary Analyses for potential categorical variables for overall deterioration

Variable	Categories	N included		Chi-Squared
		Deterioration	No Deterioration	
Gender	Male	51	1364	$\chi^2 (1) = 1.723,$ p=0.189
	Female	74	2522	
Sudden Deterioration*	Yes	39	93	$\chi^2 (1) = 315.763,$ p<0.001
	No	86	3793	
IMD category	Most deprivation	72	2223	$\chi^2 (1) = 0.030,$ p=0.862
	Least deprivation	52	1658	
Ethnicity	White	99	3038	$\chi^2 (1) = 0.059,$ p=0.808
	Not White	10	333	
Treatment format**	Group	0	3	$\chi^2 (1) = 0.097,$ p=0.756
	Individual	125	3880	
Drop out*	Yes	36	838	$\chi^2 (1) = 3.720,$ p=0.054
	No	89	3048	
Intervention	Low intensity	77	2488	$\chi^2 (2) = 0.662,$ p=0.718
	CBT	25	669	
	Counselling	23	729	
Initial PHQ-9 score*	Severe	6	968	$\chi^2 (1) = 26.637,$ p<0.001
	Below Severe	119	2918	
Initial GAD-7 score*	Moderate or above	79	2784	$\chi^2 (1) = 4.265,$ p=0.039
	Below moderate	46	1100	
Initial Employment Status*	Employed	59	2312	$\chi^2 (1) = 7.694,$ p=0.006
	Not employed	66	1568	

*denotes significant (or close to significant) variables

** denotes that chi-squared assumptions were not met (cell count too low)

Continuous Variables

Normality Testing

In order to ensure the most appropriate analyses were performed on the continuous variables, they were plotted into histograms to assess for normality (see Appendix K). The initial WSAS scores *did* fit the normal distribution curve, and so parametric tests can be applied as the assumption of normality has been met. The age of participants, and the initial 'phobia' scores *did not* fit the normal distribution, and so these variables have been analysed using non-parametric testing (Mann-Whitney).

Parametric Tests

The only continuous variable that met the assumption of normality was the initial WSAS score, however this was not significantly related to deterioration ($t(4002) = 1.034$, $p = 0.301$), and so was not utilised in the regression.

Non-Parametric Tests

Those variables that did not meet the assumption of normality (i.e., age of participants, and the initial 'phobia' scores) were analysed using Mann Whitney tests. However neither were significant in relation to deterioration within the dataset (Initial treatment phobia score, ($Mdn = 6$), $p = 0.513$; Age, ($Mdn = 40$), $p = 0.196$ and so were not included in the regression analyses.

Logistical Regression Analysis

Following on from the above preliminary analyses, five variables were included in the regression analysis. These variables were as follows: (1) sudden deterioration, (2) drop out, (3) initial PHQ-9 severe severity, (4) initial GAD-7 moderate severity, and (5) initial employment status. The variable of drop out was not significant in the preliminary analyses (chi-squared), however it did reach a significance level of $p=0.054$, and so drop out has been included as a potential variable.

Table 10 – Initial Logistical regression for overall deterioration

Variable	Included	B (Standard Error)	EXP (B) (OR)	95% Confidence interval for Exp (B)		Significance
				Lower	Upper	
Sudden Deterioration*	Yes	2.806 (0.226)	16.546	10.621	25.776	p<0.001
	No			Baseline		
Initial PHQ-9 range*	Below	1.749 (0.432)	5.749	2.465	13.410	p<0.001
	Severe			Baseline		
Initial GAD-7 range	Severe			Baseline		
	Below	0.216 (0.203)	1.241	0.833	1.848	p= 0.289
	Moderate			Baseline		
	Moderate or above			Baseline		
Drop out*	Yes	0.479 (0.216)	1.614	1.057	2.465	p= 0.027
	No			Baseline		
Initial Employment Status*	Not Employed	0.702 (0.194)	2.018	1.380	2.951	p<0.001
	Employed			Baseline		
Constant		-5.815 (0.444)	0.003			p<0.001

*Denotes significant variables

All variables, with the exception of GAD-7, were significant (see Table 10 for details). Accordingly, the GAD-7 variable was removed and the revised model is presented in Table 11. All of the significant variables were combined together with their interactions and included in the analysis. However, they did not add more to the model than the original variables. (See Appendix L for the results of these analyses).

Table 11 – Revised logistic regression model for overall deterioration

Variable	Included	B (Standard Error)	EXP (B) (OR)	95% Confidence interval for Exp (B)		Significance
				Lower	Upper	
Sudden Deterioration*	Yes	2.798 (0.226)	16.410	10.542	25.542	p<0.001
	No	Baseline				
Initial PHQ-9 range*	Below	1.816 (0.427)	6.148	2.664	14.191	p<0.001
	Severe	Baseline				
Drop out*	Yes	0.455 (0.215)	1.576	1.035	2.400	p= 0.034
	No	Baseline				
Initial Employment Status*	Not Employed	0.686 (0.193)	1.985	1.359	2.898	p<0.001
	Employed	Baseline				
Constant		-5.787 (0.442)	0.003			p<0.001

*Denotes significant variables

The model shown above in Table 11 shows all the significant predictors of deterioration within the dataset, and predicted 96.9% of the rate of deterioration. However due to the fact that deterioration is such a rare event within a large sample size, a high level of *overall prediction* was found despite 0.0% of reliable deterioration being predicted. Hence it can be seen that model was not robust for predicting deterioration, but the following variables *were* predictive of deterioration:

- Initial Employment status – Those participants who were not employed (that is retired, homemaker or unemployed) were over three times more likely to experience overall deterioration over those who were employed (full-time employed, part-time employed or unemployed).
- Drop out – Those clients who dropped out of treatment were one and a half times more likely to experience deterioration over those who did not drop out. However examination of the 95% confidence levels of *Exp (B)* shows that this variable may not have an effect.
- Initial PHQ-9 score below severe severity threshold – Participants whose initial PHQ-9 score was below the severe threshold were over six times more likely to experience deterioration.
- Sudden deterioration – Clients who experienced sudden deterioration during their treatment were 16 times more likely to also experience overall deterioration over those who did not experience sudden deterioration through treatment.

Predictors of sudden deterioration

In order to test for predictors of sudden deterioration, the same analytic strategy was employed as for predictors of overall deterioration. Hence preliminary analyses were performed to discover the variables with the most potential to be predictive of sudden deterioration. Those variables that were significant in the preliminary analyses were then utilised in a backwards logistical regression.

Preliminary Analyses

Categorical Variables

The categorical potential variables were very similar to those used for *overall deterioration*. These were 1) gender, 2) overall deterioration present or not, 3) Indices of Multiple Deprivation (separated into high and low categories based on the median), 4) Ethnicity (white or not white), 5) Treatment format (group or individual), 6) Drop out of treatment or not, 7) treatment level (low intensity, CBT, Counselling), 8) Initial PHQ-9 score in the severe range or not, 9) Initial GAD-7 score in (at least) the moderate range or not, and 10) Initial employment status (employed: full or part time employment, or student; unemployed: unemployed, homemaker, or retired).

For the categorical potential variables a chi-squared analyses was performed and the results are presented in Table 12. This shows that 'overall deterioration or not', initial PHQ-9 score above the severe severity threshold and initial GAD-7 score above the moderate threshold were the only significant categorical variables. Therefore these variables were included in the initial regression analysis.

Table 12–Preliminary Analyses for potential categorical variables for sudden deterioration

Variable	Categories	N included		Chi-Squared
		Sudden Deterioration	No Sudden Deterioration	
Gender	Male	44	1371	$\chi^2 (1) = 0.279$, p=0.597
	Female	88	2508	
Overall Deterioration*	Yes	39	86	$\chi^2 (1) = 315.763$, p<0.001
	No	93	3793	
IMD category	Most Deprivation	85	2211	$\chi^2 (1) = 1.050$, p=0.305
	Least Deprivation	51	1659	
Ethnicity	White	116	3021	$\chi^2 (1) = 3.351$, p=0.067
	Not White	6	337	
Treatment format**	Group	0	3	$\chi^2 (1) = 0.202$, p=0.653
	Individual	132	3873	
Drop out*	Yes	28	846	$\chi^2 (1) = 0.031$, p=0.861
	No	104	3033	
Intervention	Low intensity	87	2548	$\chi^2 (2) = 1.377$, p=0.502
	CBT	26	678	
	Counselling	20	739	
Initial score*	PHQ-9	Severe	9	$\chi^2 (1) = 23.161$, p<0.001
		Below Severe	124	
Initial score*	GAD-7	Severe	41	$\chi^2 (1) = 6.222$, p=0.013
		Below moderate	92	
Initial Employment Status	Employed	82	2341	$\chi^2 (1) = 0.304$, p=0.581
	Not employed	51	1609	

*denotes significant (or close to significant) variables

** denotes that chi-squared assumptions not met (cell count too low)

Continuous Variables

Parametric Tests

As described above the only normally distributed continuous variable was the initial WSAS score. In relation to sudden deterioration this was significant

($t(4079)=-2.223$, $p=0.026$), and so this variable was included in the regression analysis.

Non-Parametric Tests

The two variables that did not meet the criteria of normality were age and initial phobia score. Neither of these potential variables showed a significant difference in a Mann-Whitney test for sudden deterioration (Initial phobia score ($Mdn = 6.0$), $p= 0.373$, Age ($Mdn = 40.0$) $p= 0.741$) and so were not included in the regression.

Logistical Regression Analysis

There were four variables that were shown to be significantly related to sudden deterioration, and so these were the only potential variables included in the regression analysis. These variables were deterioration, initial PHQ-9 score below the severe range, initial GAD-7 score below the severe range, and the initial WSAS score.

Table 13 – Initial logistical regression for sudden deterioration

Variable	Included	B (Standard Error)	EXP (B) (OR)	95% Confidence interval for Exp (B)		Significance
				Lower	Upper	
Deterioration	Yes	2.768 (0.222)	15.928	10.303	24.624	p<0.001
	No	Baseline				
Initial PHQ-9 range	Below	1.170 (0.380)	3.224	1.531	6.789	p= 0.002
	Severe	Baseline				
Initial GAD-7 range	Below	0.051 (0.215)	1.053	0.691	1.605	p= 0.811
	Severe	Baseline				
Initial WSAS score		-0.004 (0.011)	0.996	0.974	1.018	p= 0.702
Constant		-4.654 (0.440)	0.010			p<0.001

It can be seen from Table 13 above that the only variables that were significant were initial PHQ-9 score below the severe threshold and overall deterioration. All other variables were removed from the model, and a revised model is presented below in Table 14.

Table 14 – Revised logistical regression model for sudden deterioration

Variable	Included	B (Standard Error)	EXP (B) (OR)	95% Confidence interval for Exp (B)		Significance
				Lower	Upper	
Deterioration*	Yes	2.764 (0.222)	15.858	10.267	24.494	p<0.001
	No	Baseline				
Initial PHQ-9 range*	Below	1.243 (0.352)	3.467	1.740	6.908	p<0.001
	Severe	Baseline				
	Severe	Baseline				
Constant		-4.753 (0.337)	0.009			p<0.001

* Denotes significant variables

The model presented in Table 14 above shows the significant predictors of sudden deterioration within the dataset, and this showed a high level of prediction of 96.7%. No incidences of sudden deterioration were predicted at all, and so this model is unlikely to represent a robust model for predicting sudden deterioration. The variables included however do have some predictive validity in that experiencing overall deterioration showed almost 16 times more likely to also have experienced sudden deterioration. Also having an initial PHQ-9 score below the severe range (19 or below) indicated almost three and a half times more likely to experience sudden deterioration.

Summary

The aim of the present research was to investigate deterioration, both sudden and overall, in an IAPT dataset. The process of defining sudden deterioration has been outlined and illustrated, resulting in the proposed definition of a deterioration of at least the RCI for the outcome measure between sessions, with no sudden gain within the treatment. Rates of sudden deterioration and overall deterioration were found to be 3.4% and 3.1% respectively, with no significant differences between treatment levels. Logistical regression analyses were performed using appropriate variables, which demonstrated predictors for overall deterioration were initial employment status, drop out, initial PHQ-9 score below the severe range, and if sudden deterioration were present. The only predictors found for sudden deterioration were the presence of overall deterioration and initial PHQ-9 score below the severe range.

Discussion

The aims of this study were to investigate both deterioration rates and predictors of deterioration during psychotherapy, specifically within the IAPT initiative. The study also attempted to explore whether there was any evidence that the previously unexamined phenomenon of sudden deterioration existed, and again to determine if there were any predictors of this.

The results demonstrated a relatively low deterioration rate of 3.1%, which is lower than the general average reported by both Mohr (1995) and Lambert & Ogles (2004) of 5-10%. Within this group, 37.6% experienced deterioration that was clinically significant, meaning that the score moved from the functional range to the clinical range, indicating that after treatment clients were more similar to those who are depressed than to those without. Lower than average rates of deterioration are not unique, and have been found by researchers previously, for example Ford, Fisher & Larson (1997), who reported a rate of between 0 and 3%. However, this remains uncommon in routine practice. The phenomenon of sudden deterioration was investigated using different definitions, and found to be present within the sample at a rate of 3.4%, a rate very similar to the rate of overall deterioration. A number of predictors of deterioration and *sudden deterioration* were identified. However, the models were poor at predicting *overall deterioration* and *sudden deterioration* as opposed to predicting these phenomena not being present.

Sudden deterioration has been shown to exist within this routine practice population of the IAPT service. This was defined as a reliable deterioration,

using the concept of RCI, which occurred between two consecutive treatment sessions. An additional stipulation was added in order to ensure the sudden deterioration is not as a part of variability in the process for that client. This was done through excluding any treatment episode where there was also a reliable session-by-session decrease (improvement or sudden gain) in score. *Sudden deterioration* was found to be present in a small percentage (3.4%) of clients, and this research has demonstrated a significant relationship between *overall deterioration* and *sudden deterioration*. This has been shown both with the regression analyses, and the fact that 29% of those clients who experienced *sudden deterioration* ended treatment having reliably deteriorated. Hence this may be a beneficial area for further research as early identification of deterioration.

Lambert and colleagues have been the primary researchers in the area of prevention of deterioration. They have proposed that outcome measures should be utilised at every session, and therapists (and sometimes clients also e.g. Lambert, 2010) are provided with feedback, and should negative change be detected a specific guided intervention advised. They identify deterioration via an expected recovery rate, and should the client's progress differ significantly from this, an alert is raised and guidance provided to the therapists. They have found that deterioration can be significantly reduced through this process. The present research into the phenomenon of sudden deterioration has provided further evidence that a sudden deterioration is related to deterioration overall, and the policy of providing feedback can potentially be beneficial. However, the definition of sudden deterioration would be significantly more manageable in

routine clinical settings, without requiring skills in statistical analyses as opposed to calculating change from expected recovery rates. This may be especially relevant for large scale group interventions such as those in the low intensity treatments that are often provided by professions with less training in statistics.

There has been some effort by researchers into predictors of deterioration in the hopes of identifying clients who may be more likely to experience deterioration so that it may be avoided. However, there have been few advances made in this area, and there has been minimal advancement as a very complex and muddled picture has emerged. Unfortunately, this research has also not been able to provide a clear model for prediction of either deterioration or sudden deterioration, although certain variables reached significance and can add to the current knowledge in this area.

Initial symptom severity has been shown previously to have predictive qualities for deterioration. Sotsky et al. (1991) found that low depression severity predicted poorer outcome, however Prestano et al. (2008) reported deterioration in the participant with the lowest initial level of psychopathology. In the present research, initial symptom severity was also shown to be a significant variable of both overall deterioration and sudden deterioration, specifically that starting treatment with a PHQ-9 score below the 'severe' range significantly increases the likelihood of experienced deterioration. Although it is possible that this is simply a factor of the outcome measure itself, in that the maximum score is 27, and to be in the severe range a score must be 20 or above, and the RCI for the

PHQ-9 is 6. Therefore, there is limited scope for deterioration if initial scores are already within the severe range.

It has previously been found that a high level of instability in a person's life can also predict deterioration by Moos (2005). He was referring to both unemployment and being single. Marital status was unavailable for this dataset, but initial employment status was and was categorised into employed (full or part time employed and student status) and not employed (unemployed, retired, or full time homemaker). It was found that those who were not employed were almost twice as likely to experience deterioration, which further suggests that instability related to occupation can predict deterioration. This is of great significance to the IAPT initiative which has an aim of getting people back to work, and indicates that a heightened awareness of the potential for deterioration for this group would be beneficial for clinicians to adopt.

Another variable that had some predictive value for deterioration was drop out. Those clients who dropped out of treatment were one and a half times more likely to experience deterioration. However, consideration of the lower 95% confidence interval around this variable of approximately 1:1 within the regression suggests it would be useful for future research to attempt to replicate this. It is possible that the experience of the deterioration caused the clients to drop out, or that the process of treatment was too much for the client, or related to life events outside of therapy. However, it is of note that drop out was not significantly related to sudden deterioration, suggesting that a gradual deterioration is more likely to result in a client dropping out of therapy.

Limitations

There were a number of limitations to this study. In particular the data used were collected from routine practice of clinicians working within the IAPT initiative. There are obvious inherent benefits to this in that it represents the experiences of clients in the real world, rather than only in clinical trials. However this also means that there is no control group, as there was no wait list provision in the IAPT model. Also there were many missing data points, which is less likely to occur in clinical trials. Individuals who access IAPT services do so within a primary care model, and therefore tend to have mild to moderate mental health diagnoses. This group represents the majority of people receiving mental health care, which is also the basis for the IAPT initiative (Layard, 2006.) However individuals with severe or more complex disorders, such as personality disorder, psychosis or bi-polar affective disorder, are not represented within the population examined in this research, and so no conclusions about this group can be inferred from the results. In addition the IAPT service primarily provides treatment based on CBT or counselling models, and so conclusions cannot be made regarding other models of types of psychotherapy. However it has previously been noted in deterioration literature that there is currently minimal evidence to suggest that there are significant differences between therapy models (Mohr, 1995.)

This study was also part of a larger research group, and so there were certain limitations in the areas that could be examined in the present study. In particular the impact of specific therapists on outcome was a separate research stream,

and so it was considered inappropriate to also explore therapist effects on the occurrence of general deterioration or sudden deterioration. However the minimal literature available regarding the effect of individual therapists on general deterioration suggests the potential for this being a significant variable, hence it not being included in this study limits the cohesiveness of the results.

As a result of routine data collection, in this dataset there was not access to reliable information regarding previous treatment. The only accurate data was regarding whether or not the client had more than one episode, which did show an initial significance in the preliminary analyses. It was decided to exclude this variable as it only referred to previous treatment within the IAPT service as previous treatment other than this was not recorded. However this suggests it may be beneficial for future studies into either deterioration or sudden deterioration to consider the inclusion of this variable should it be available.

Another significant limitation is the fact that both deterioration and sudden deterioration are very rare events. This significantly limits the statistical analyses when using regression analyses in such a large dataset. In particular the predictors found in the regression models for both sudden deterioration and deterioration was able to predict 0.0% of these aspects, despite accurately predicting over 96% of the responses.

In this study only one outcome measure was used as the focus was primarily on depression, however this also limits the research. The PHQ-9 is a self report measure, and so considered only one perspective, and only one aspect of the participants' overall functioning was therefore examined. It cannot be assumed

that similar deterioration occurred in other aspects of participants' functioning¹⁰. In this population, session-by-session data was also available for the GAD-7, which if included could have allowed a broader understanding of the deterioration by encompassing experiences of symptoms of anxiety also. However counselling is not recommended by NICE guidelines for the treatment of anxiety, and so comparison across the treatment types, as was performed in the present study, would be untenable. Therefore there are a number of limitations to this research, which significantly limits the generalisability of the results. The use of IAPT data, the use of only one outcome measure targeting symptoms of depression, the amount of potential participants due to the lack of complete data, all impact on the confidence with which conclusions can be drawn. Hence further research is indicated in both general deterioration and sudden deterioration.

Clinical and Theoretical Implications

The development of increased understanding regarding deterioration during the course of psychotherapeutic interventions has a number of clinical implications. This is particularly important for new interventions such as the IAPT initiative, especially considering the cost of the initiative and the early criticism. It was intended that the results of this study would allow useful information to refine service standards about the IAPT model as a whole, and help to identify clients who may be at an increased risk of deterioration. The more information clients,

¹⁰ For further discussion on the limits of outcome measures in deterioration research see Mohr (1995).

clinicians and service commissioners have available to them around expected recovery and predictors of deterioration, the more potential to improve assessment procedures, treatments, and reduce the potential of therapy having a negative effect on clients. The results suggest that there is a low rate of overall deterioration, and there were no significant differences among the three treatment levels. This indicates that the IAPT initiative is a relatively strong treatment model with regards to overall deterioration. However, this research was a single site study and in order to allow for generalisation it would be helpful for future researchers to both determine and report rates of overall deterioration in the IAPT initiative and in other treatment models. Also it can be of significant benefit to the psychotherapy professions as a whole to ensure treatments provided are the most appropriate and the negative impacts of such have been thoroughly explored and weaknesses minimised.

Deterioration is an important and under-researched issue. Therefore, work in this area also has significant theoretical implications. The predictors of deterioration found in these data suggests that clinicians could pay particular attention to the monitoring of the progress of particular clients who experience a *sudden deterioration*, have an initial PHQ-9 score below the severe range, and have an initial employment status of unemployed, retired or homemaker. These factors have been shown previously in the literature to have an impact on deterioration and have been replicated in this study and so have sufficient evidence to affect clinical practice.

Sudden deterioration has been investigated for the first time in this research, and it has been found that the phenomenon exists within the IAPT data, and was equally present in the three treatment types of low intensity, CBT and counselling. Hence, it can tentatively be suggested that this phenomenon, while a rare event, may be present in a number of different therapy models. *Sudden deterioration* alone may be considered a part of the process and not a cause for concern provided improvement occurs, however the fact that *sudden deterioration* is so closely linked to *overall deterioration* indicates that this would be a useful avenue for future research. In particular further research is indicated in other therapy models, and for other disorders not encompassed in primary care settings. In order to account for some of the limitations of this study, it would also be beneficial to use a robust assessment strategy that allows a broad understanding of deterioration or improvement in a number of different areas of participants' wellbeing and functioning, and also to explore the potential impact and effects of therapist variables. There may be benefit in qualitative studies in order to further understand the experience of these deterioration or sudden deterioration events from the clients', therapists' and also family members' viewpoints.

Conclusions

- A balanced definition of *sudden definition* was derived. That being a reliable between session deterioration that is not allied to sudden gains within the same treatment episode.
- The rate of *sudden deterioration* found within the dataset was 3.4%.
- The rate of *overall deterioration* found within the dataset was 3.1%. Compared to the reported average (5-10%), this is relatively low.
- There were no differences in the rates of *sudden deterioration* or *overall deterioration* between the three treatment levels.
- Low rates of deterioration and similar rates across treatment levels suggest that the IAPT is a safe and appropriate treatment model with regards to deterioration.
- Predictive models were not found for either *sudden deterioration* or *overall deterioration*.
- Predictors of *overall deterioration* were found to be initial employment status, drop out, initial PHQ-9 score below the severe range, and if *sudden deterioration* was present.
- Predictors for sudden deterioration were the presence of overall deterioration and initial PHQ-9 score below the severe range.
- Future research into *sudden deterioration* could be a useful addition to the deterioration literature.

References

- Arroll, B., Goodyear-Smith, F., Crengel, S., Gunn, J., Kerse, N., & Fishman, T., et al. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine*, 8, 348 – 353, [doi:10.1370/afm.1139](https://doi.org/10.1370/afm.1139)
- Barlow, D. H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist*, 65, 13-20, [doi:10.1037/a0015643](https://doi.org/10.1037/a0015643)
- Boisvert, C. M. (2010). Negative treatment effects: Is it time for a black box of warning? *American Psychologist*, 65, 680 -681, [doi:10.1037/a0020078](https://doi.org/10.1037/a0020078)
- Boisvert, C.M., & Faust, D. (2006). Practicing psychologists' knowledge of general psychotherapy research findings: Implications for science –practice relations. *Professional Psychology: Research and Practice*, 37, 708-716, [doi:10.1037/0735-7028.37.6.708](https://doi.org/10.1037/0735-7028.37.6.708)
- Bower, P., & Gilbody, S., (2005). Stepped care in psychological therapies: access, effectiveness and efficiency: Narrative literature review. *The British Journal of Psychiatry*, 186, 11-17, [doi:10.1192/bjp.186.1.11](https://doi.org/10.1192/bjp.186.1.11)
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving Access to Psychological Therapy: Initial evaluation of two UK demonstration sites. *Behaviour, Research, and Therapy*, 47, 910 – 920, [doi:10.1016/j.brat.2009.07.010](https://doi.org/10.1016/j.brat.2009.07.010)

- Cromby, J., Diamond, R., Kelly, P., Moloney, P., Priest, P., & Smail, D., et al. (2008). Our Big Fat Multi-Million Pound Psychology Experiment. *Clinical Psychology Forum*, 181.
- Dimidjian, S. & Hollon, S.D. (2010). How would we know if psychotherapy were harmful? *American Psychologist*, 65, 21-33, [doi:10.1037/a0017299](https://doi.org/10.1037/a0017299)
- Field, A. (2009). *Discovering statistics Using SPSS: Third Edition*. Sage Publications, LA, USA.
- Ford, J., Fisher, P., & Larson, L. (1997). Object relations as a predictor of treatment outcome with chronic Posttraumatic Stress Disorder. *Journal of Consulting & Clinical Psychology*, 65, 547-559, [doi:10.1037/0022-006X.65.4.547](https://doi.org/10.1037/0022-006X.65.4.547)
- Forsythe, A., May, P., & Engelman, L. (1971). Prediction by multiple regression how many variables to enter? *Journal of Psychiatric Research*, 8, 119-126, [doi:10.1016/0022-3956\(71\)90013-6](https://doi.org/10.1016/0022-3956(71)90013-6)
- Fournier, J., DeRubeis, R.J., Shelton, R.C., Hollon, S.D., Amsterdam, J.D. & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting and Clinical Psychology*, 77, 775-787, [doi:10.1037/a0015401](https://doi.org/10.1037/a0015401)
- Fricke, S., Moritz, S., Andresen, Jacobsen, B., Kloss, M., & Rufer, M., et al. (2006). Do personality disorders predict negative treatment outcome in obsessive-compulsive disorders? A prospective 6-month follow-up study. *European Psychiatry*, 21, 319-324, [doi:10.1016/j.eurpsy.2005.03.010](https://doi.org/10.1016/j.eurpsy.2005.03.010)

Glover, G, Webb, M., & Evison, F. (2010). Improving Access to Psychological Therapies: A review of the progress made by sites in the first roll-out year. North East Public Health Observatory.

Hartman, A., Orlinsky, D., Weber, S., Sandholtz, A., & Zeeck, A. (2010). Session and intersession experience related to treatment outcome in bulimia nervosa. *Psychotherapy Theory, Research, Practice, Training*, 47, 355-370, [doi:10.1037/a0021166](https://doi.org/10.1037/a0021166)

Hosmer, D. & Lemeshow, S. (1989). Applied logistic regression. *Wiley, New York*.

Jacobson, N. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19, [doi:10.1037/0022-006X.59.1.12](https://doi.org/10.1037/0022-006X.59.1.12)

Kellett, S., Clarke, S., & Matthews, L. (2007). Delivering group psychoeducational CBT in Primary Care: Comparing outcomes with individual CBT and individual psychodynamic-interpersonal psychotherapy. *British Journal of Clinical Psychology*, 46, 211-222. [doi:10.1348/014466506X146188](https://doi.org/10.1348/014466506X146188)

Kroenke, K., Spitzer, R., & Williams, J., (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613, [doi:10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)

- Lambert, M. (2007). Presidential Address: What have we learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1-14, [doi:10.1080/10503300601032506](https://doi.org/10.1080/10503300601032506)
- Lambert, M. (2010). *Prevention of Treatment Failure: The Use of Measuring, Monitoring, and Feedback in Clinical Practice*. American Psychological Association, Washington, USA.
- Lambert, M., Whipple, J.L., Smart, D.W., Vermeersch, D.A., Nielsen, S.L., & Hawkins, E.J., (2001). The effects of providing therapists on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, 11, 49-68.
- Lambert, M., & Ogles, B. (2004). The efficacy and effectiveness of psychotherapy. In M. Lambert (Ed.) *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. (5th ed., 139-193) New York: Wiley.
- Lambert, M., Harmon, C., Slade, Whipple, J.L., & Hawkins, E.J. (2005). Providing feedback to Psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology*, 61, 165-174.
- Layard, R. (2004). Mental Health: Britain's biggest social problem. www.strategy.gov.uk/downloads/files/mh_layard.pdf.
- Layard, R. (2006). The depression report: A new deal for depression and anxiety disorders.

- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2, 53 – 70, [doi:10.1111/j.1745-6916.2007.00029.x](https://doi.org/10.1111/j.1745-6916.2007.00029.x)
- Lowe, B., Decker, O., Moore, S., Brahler, E., Schellberg, D., & Herzog, W., et al. (2008). Validation and standardization of the Generalised Anxiety Disorder screener (GAD-7) in the general population. *Medical Care*, 46, 266 – 274, [doi:10.1097/MLR.0b013e318160d093](https://doi.org/10.1097/MLR.0b013e318160d093)
- Lunnen, K., Ogles, B., & Pappas, L. (2008). A multiperspective comparison of satisfaction, symptomatic change, perceived change, and end-point functioning. *Professional Psychology: Research & Practice*, 39, 145-152, [doi:10.1037/0735-7028.39.2.145](https://doi.org/10.1037/0735-7028.39.2.145)
- Manning, P., Hardy, G., & Kellett, S., (2010). Reversals of sudden gains made during cognitive therapy with depressed adults: A preliminary investigation. *Behavioural and Cognitive Psychotherapy*, 38, 491-495, [doi:10.1017/S1352465810000123](https://doi.org/10.1017/S1352465810000123)
- McMillan, D., Gilbody, S., & Richards, D., (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders*, 127, 122 – 129, [doi:10.1016/j.jad.2010.04.030](https://doi.org/10.1016/j.jad.2010.04.030)
- Mohr, D., Beutler, L., Engle, D., Shoham-Salomon, V., Bergan, J., Kaszniak, A., & Yost, E. (1990). Identification of patients at risk for nonresponse and

- negative outcome in psychotherapy. *Journal of Consulting and Clinical Psychology*, 58, 622-628, [doi:10.1037/0022-006X.58.5.622](https://doi.org/10.1037/0022-006X.58.5.622)
- Mohr, D.C., (1995.) Negative outcomes in psychotherapy: A critical review. *Clinical Psychology: Science & Practice*, 2, 1-27.
- Moos, R.H., (2005). Iatrogenic Effects of Psychosocial Interventions for Substance Use Disorders: Prevalence, Predictors, Prevention', *Addiction*, 100, 1240-1248.
- Mundt, J.C., Marks, I.M., Shear, M.K. & Greist, J.H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180, 461-464, [doi:10.1192/bjp.180.5.461](https://doi.org/10.1192/bjp.180.5.461)
- Parry, G., Barkham, M., Brazier, J., Dent-Brown, K., Hardy, G., Kendrick, T., et al. (2011). An evaluation of a new service model: Improving Access to Psychological Therapies demonstration sites 2006-2009. Final report. NIHR Service Delivery and Organisation programme.
- Prestano, C., Lo Coco, G., Gullo, S., Lo Verso, G., (2008). Group analytic therapy for eating disorders: Preliminary results in a single-group study. *European Eating Disorders Review*, 16, 302-310, [doi:10.1002/erv.871](https://doi.org/10.1002/erv.871)
- Reich, J. (2003). The effect of axis II disorders on the outcome of treatment of anxiety and unipolar depressive disorders: A review. *Journal of Personality Disorders*, 17, 387-405, [doi:10.1521/pedi.17.5.387.22972](https://doi.org/10.1521/pedi.17.5.387.22972)

- Richards, D.A., & Suckling, R., (2008). Improving access to psychological therapy: The Doncaster demonstration site organisational model. *Clinical Psychology Forum*, 181, 9-16.
- Scogin, F., Floyd, M., Jamison, C., Ackerson, J., Landreville, P. & Bissonnette, L. (1996). Negative outcomes: What is the evidence on self-administered treatments? *Journal of Consulting & Clinical Psychology*, 64, 1086-1089. [doi:10.1037/0022-006X.64.5.1086](https://doi.org/10.1037/0022-006X.64.5.1086).
- Sotsky, S.M., Glass, D.R., Shea, M.T., Pilkonis, P.A., Collins, J.F., Elkin, I. et al. (1991). Patient predictors of response to psychotherapy and pharmacotherapy: Findings in the NIMH treatment of depression collaborative research program. *American Journal of Psychiatry*, 148, 997-1008.
- Swift, J. K., Callahan, J. L., Heath, C. J., Herbert, G. L., & Levine, J. C. (2010). Applications of the psychotherapy phase model to clinically significant deterioration. *Psychotherapy Theory, Research, Practice, Training*. 47, 235 – 248, [doi:10.1037/a0019787](https://doi.org/10.1037/a0019787)
- Tang, T.Z., & DeRubeis, R.J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67, 894-904, [doi:10.1037/0022-006X.67.6.894](https://doi.org/10.1037/0022-006X.67.6.894)
- Tang, T.Z., DeRubeis, R.J., Hollon, S.D., Amsterdam, J., & Shelton, R., (2007). Sudden gains in cognitive therapy of depression and depression

relapse/recurrence. *Journal of Consulting and Clinical Psychology*, 75, 404-408, [doi:10.1037/0022-006X.75.3.404](https://doi.org/10.1037/0022-006X.75.3.404)

Woody, G.E., McLellan, A.T., Luborsky, L., O'Brien, C.P., Blaine, J, Fox, S. et al., (1984). Severity of psychiatric symptoms as a predictor of benefits from psychotherapy: the Veterans Administration-Penn study. *American Journal of Psychiatry*, 141, 1172-1177.

Appendix A – Approval Letter for Journals



Department Of Psychology. Clinical Psychology Unit.

Doctor of Clinical Psychology (DClin Psy) Programme
Clinical supervision training and NHS research training
& consultancy.

**Clinical Psychology Unit
Department of Psychology
University of Sheffield
Western Bank
Sheffield S10 2TP UK**

Telephone: 0114 2226550
Fax: 0114 2226610
Email: c.harrison@sheffield.ac.uk

20 September 2011

Leigh Thorpe
Third year trainee
Clinical Psychology Unit
University of Sheffield

Dear Leigh

I am writing to indicate our approval of the journal(s) you have nominated for publishing work contained in your research thesis.

Literature Review: Clinical Psychology: Science and Practice

Research Report: British Journal of Clinical Psychology

Please ensure that you bind this letter and copies of the relevant Instructions to Authors into an appendix in your thesis.

Yours sincerely

A handwritten signature in black ink, appearing to read 'Rebecca Knowles', on a light blue background.

Dr Rebecca Knowles
Research Tutor

Appendix B i) Literature Review Journal - Current notes for contributors

Author Guidelines – Clinical Psychology: Science and Practice

Clinical Psychology: Science and Practice follows publication policies and ethical principles of the American Psychological Association (APA). Authors are assumed to be familiar with and are responsible for adherence to the policy. Among the tenets, the policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications or from publishing any manuscript that has already been published in whole or substantial part elsewhere. Authors are obligated to consult journal editors concerning prior publication of any material upon which their article depends.

Although the vast majority of papers will be review and discussion articles, occasionally papers representing programs of research or papers drawing on such research will be published. Compliance with APA ethical standards in the treatment and protection of the sample as elaborated in the APA Ethical Principles is expected by the journal.

Manuscript Format

Manuscripts are to be prepared in accordance with the Publication Manual of the American Psychological Association (5th ed.). Typing instructions (all copy must be double-spaced) and instructions for preparing tables, figures, references, metrics, and abstracts appear in the manual. Manuscripts of regular articles are to be accompanied by an abstract containing a maximum of 960 characters and spaces (which is approximately 120 words), followed by three to six key words. Abstracts, tables, and figure captions should be typed on separate pages, and manuscript pages for any tables or figure captions should be placed at the end of the manuscript for production purposes.

Appendix B ii) Empirical Report Journal - Current notes for contributors

Author Guidelines – British Journal of Psychology

The Editorial Board of the British Journal of Psychology is prepared to consider for publication:

- (a) reports of empirical studies likely to further our understanding of psychology
- (b) critical reviews of the literature
- (c) theoretical contributions Papers will be evaluated by the Editorial Board and referees in terms of scientific merit, readability, and interest to a general readership.

1. Circulation

The circulation of the Journal is worldwide. Papers are invited and encouraged from authors throughout the world.

2. Length

Papers should normally be no more than 8000 words (excluding the abstract, reference list, tables and figures), although the Editor retains discretion to publish papers beyond this length in cases where the clear and concise expression of the scientific content requires greater length.

3. Submission and reviewing

All manuscripts must be submitted via <http://www.editorialmanager.com/bjp/>. The Journal operates a policy of anonymous peer review.

4. Manuscript requirements

- Contributions must be typed in double spacing with wide margins. All sheets must be numbered.
- Manuscripts should be preceded by a title page which includes a full list of authors and their affiliations, as well as the corresponding author's contact details. A template can be downloaded from [here](#).
- Tables should be typed in double spacing, each on a separate page with a self-explanatory title. Tables should be comprehensible without reference to the

text. They should be placed at the end of the manuscript with their approximate locations indicated in the text.

- Figures can be included at the end of the document or attached as separate files, carefully labelled in initial capital/lower case lettering with symbols in a form consistent with text use. Unnecessary background patterns, lines and shading should be avoided. Captions should be listed on a separate sheet. The resolution of digital images must be at least 300 dpi.
- All articles should be preceded by an Abstract of between 100 and 200 words, giving a concise statement of the intention, results or conclusions of the article.
- For reference citations, please use APA style. Particular care should be taken to ensure that references are accurate and complete. Give all journal titles in full.
- SI units must be used for all measurements, rounded off to practical values if appropriate, with the imperial equivalent in parentheses.
- In normal circumstances, effect size should be incorporated.
- Authors are requested to avoid the use of sexist language.
- Authors are responsible for acquiring written permission to publish lengthy quotations, illustrations, etc. for which they do not own copyright. For guidelines on editorial style, please consult the [APA Publication Manual](#) published by the American Psychological Association.

5. Supporting Information

BJOP is happy to accept articles with supporting information supplied for online only publication. This may include appendices, supplementary figures, sound files, videoclips etc. These will be posted on Wiley Online Library with the article. The print version will have a note indicating that extra material is available online. Please indicate clearly on submission which material is for online only publication. Please note that extra online only material is published as supplied by the author in the same file format and is not copyedited or typeset. Further information about this service can be found at <http://authorservices.wiley.com/bauthor/suppmat.asp>

6. Copyright

Authors will be required to assign copyright to The British Psychological Society. Copyright assignment is a condition of publication and papers will not be passed to the publisher for production unless copyright has been assigned. To assist

authors an appropriate copyright assignment form will be supplied by the editorial office and is also available on the journal's website at http://www.blackwellpublishing.com/pdf/CTA_BPS.pdf. Government employees in both the US and the UK need to complete the Author Warranty sections, although copyright in such cases does not need to be assigned.

7. Colour illustrations

Colour illustrations can be accepted for publication online. These would be reproduced in greyscale in the print version. If authors would like these figures to be reproduced in colour in print at their expense they should request this by completing a Colour Work Agreement form upon acceptance of the paper. A copy of the Colour Work Agreement form can be downloaded [here](#).



National Research Ethics Service

South Yorkshire Research Ethics Committee

Millside
Mill Pond Lane
Meanwood
Leeds
LS6 4RA

Telephone: 0113 305 0128

12 October 2010

Dr Stephen Kellett
Programme Director and Consultant Clinical Psychologist
Clinical Psychology Unit
University of Sheffield
S10 2TN

Dear Dr Kellett

Study Title: Identifying patient and service factors that contribute to the degree and shape of change in patient outcomes for patients within the Sheffield IAPT service
REC reference number: 10/H1310/69

The Research Ethics Committee reviewed the above application at the meeting held on 30 September 2010. Thank you for attending to discuss the study.

Ethical opinion

Members agreed that this was a very well-constructed proposal and did not identify any ethical issues.

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is

This Research Ethics Committee is an advisory committee to Yorkshire and The Humber Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within
the National Patient Safety Agency and Research Ethics Committees in England

available in the Integrated Research Application System or at <http://www.rdforum.nhs.uk>. Where the only involvement of the NHS organisation is as a Participant Identification Centre, management permission for research is not required but the R&D office should be notified of the study. Guidance should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Investigator CV		
Protocol	1	13 August 2010
Conditional agreement to be the project's research governance sponsor		
REC application		
Covering Letter		
Review of research proposal		
Evidence of insurance or indemnity		11 August 2010

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

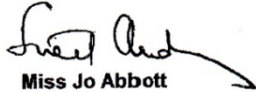
We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

10/H1310/69

Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely



Miss Jo Abbott
Chair

Email: Sinead.audsley@leedspft.nhs.uk

Enclosures: List of names and professions of members who were present at the meeting and those who submitted written comments "After ethical review – guidance for researchers"

Copy to: Ms Lauren Smaller, University of Sheffield

South Yorkshire Research Ethics Committee

Attendance at Committee meeting on 30 September 2010

Committee Members:

<i>Name</i>	<i>Profession</i>	<i>Present</i>	<i>Notes</i>
Sinead Audsley	REC Co-ordinator	Yes	
Miss Jo Abbott	Consultant in Public Health	Yes	
Dr A H Abdelhafiz	Consultant Physician, Elderly Medicine	No	
Miss Stephanie Andrews	Drug & Alcohol Action Team Information Manager	Yes	
Reverend Joan Ashton	Co-ordinator of Chaplaincy Services	Yes	
Miss Helen Barlow	Knowledge Service Manager	Yes	
Professor Nigel Beail	Consultant Clinical Psychologist & Professor of Psychology	Yes	
Mr Ian Cawthorne	Chief Pharmacist	No	
Ms Susan Hampshaw	New Deal for Communities Evaluation Unit Manager	Yes	
Mr Neil Marsden	Police Staff	No	
Dr Anton Mayer	Consultant in Paediatric Intensive Care	Yes	
Mrs Andrea Porritt	District Nurse/Practice Educator	Yes	
Dr Ganesh Rao	Consultant Clinical Neurophysiologist	No	
Mr Jaydip Ray	Consultant ENT Surgeon	No	
Ms Stephanie Rhodes	Neonatal Sister	Yes	
Dr Salil R Sen	General Practitioner	Yes	
Dr Paul Spencer	Consultant Radiologist	Yes	
Mrs Carole Taylor	Deputy Chief Pharmacist	Yes	
Dr Jonathan Train	Consultant Anaesthetist	No	

Appendix D – Ethics Amendment Approval Letter



National Research Ethics Service

NRES Committee Yorkshire & The Humber - South Yorkshire

Millside
Mill Pond Lane
Meanwood
Leeds
LS6 4RA

Tel: 0113 305 0116

08 August 2011

Dr Stephen Kellett
Programme Director and Consultant Clinical Psychologist
University of Sheffield
Clinical Psychology Unit
Department of Psychology
University of Sheffield
S10 2TN

Dear Dr Kellett

Study title: Identifying patient and service factors that contribute to the degree and shape of change in patient outcomes for patients within the Sheffield IAPT service
REC reference: 10/H1310/69
Amendment number: 1
Amendment date: 29 June 2011

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMPs)	1	29 June 2011

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

This Research Ethics Committee is an advisory committee to the Yorkshire and The Humber Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within the National Patient Safety Agency and Research Ethics Committees in England

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

10/H1310/69:	Please quote this number on all correspondence
---------------------	---

Yours sincerely

Jo Abbott
Chair

E-mail: Claire.kelly4@nhs.net

Copy to:

*Lauren Smaller, University of Sheffield
Dr Adrian Carr, Sheffield Health and Social Care NHS Foundation
Trust*

Appendix E – PHQ-9

PHQ-9 questionnaire has been excluded from this thesis due to copyright restrictions.

Appendix F – General Anxiety Disorder – 7 (GAD-7)

GAD-7 questionnaire has been excluded from this thesis due to copyright restrictions.

Appendix G – Work and Social Adjustment Scale (WSAS)

WSAS questionnaire has been excluded from this thesis due to copyright restrictions.

Appendix H - IAPT Phobia Questions

Phobia questions have been excluded from this ethesis due to copyright restrictions.

Appendix I – Table of breakdown of search terms

[] = Articles after basic electronic screening based on category. () = articles after manual title/abstract search – in marked search

Agg = aggregated number of articles

	Web of Knowledge	PsychInfo
Deteriorat*, Psychotherap*	149 [149] (55)	1546 [606] (122)
Deteriorat*, therap*	9657 [659] (76 – agg)	1546 [122] (122 – agg)
Deteriorat*, counselling	74 [8] (77 – agg)	211 [211] (133 - agg)
Deteriorat*, counseling	97 [20] (79 - agg)	55 [22] (135 - agg)
Negative outcome, psychotherap*	15 [15] (88 – agg)	31 [23] (138 - agg)
Negative outcome, therap*	140 [45] (96 - agg)	29 [23] (140 - agg)
Negative outcome, counselling	3 [0] (96 – agg)	61 [59] (148 - agg)
Negative outcome, counselling	2 [0] (96 – agg)	437 [366] (175 - agg)
Negative respon*, psychotherapy*	9 [9] (99 – agg)	11 [6] (175 - agg)
Negative respon*, therap*	151 [24] (102 - agg)	12 [7] (176 - agg)
Negative respon*, counselling	3 [0] (102 – agg)	1 [1] (176 - agg)

Negative respon*, counselling	5 [0] (102 - agg)	0 [0] (176 - agg)
Harm, psychotherap*	227 [182] (133 - agg)	267 [181] (188 - agg)
Harm, therap*	2276 [430] (158 - agg)	664 [541] (194 - agg)
Harm, counselling	89 [44] (160 - agg)	60 [60] (194 - agg)
Harm, counseling	155 [66] (162 - agg)	216 [183] (194 - agg)
Harmful treatment, psychotherapy*	49 [36] (177 - agg)	3 [1] (194 - agg)
Harmful treatment, therap*	1316 [150] (181 - agg)	6 [3] (195 - agg)
Harmful treatment, counselling	23 [13] (182 - agg)	1 [1] (195 - agg)
Harmful treatment, counselling	30 [20] (182 - agg)	0 (195 - agg)
Deteriorat*, predict*	7540 [606] (186 - agg)	11 [10] (196 - agg)
Negative outcome, predict*	234 [83] (191 - agg)	2 [2] (197 - agg)
Negative respon*, predict*	236 [74] (191 - agg)	1 [0] (197 - agg)
Deteriorat*, predict*, client	23 [13] (193 - agg)	5 [4] (199 - agg)
Deteriorat*, predict*, treatment	1327 [172] (200 - agg)	34 [30] (199 - agg)
Treatment failure, psychotherapy*	55 [48] (212 - agg)	56 [38] (211 - agg)

Treatment failure, therap*	5010 [239] (221 - agg)	131 [97] (218 - agg)
Treatment failure, counselling	12 [1] (221 - agg)	2 [2] (218 - agg)
Treatment failure, counselling	32 [9] (221 - agg)	12 [12] (220 - agg)
Iatrogenic treatment effects, psychotherap*	12 [10] (223 - agg)	43 [15] (221 - agg)
Iatrogenic treatment effects, therap*	201 [30] (223 - agg)	132 [78] (224 - agg)
Iatrogenic treatment effects, counselling	2 [0] (223 - agg)	0 (224 - agg)
Iatrogenic treatment effects, counselling	0 [0] (223 - agg)	0 (224 - agg)

Addendum to exclusion criteria

Articles relating to 'interventions that cause harm' refers to the work of Lilienfeld (2007). This list he provided was Critical incident stress debriefing; Scared Straight interventions; Facilitated communication; Rebirthing; Recovered-memory techniques; DID-oriented therapy; Grief counselling for individuals with normal bereavement reactions; Expressive-Experiential therapies; Boot-camp interventions for conduct disorder; DARE programs.

Appendix I continued

Examples of Electronic Category Search terms

After initial searching in the databases, irrelevant categories were excluded electronically. Listed below are examples of the irrelevant categories that were excluded. These lists are not exhaustive.

Excluded: Toxicology; medical laboratory technology; mathematical computational biology; veterinary science; zoology; nuclear science technology; cell biology; criminology penology; transplantation; history.

All categories with potential relevance to the literature search were retained. Presented below is a list of some examples of the categories retained for the manual search.

Included: Psychiatry; Health care sciences services; neurosciences neurology; public environmental occupational health; psychology; substance abuse; behavioural sciences; nursing; social work.

Appendix J – Account of Rating System

The following is a step by step account of the processes employed for the rating system for the literature identified in the review.

- Recommendations by Mohr (1995) were described (these are described within the main text)
- Recommendations by Mohr were delineated into specific criteria, and scores (0, 1, or 2) assigned to the levels to which studies can meet these criteria (see appendix J (i)).
- Articles were ordered chronologically, and the alternate articles assigned to two groups. Group 1 was ordered from oldest to newest, and Group 2 ordered from newest to oldest.
- The primary rater (LT) then rated all the 28 articles for the 6 categories in the assigned order.
- Instructions were then developed for an independent rater to be able to follow the same process as used by LT (see Appendix J (ii)).
- The articles were then ordered according to the overall rating given by LT, separated into 3 approximate groups, allowing for those with the same scores to be in the same group. The highest, mid and lowest rated article in each group made up a subgroup for rating by the independent rater. This subgroup was arranged alphabetically
- The independent rater was provided with 2 hours training by LT to ensure sufficient understanding of the concepts surrounding the rating, and the processes involved in rating.
- The independent rater then rated the articles.

- LT and the independent rater met to compare and discuss the ratings. All differences were discussed, and articles reread until full agreement was reached. (A table displaying these ratings can be found in Appendix J (iii))
- The scores for each criterion were then squared (i.e. 1→1, 2→ 4, 3→ 9) for further differentiation between ratings.

Appendix J (i) – Quality Rating Scale

General Scoring:

1 = no efforts made by investigators to follow recommendation.

2 = partial efforts made but not implemented methodically.

3 = Investigators methodically followed recommendation.

General Notes:

- If more than one study combined within an article choose lowest level achieved.
- If outcome measure is an interview (for criterion C) consider the instrument used and where input comes from – Is it primarily self-report, recorded by therapist – score 1, Does the therapist use own judgement to rate self-report by client – score 2, Does the therapist use own judgement to rate based on interviews with client and others – score 3.
- ‘Outcome Measure’ refers to either psychometric assessment, or other method of determining outcome.

Methodological Recommendations:

A) Identification of negative responders:

1 = Negative responders identified in a vague manner without any statistical validity – (e.g. clients described themselves as ‘worse’).

2 = Efforts made to consider reliability of change in identification of negative responders, but not methodically – (e.g. either statistical or clinical significance considered, although not necessarily used, not both)

3 = Statistical and clinically relevant change considered using documented and reported principles to identify negative responders, although not necessarily used (e.g. Reliable Change Index and Clinically Significant Change both employed).

B) 1) Assessment strategy – outcome measurement (global or specific)

1 = Only global outcome measures used.

- 2 = Only specific outcome measures used.
- 3 = Both specific and global outcome measures used.

B) 2) Assessment strategy – outcome measurement (number of measures)

- 1 = Only one outcome measure used.
- 2 = Two outcome measures used.
- 3 = Three or more outcome measures used.

C) Assessment strategy – measured by more than one perspective

- 1 = Outcome measure from one direct source (e.g. therapist or client)
- 2 = Outcome measures from two direct sources (e.g. therapist and client)
- 3 = Outcome measures from direct sources and others (e.g. client, therapist, family member, observer etc.)

D) Timing of assessment – pre/post, follow up, during treatment

- 1 = Inappropriate timings of assessments which may not be indicative of deterioration (e.g. pre treatment and follow up, as this may be relapse)
- 2 = Appropriate timings, although not at all key stages of treatment for assessments (e.g. pre treatment and post treatment, with no follow up or during treatment assessments)
- 3 = Outcome measures used at all key stages of treatment (e.g. pre treatment, during treatment, post treatment and follow up)

E) Recognition of Negative Responders as separate group

- 1 = Negative responders included with non-responders or those who drop out.
- 2 = N/A
- 3 = Negative responders, non-responders, and positive responders considered as separate groups.

Appendix J (ii) - Instructions to Independent Rater

Thank you for agreeing to be involved in the quality rating of literature for this review.

In the sample articles package you will find:

- Instructions for independent rater of sample articles
- 9 articles, in the order it is intended that you will rate them.
- Recommendations from Mohr (1995)
- Ratings for recommendations from Mohr (1995)
- Summary Table – Quality ratings of sample literature (independent rater)

Please begin by reading this instruction sheet in full before beginning the task. Then read 'Recommendations from Mohr (1995)' and ensure you fully understand this. This is the basis on which the rating scale has been developed and provides some context to the rating scale. Then read the 'Ratings for recommendations from Mohr (1995)' and ensure full understanding. If you are unclear on any point, please contact the author of the literature review, LT, to discuss.

Once you are sure you understand the task, begin reading the first article. Use the 'Summary Table' to record your ratings, and use the sheet entitled 'Ratings for recommendations from Mohr (1995),' in conjunction with 'Recommendations from Mohr (1995)' for information on the criteria being rated and the specifications required for each rating.

Thank you again for your participation.

Appendix J (iii) Summary Table – Quality ratings of literature (ordered chronologically)

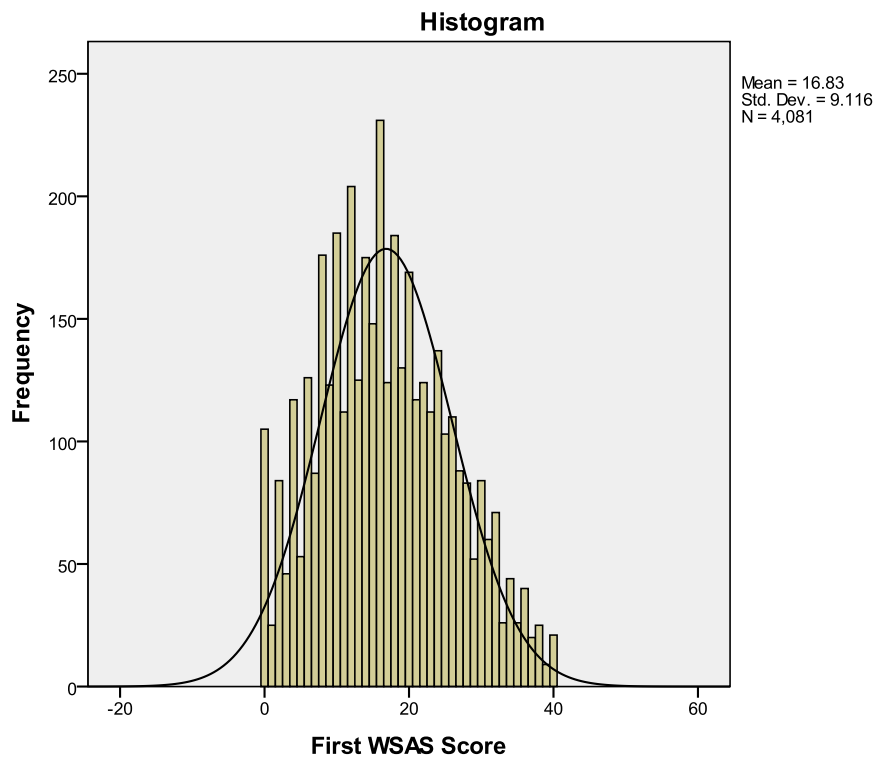
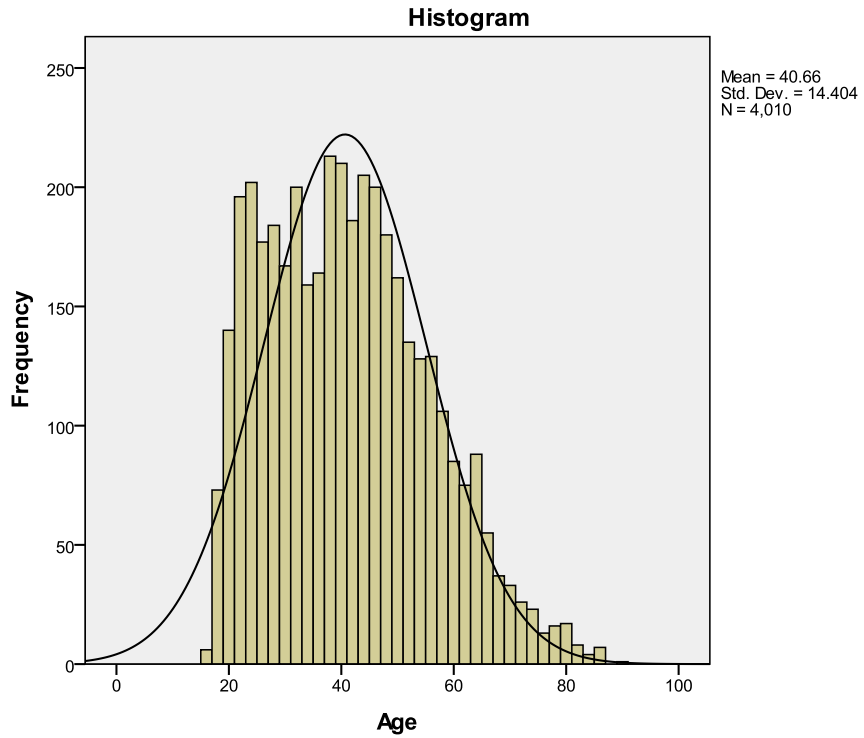
Article	Criterion A	Criterion B 1)	Criterion B 2)	Criterion C	Criterion D	Criterion E	Total Score (%)
Ogles, Lambert & Sawyer (1995)	3	3	3	2	2	3	81.5
Pekarik & Wolff (1996)	3	2	2	2	2	3	63.0
Scogin, Floyd, Jamison, Ackerson, Landreville & Bissonette (1996)	2	2	2	2	2	3	42.6
Ford, Fisher & Larson, (1997).	3	3	3	2	2	3	81.5
Samstag, Batchelder, Muran, Safran, & Winston, (1998.)	2	1	2	2	2	1	33.3
Pekarik & Guidry (1999)	3	3	3	2	3	3	90.7
Tarrier, Pilgrim, Sommerfield, Faragher, Reynolds & Graham et al. (1999)	1	3	3	2	2	1	51.9
Moos, Moos, & Finney (2001)	2	2	1	2	1	3	42.6

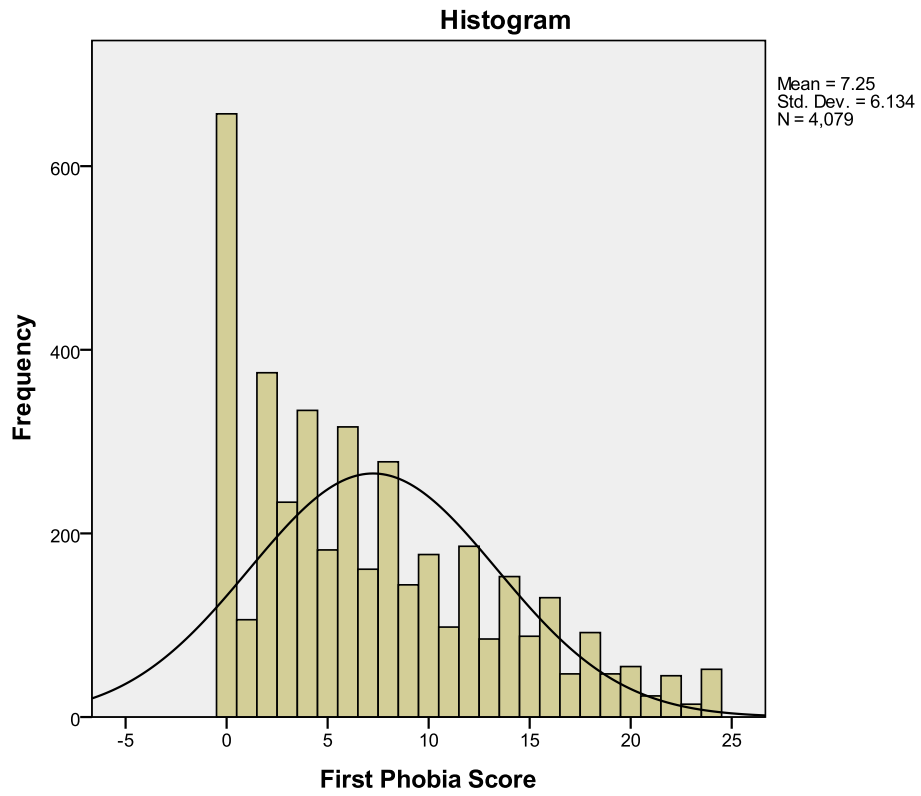
Moos, Nichol & Moos (2002)	2	2	1	2	1	3	42.6
Pike, Walsh, Vitousek, Wilson, & Bauer (2003).	2	2	1	1	3	1	37.0
Moritz, Fricke, Jacobsen, Kloss, Wein, Rufer et al. (2004)	2	2	3	2	2	1	48.1
Beutal, Hoflich, Kurth, Reimer (2005)	1	3	3	2	2	3	66.7
Lincoln, Rief, Hahlweg, Frank, Witzleben, Schroeder (2005)	2	3	3	2	2	3	72.2
Fricke, Moritz, Andresen, Jacobsen, Kloss, & Rufer et al. (2006)	2	3	3	2	2	1	57.4
Ilgen & Moos (2006)	1	3	2	1	2	3	51.9
Kanter, Landes, Busch, Rusch, Brown, & Baruch (2006)	1	2	2	2	2	1	33.3
Rufer, Fricke, Moritz, Kloss & Hand, (2006)	2	2	2	2	2	1	38.9
Kellett, Clarke & Matthews (2007)	3	3	3	1	2	3	75.9

Lunnen, Ogles, Pappas (2008)	3	1	2	3	2	3	59.3
Prestano, Lo Coco, Gullo, & Lo Verso (2008)	3	3	3	1	2	3	75.9
Samstag, Muran, Wachel, Slade, Safren & Winston (2008)	3	1	2	2	2	1	42.6
Von der Lippe, Monsen, Ronnestad, & Eilertsen (2008)	3	3	3	2	2	3	81.5
Callahan, Almstrom, Swift, (2009)	3	3	2	1	2	3	66.7
Fournier, DeRubeis, Shelton, Hollon, Amsterdam, & Gallop (2009)	2	2	1	1	2	1	27.8
Hartman, Orlinsky, Weber, Sandholtz, Zeeck (2010)	2	3	3	3	3	1	75.9
Jensen, Mortensen, & Lotz, (2010)	3	1	1	1	2	3	46.2

Lindgren, Werbart, Philips (2010)	3	2	3	2	2	3	72.2
Lopez-Goni, Fernandez-Montalvo, Menendez, Yudego, Garcia, Esarte (2010)	3	2	1	2	1	3	51.9

Appendix K – Histograms displaying Distributions for continuous variables





Appendix L – Output for analyses of interactions in logistic regressions

General Deterioration

(From logistic regression backwards LR)

		Exp (B)	Significant?
First Employment Code	PHQ sev or not	1.184 (1), p=0.852	Not Sig
	PHQ mod or not	0.845 (1), p=0.688	Not Sig
	Drop out or not	1.431 (1), p=0.444	Not Sig
	Sud det without gains	0.572 (1), p=0.237	Not Sig
Drop out or not	PHQ sev or not	B(1)=16.939, p=0.994	Not Sig
	PHQ mod or not	0.885 (1), p=0.789	Not Sig
	Sud det without gains	.0588 (1), p=0.324	Not Sig