Can Therapists Estimate Current Patient Progress and Predict Final Outcomes with the Provision of Feedback?

Nicola K Flood
Bachelor of Arts (Honours), MPsysch (Clinical)

This thesis is presented for the degree of Doctor of Philosophy, and in partial fulfilment of the requirements for the Master of Psychology (Clinical) degree, at The University of Western Australia

School of Psychological Science
2018
THESIS DECLARATION

I, Nicola Kim Flood, certify that:

This thesis has been substantially accomplished during enrolment in the degree.

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

The research involving human data reported in this thesis was assessed and approved by The University of Western Australia Human Research Ethics Committee. Approval #: RA/4/1/6423

Written patient consent has been received and archived for the research involving patient data reported in this thesis.

The work described in this thesis was funded by the ARC Linkage grant LP150100503.

This thesis contains work prepared for publication, some of which has been co-authored.

Signature: [blank]

Date: 4\textsuperscript{th} April 2018
ABSTRACT

Cognitive Behaviour Therapy (CBT) is effective in treating the majority of patients, with a variety of disorders. However, despite evidence of CBT’s efficacy and effectiveness, a substantial proportion of patients do not improve following CBT, and a smaller proportion get worse. To address the problems of no change and deterioration in real time, we need to identify, during treatment, how patients are progressing. Unfortunately, therapists tend to overestimate positive progress and outcomes while underestimating no change and deterioration. Further, empirical methods of prediction, which use statistical methods to predict outcomes, although superior to therapists, are still imperfect. Feedback of the results of routine outcome monitoring to the therapist alone, and to both the therapist and patient, improves clinical outcomes for patients at risk for treatment failure. However, to our knowledge, no studies assessing therapists’ judgements of progress and outcomes include feedback to either the therapist alone or to both the therapist and patient. Thus, a primary aim of the present thesis was to determine whether feedback can improve therapists’ estimates of current patient progress and predictions of final outcomes. A secondary aim was to determine whether an empirical method of prediction would be superior to therapists at predicting no change and deterioration.

The first study of this thesis (Chapter 2) assessed the effects of feedback to the therapist only, on therapists’ estimates of progress and predictions of outcomes. Consistent with previous research, it was found that therapists overestimated positive and underestimated negative progress and outcomes overall. However, therapists correctly predicted the proportion of patients who recovered and deteriorated. Therapists also correctly estimated the progress of 50% of the deteriorated patients after
feedback was provided to them. The empirical method used in this study was superior to therapists at predicting no change and deterioration.

In the second study (Chapter 4), therapists were asked to make their estimates and predictions following feedback to both the therapists and patients. Further benefits of feedback on therapists’ judgements of progress and outcomes were found in this study. It was found that therapists accurately predicted the proportion of patients who would go on to achieve a positive outcome, remain unchanged, or deteriorate in their symptoms. Therapists also accurately estimated the proportion of patients who had made positive progress, remained unchanged, or deteriorated for symptoms, at the time of the estimates. However, agreement between therapists’ predictions/estimates and actual outcomes/progress was less accurate for the unchanged and deteriorated categories. Inconsistent with past research was that the empirical method was not necessarily superior to therapists at predicting outcomes, in that therapists outperformed the empirical method of prediction in some areas.

The results from the first two studies suggest that the empirical method was not necessarily superior to therapists at predicting outcomes. Further, although feedback appeared to assist therapists’ judgements of progress and outcomes, the results suggested that therapists still require assistance to identify and predict negative progress and outcomes. Therefore, the final study of this thesis (Chapter 6), assessed whether early change as defined by longitudinal modelling could predict negative outcomes more effectively than the clinical significance early change method, which was used in the first two studies. It was found that the simpler clinical significance early change method was superior at predicting negative outcomes than longitudinal modelling.

Feedback was found to assist therapists in their judgements of progress and outcomes and the simpler clinical significance early change method was superior at
predicting outcomes compared to longitudinal modelling. However, therapists and the clinical significance early change method are not perfect at predicting negative outcomes. Therefore, future research is needed to develop and test better methods of identifying negative progress in order to prevent future negative outcomes.
# TABLE OF CONTENTS

Acknowledgements ........................................................................................................... ix
Authorship Declaration ........................................................................................................ xi
Preamble to the Thesis .......................................................................................................... xiii

Chapter

1. General Introduction ........................................................................................................ 1
   A. The Rise of Patient-Focused Research ................................................................. 4
      i. Clinical significance .......................................................................................... 6
      ii. Negative outcomes in psychotherapy ............................................................. 7
   B. Therapists’ Estimates of Current Progress and Predictions of Outcomes ... 9
   C. Empirical Predictions of Outcomes ..................................................................... 12
   D. Routine Outcome Monitoring and Feedback ..................................................... 13
   E. Thesis Aims and Overview ................................................................................ 23

   A. Abstract .................................................................................................................. 30
   B. Introduction ............................................................................................................. 31
   C. Method ...................................................................................................................... 37
   D. Results ...................................................................................................................... 42
   E. Discussion ................................................................................................................. 48

3. The Addition of Feedback to the Patient and Clinical Significance
   Categorisation Issues ................................................................................................. 55

4. Therapists’ Ability to Judge Patient Outcomes and Progress with the Provision of Feedback to both the Therapist and Patient .................................................... 61
   A. Abstract .................................................................................................................. 62
ACKNOWLEDGEMENTS

I would firstly like to express sincere gratitude to my supervisor, Andrew Page. Thank you for being so approachable and engaging. Your wisdom, creative way of thinking, timely feedback, and overall support has helped me immensely throughout this experience. You constantly went out of your way to ensure I was making good progress (including reading drafts while travelling on planes, trains etc.). I feel so lucky to have had such a dedicated and helpful supervisor who made this experience a very rewarding and enjoyable one. Thank you also to the University of Western Australia for providing me with the University Postgraduate Award (UPA) and UWA Safety Net Top-Up scholarships.

I would also like to sincerely thank Geoffrey Hooke, Moira Monroe, and the clinical and research teams at Perth Clinic. Thank you for your continued support and for providing the funding for my research with the ARC Linkage grant as well as providing me with the Perth Clinic Ad Hoc Top-Up Scholarship. Geoff, your hilarious lunch time “high-brow” conversations kept me positive through tough times. A special mention also to Shraddha Kashyap and Gaby Palacios for contributing to these silly lunch time conversations and always picking me up when I felt overwhelmed. Moira, thank you for always being a strong supporter of research at Perth Clinic and providing me with the data, desk space, and of course delicious scones and occasional 5-star dining experiences. A big thank you also to the therapists who always showed an interest in my research and were happy to complete my questionnaires each fortnight, despite my pestering.

Thank you to my parents and their partners, my sister (Angela) and brother (Cameron) who have all provided me with unconditional love and support, and of course delicious meals. Mum, you basically taught me how to write, and our regular
Monday dinners and chats really helped me through any obstacle I was faced with. Dad, thanks for always being there for me and encouraging me to aim high, your support means so much. Angela, although you live in another city, I have always felt loved and supported by you. Our regular phone chats filled with laughs and silliness always brightened my day. Cam, thanks for making me laugh and always believing in me. To my partner, Gus; thank you so much for your patience, encouragement, love, support, and selflessness. I really appreciate you making changes to your work schedule so that you could be here for me during tough and stressful times. Your help getting me over the finish line means so much.

I would like to thank all of my friends and housemates over the years, especially Bree Carter, Nikki Graham, Kate Miller, Bec Kieft, Lauren Hearty, and Gab Barnes. Thank you for all of the fun distractions from postgrad life and being there for me during the difficult and stressful times. Thank you to all of my uni friends, especially my office buddy and arch nemesis Elise Hartley. Thanks also to Matthew Herbert, Georgina Mann, Stephanie Stevens, Henry Austin, and Blake. I would not have been able to complete this journey without your support, reassurance, advice, and of course all of the laughs. You really kept me sane throughout this experience (or did you; you kewl mules). Thanks also to my wonderful grandparents (Eddy and Joan) who provided me with lots of encouragement, love, and support. Thanks to all of the Beatty Park Leisure Centre lifeguard crew and finally, thank you to everyone involved in the Claremont Women’s Football Club. Footy has been great for keeping my stress levels in check and letting off steam with a bunch of amazing girls. So thank you to all the footy girls, coaches, and volunteers at the club for your comradery and support throughout this experience.
AUTHORSHIP DECLARATION

This thesis contains work that has been prepared for publication.

**Details of the work:**
Submitted revisions to the “Journal of Clinical Psychology” as a research paper on 29th January 2018: Can Therapists Predict Patient Outcomes and Estimate Progress in Group Cognitive Behaviour Therapy when Feedback is Available to the Therapist?

**Location in thesis:**
Chapter 2, Study 1

**Student contribution to work:**
The data were collected as part of routine hospital data collection at Perth Clinic and provided to the student by Geoffrey Hooke.
The data were formatted to allow analyses that were analysed, interpreted and presented by the student.
The literature review, method, results, and discussion sections were written by the student and feedback on content was provided by Andrew Page (principal supervisor). Feedback was also provided by international researchers, Wolfgang Lutz and Michael Barkham. Wolfgang and Lutz and Michael Barkham are located outside of Australia so we were unable to obtain their signatures.

Co-author signatures and dates:

---

**Details of the work:**
Submitted revisions to the “Psychotherapy Research” journal as a research paper on 8th March 2018: A Comparison between the Clinical Significance and Growth Mixture Modelling Early Change Methods at Predicting Negative Outcomes

**Location in thesis:**
Chapter 6, Study 3

**Student contribution to work:**
The data were collected as part of routine hospital data collection at Perth Clinic and provided to the student by Geoffrey Hooke.
The data were formatted to allow analyses that were analysed, interpreted and presented by the student.
The literature review, method, results, and discussion sections were written by the student and feedback on content was provided by Andrew Page (principal supervisor).

Co-author signatures and dates:

---

Student signature: [Blacked out] Date: 4th April 2018
I, Andrew Page, certify that the student’s statements regarding their contribution to each of the works listed above are correct.

Coordinating supervisor signature: [Redacted]

Date: 4th April 2018
PREAMBLE TO THE THESIS

Consistent with the University of Western Australia’s guidelines for PhD submission, this thesis is presented as a series of papers. Studies 1 and 3 (Chapters 2 and 6) have been prepared and submitted for publication (see Authorship Declaration for details). Presentation of this thesis as a series of papers may lead to some repetition and will require some linking chapters to develop the overall argument. The linking chapters have not been submitted for publication. Figures and tables have been inserted into the text to facilitate reading.
Chapter 1

General Introduction
Chapter 1: General Introduction

CHAPTER 1

General Introduction

Decades of research has shown that Cognitive Behaviour Therapy (CBT) is an effective treatment for many patients experiencing a variety of disorders (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012; Westbrook & Kirk, 2005). However, a patient-focused research approach has drawn attention to the fact that some patients make no change during CBT and others even get worse (Lambert, 2013b). Patient-focused research encourages attention to be allocated to the progress of individual patients because these patterns may be masked by group averages. The absence of change for some patients is a problem not only for the patients themselves, but for mental health services more generally. If patients are leaving treatment unchanged or worse off than when they entered, more pressure is placed on mental health services in terms of waitlists and cost. Agencies who fund mental health services want evidence-based treatments that are effective and efficient (McHugh & Barlow, 2010; Page & Stritzke, 2014) and patients themselves want to be treated effectively in the shortest time.

To address the problems of no change and deterioration, we need to be able to identify, during treatment, how patients are progressing. One way of doing this would be to ask the therapist. Unfortunately, therapists appear to be poor at both identifying and predicting no change and deterioration. Specifically, it appears that therapists tend to overestimate positive progress and underestimate no change and deterioration, and the same pattern is observed when therapists are asked to predict final outcomes (Chapman et al., 2012; Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010). Therefore, patient-focused researchers have explored methods of improving outcomes for patients at risk for a negative outcome.
An important premise of the attempts to target appropriate treatment towards patients who are not changing or deteriorating, is the observations that empirical methods are superior to clinical judgements. Empirical methods of prediction use specific statistical techniques performed on self-report measures to predict the outcomes of patients (Spielmans, Masters, & Lambert, 2006). For example, Hannan et al. (2005) found that an empirical method identified all of the patients (100%) who had deteriorated by the end of treatment, whereas therapists correctly identified none. However, the empirical method heavily overpredicted deterioration, therefore this method was still imperfect. Another method of improving outcomes for patients at risk for treatment failure is the use of routine outcome monitoring and providing the monitoring outcomes to therapists as feedback about patient progress. Routine outcome monitoring typically involves the repeated administration of reliable and valid measures that the patient completes at each session (Howard, Moras, Brill, Martinovich, & Lutz, 1996). Feedback involves providing this information in some way to either the therapist alone or to both the therapist and patient. Lambert and colleagues examined the effects of feedback on two different groups of patients; patients who were not on track and patients who were on track. Not on track patients can be described as patients who are at risk of making no change or deteriorating, while on track patients are those patients who are on track for a positive outcome. Feedback has been found to reduce rates of no change and deterioration, at least among those who were not on track for an optimal outcome (Lambert, Whipple, et al., 2001; Newnham, Hooke, & Page, 2010b; Shimokawa, Lambert, & Smart, 2010).

Thus, there are two lines of research that have not yet been brought together. On the one hand, some studies have investigated the effects of feedback; while the others have examined the effectiveness of therapists’ abilities to detect not on track patients.
That is, the studies that assessed therapists’ estimates of current progress and predictions of outcomes did not include feedback to the therapists. Therefore, the present thesis asks a series of questions. First, can feedback improve therapists’ estimates of current progress and predictions of outcomes? If so, then upon identifying no change or deterioration during treatment, therapists could initiate some form of action during therapy to guide the patient on track for a positive treatment outcome. If not, then can we identify a better way of detecting and predicting no change and deterioration in order to reduce rates of negative outcomes?

**The Rise of Patient-Focused Research**

In the early days of psychotherapy research, researchers sought to determine whether a specific treatment worked by comparing the improvements of a sample of patients, often with a single diagnosis, who receive the treatment of interest to a sample of patients with the same diagnosis who receive a different treatment, no treatment, or a control treatment (Lambert, Hansen, & Finch, 2001). The active treatment is deemed efficacious if the sample of patients receiving the treatment of interest, on average, improve more than the sample of patients in the control condition. This type of research is typically called efficacy research and is frequently conducted using Randomised Controlled Trials (RCTs) (Chambless & Hollon, 1998). The focus in this efficacy research is on average patient change, as a result of the treatment, within ideal circumstances (Howard et al., 1996). Once the treatment has shown positive effects in a large number of studies, including meta-analyses, the treatment can move to the next step. The clear next step for determining whether a new treatment works was to test this new treatment out in a “real world” clinical setting. This type of research is typically called effectiveness or practice-based research and aims to gain ecological validity for the treatment of interest (Lambert, 2001). Effectiveness research, like efficacy research,
is still focused on average patient change, as a result of the treatment, but within a naturalistic setting (Howard et al., 1996). Once the treatment has been through both of these stages, it may not be deemed evidence-based practice; rather, evidence-based practice is accomplished by developing a body of research that meets certain well-specified criteria (Satterfield et al., 2009). Further, in order for the treatment to remain “evidence-based practice”, therapists must, using their clinical expertise, take into account patient characteristics, culture, and preferences. Further, therapists must frequently monitor their treatments’ effectiveness on each of their individual patients to ensure that it remains evidence-based (Chambless & Hollon, 1998).

One such treatment that has been through a large number of efficacy and effectiveness trials is Cognitive Behaviour Therapy (CBT), which was pioneered by A. T. Beck (1970) and Ellis (1962). CBT posits that psychiatric disorders such as depression develop due to unhelpful cognitions and maladaptive behaviours which perpetuate emotional distress and behavioural issues (Hofmann et al., 2012). CBT is understood to work by challenging unhelpful cognitions, increasing positive behaviours, and modifying maladaptive behaviours, thus leading to improvements in mood and wellbeing (A. T. Beck, 1970; Ellis, 1962). Efficacy and effectiveness studies have shown CBT to be effective in treating a number of different psychiatric disorders in a variety of settings (Hofmann et al., 2012; Lambert, 2013a).

The problem with concluding that a treatment such as CBT is effective, after exposing it to efficacy and effectiveness trials, is that the focus is on average patient change. An average generally tells us about the majority of the patients depending on the distribution of scores but not, for instance, those few patients who have worsened, or the larger group who have made no significant change (Newnham & Page, 2007). Therefore, the therapist, as an evidence-based practitioner, might be asking “how is my
individual patient progressing, after today’s session?” Evidence-based practice in psychology requires that researchers continue to search for better methods to help individual patients, and therapists monitor patient outcomes in order to strive for continuous improvement and evidence-based practice. Therapists should then evaluate their own interventions, and report these findings to the professional and scientific communities (Hayes, Barlow, & Nelson-Gray, 1999; Page & Stritzke, 2014). The scientist-practitioner model therefore goes hand in hand with patient-focused research which aims to improve the outcomes of individual patients by monitoring their progress (Newnham & Page, 2007). The process of monitoring individual patients’ progress and outcomes in patient-focused research brought to light the fact that not all patients benefit from treatment. The research suggests that while some patients experience no change, others get worse (deteriorate) as a result of treatment (Lambert, 2013b). Researchers therefore needed a method to define these different types of patients’ outcomes.

Clinical significance. Clinical significance was developed in response to perceived flaws associated with the traditional methods used in efficacy and effectiveness research: comparing average pre- and post-treatment scores. An issue with the traditional pre-post comparison is that it focuses on average patient change and as such, does not take into account the different patterns of change between patients (e.g., no change and deterioration). Further, statistical significance does not necessarily equate to a meaningful clinically significant change (Jacobson & Truax, 1991). While there are a variety of different methods of calculating clinical significance (Lambert, Hansen, & Bauer, 2008; Ronk, Hooke, & Page, 2012), the Jacobson-Truax clinical significance method is arguably the most popular method for defining clinical outcomes (Jacobson,
This method classifies patient outcomes based on two criteria:

1. Whether the patient has moved from a pre-determined dysfunctional population range to the functional population range by the end of treatment and;
2. Whether the change was statistically significant (Ronk, Korman, Hooke, & Page, 2013).

Combining these two criteria produces a minimum of four clinical significance categories for those patients beginning treatment within the dysfunctional range:

1. **Recovered:** the patient has moved from the dysfunctional to the functional population range and the change was significant.
2. **Improved:** The patient remains in the dysfunctional population range but the change was positive and significant.
3. **Unchanged:** The patient remains in the dysfunctional population range and the change was not significant.
4. **Deteriorated:** The patient remains in the dysfunctional population range and has made a significant negative change (Jacobson & Truax, 1991; Ronk et al., 2013).

Recovered and improved outcomes could therefore be considered a positive outcome; and unchanged and deteriorated outcomes could be considered a negative outcome.

**Negative outcomes in psychotherapy.** Rates of no change and deterioration vary from study to study. For example, Cahill, Barkham, and Stiles (2010) conducted a systematic review of practice-based research conducted in primary care settings in the United Kingdom (UK). The authors found that rates of unchanged patients ranged from 19% to 31% with a mean of 25.7% and deterioration rates ranged from 1% to 3% with a
mean of 1.5% (Cahill et al., 2010). A study conducted in secondary care settings in the UK found that 40% of patients remained unchanged and 6% deteriorated following treatment (Barkham et al., 2001). The rates for these two negative outcomes tend to be higher within the United States (US). For example, Hansen, Lambert, and Forman (2002) found a mean rate of 56.8% ranging from 45.6% to 60.7% for unchanged patients, and a mean rate of 8.2% ranging from 3.2% to 14.1% for deteriorated patients, across a variety of different clinical settings in the US.

Newnham, Harwood, and Page (2007) examined the rates of no change and deterioration in 1830 patients from a private psychiatric hospital in Australia. The rates were more comparable to the UK than the US rates, as the authors found that 43.6% of patients made no change and 1.7% deteriorated (Newnham et al., 2007). A recent study conducted by Saxon, Barkham, Foster, and Parry (2017) examined the rates of deterioration in a large sample of patients (N = 10521) from various counselling and clinical psychology services in the UK. The authors used the CORE Assessment form (CORE-OM) (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005) to assess outcomes and found that the rate of reliable deterioration was less than 1% of the whole sample. However, the rate of any deterioration (i.e., any negative shift in score on the CORE-OM) was approximately 11%. This finding is consistent with previous research in the UK and other countries (Cahill et al., 2010; Hansen et al., 2002; Lambert, 2010; Newnham et al., 2007; Newnham & Page, 2007). Despite the varying rates of no change and deterioration, researchers tend to agree that the rate of no change is usually around 50% and rates of deterioration usually range from 5% to 10% (Lambert, 2010; Newnham & Page, 2007; Saxon et al., 2017).
Therapists’ Estimates of Current Progress and Predictions of Outcomes

Regardless of the rates of no change and deterioration, patients with these negative outcomes exist within mental health services. Therefore, methods of detecting and subsequently preventing negative outcomes need to be identified. As previously noted, researchers first looked to the professionals and experts in the field of psychotherapy; the therapists. Hatfield et al. (2010) for example, investigated therapists’ ability to detect patient deterioration by examining the therapists’ progress notes over the course of treatment. Deterioration was defined based on the reliable change index of the OQ-45, which meant that deterioration was flagged if the patient’s OQ-45 score worsened by 14 points or more since intake. It was found that therapists mentioned deterioration in the progress notes of only 21% of patients who had shown deterioration at some point in treatment. This suggests that the therapists either did not notice the deterioration in the other 79% of patients, or they failed to mention it in their progress notes. Progress notes should include information about the patient’s response to treatment and any changes in the patient’s symptoms and/or wellbeing (Perkinson, 2016). Therefore, it is more likely that these therapists were not aware of deterioration in the majority of the deteriorated patients.

Hannan et al. (2005) also assessed therapists’ ability to estimate current patient progress. The authors asked therapists to decide whether their patient was recovered, improved, unchanged, or deteriorated (clinical significance categories) as of “today’s” session, using their clinical judgement. Therapists were found to be optimistic in their estimates of current patient progress. Specifically, therapists overestimated the proportion of patients who had made positive progress and underestimated the proportion of patients who had made no change or had deteriorated. This is consistent
with the findings of Hatfield et al. (2010) who found that patients underestimated deterioration.

Despite therapists’ poor performance in identifying current patient progress, researchers have also assessed therapists’ ability to predict patient outcomes. For example, in a study conducted by Hannan et al. (2005), therapists were asked to predict their patients’ outcomes by deciding whether their patient would be recovered, improved, unchanged, or deteriorated by the end of treatment. Not surprisingly, the authors found that therapists were optimistic in their predictions of patient outcomes as well. Specifically, therapists predicted that the majority of patients would achieve a positive outcome, while predicting that only 0.01% would deteriorate. It could be reasonable to assume that therapists may not have been aware that patient deterioration occurs, and of the typical rates of deterioration. However, therapists were informed before participating in the study that the rate of deterioration at the clinic had been consistently around 8%.

Chapman et al. (2012) conducted a similar study in which therapists were asked to predict their patients’ outcomes in a group therapy setting. Therapists were required to predict patient outcomes based on whether they believed the patient would leave treatment reliably improved, having made no reliable change, or reliably worse. Consistent with the findings from Hannan et al. (2005), therapists overpredicted the proportion of patients who would achieve a positive outcome and underpredicted the proportion who would make no change or deteriorate. Also consistent in both the Hannan et al. (2005) and Chapman et al. (2012) is that the patients who therapists predicted would deteriorate, usually did not end treatment with a deteriorated outcome. Therefore, there was little agreement between therapists’ predictions of outcomes and actual patient outcomes. These findings are consistent with a broader literature that has
shown that therapist experience made little to no difference to therapists’ prediction accuracy (Grove & Meehl, 1996), and when therapists were asked to predict outcomes later in therapy, the results were also poor (Breslin, Sobell, Sobell, Buchan, & Cunningham, 1997).

Researchers have then asked the question; why are therapists poor at identifying patient progress and predicting patient outcomes? Some researchers suggest that therapists may be optimistic about patient outcomes because they are no less prone to cognitive biases than the average person. For example, Walfish, McAlister, O'Donnell, and Lambert (2012) suggested that therapists have a self-assessment bias, which is the inflated belief in one’s own abilities (Myers & Ridl, 1979). The authors found that on average, therapists rated themselves at the 80th percentile compared to other therapists, while only 8.4% rated themselves as below average. This means that despite therapists being told about patient deterioration in the Hannan et al. (2005) study, they may have assumed that these patients were being treated by other therapists, not themselves. Walfish et al. (2012) also found that therapists believed that 77% of their patients improved and just under half of the therapists indicated that none of their patients deteriorated. A limitation of this study however, was that the authors did not compare therapists’ estimates/predictions with actual patient outcomes. Therefore, therapists’ judgement accuracy is unclear. Based on typical rates of negative outcomes however, if the comparison between therapists’ predictions and actual outcomes was made, it is likely that therapists would have overpredicted positive outcomes and underpredicted deterioration.

The research in the area of therapists’ judgements of progress and outcomes is relatively new, and therefore limited. However, one could conclude from the results of the above studies that therapists believe most of their patients will achieve a positive
outcome, and struggle to identify negative progress and predict those patients who will leave treatment either unchanged or deteriorated. One could argue that it is reasonable for therapists to think optimistically about patient outcomes due to certain cognitive biases/heuristics. For example, therapists may have a self-assessment bias, the inflated belief in their own ability to treat their patients, which could explain their optimism (Walfish et al., 2012). Therapists may also think optimistically regarding their patients’ outcomes because the results of efficacy and effectiveness research suggest that the majority of patients receiving CBT will achieve a positive outcome (Hofmann et al., 2012; Westbrook & Kirk, 2005). In fact, it would also be reasonable for therapists practicing other evidence-based treatments to be optimistic about their patients’ outcomes, because other evidenced-based treatments have similar outcomes to CBT (Wampold et al., 2017). Despite it being reasonable for therapists to expect positive outcomes, some patients remain unchanged or get worse following treatment. Therefore, another method of identifying and predicting negative progress and outcomes is needed.

Empirical Predictions of Outcomes

As a result of the limitations of therapists’ judgements of progress and outcomes, researchers have looked to empirical methods of identifying and predicting no change and deterioration, and some have compared these methods to therapists’ abilities. Comparisons of clinical versus statistical (empirical) predictions of outcomes date as far back as the 1950s, e.g., Meehl (1954). Since then, studies have consistently found that empirical methods of prediction are typically as accurate or superior to clinical predictions (Grove, 2005; Grove, Zald, Lebow, Snitz, & Nelson, 2000). For example, Hannan et al. (2005) compared therapists’ predictions of outcomes to an empirical method. This empirical method divided patients into groups based on their
initial severity. The empirical method identified 50 groups based on intake scores and each group contained at least 220 patients (approximately 2% of the total sample). The data were then analysed to generate a linear model with tolerance intervals for expected treatment response. It was found that the majority of patients (≈70%) predicted by the empirical method to deteriorate, did not. However, the majority of these patients were unchanged (74%), which would be perceived as a negative outcome. Further, all of the patients (100%) who had deteriorated by the end of treatment were identified by the empirical method.

The results from Hannan et al. (2005) suggest that in contrast to therapists, the empirical method overpredicted the proportion of patients who would deteriorate, but correctly identified all of the patients who deteriorated. Therefore, the empirical method was superior to therapists at predicting deterioration. Although the empirical method overpredicted deterioration, the authors argued that as most of these patients still had a negative outcome (unchanged), these patients would still be at risk for treatment failure (Hannan et al., 2005). An empirical method that can predict negative outcomes during psychotherapy is a useful tool for therapists. Yet, what should therapists do once they become aware that a patient is predicted to have a negative outcome, in order to prevent this negative outcome from occurring?

**Routine Outcome Monitoring and Feedback**

A group of researchers introduced the concept of providing feedback of routine outcome monitoring of patient progress, to therapists, as an integral aspect of patient-focused research (Howard et al., 1996). Routine outcome monitoring is the repeated administration of a reliable and valid measure, and feedback is simply the provision of this information to either the therapist alone or to both the therapist and patient (Hawkins, Lambert, Vermeersch, Slade, & Tuttle, 2004; Howard et al., 1996). There are
four possible options of what can be done when routine outcome monitoring is applied in a clinical setting. The first option is that neither the therapist nor patient views their scores on the measure of interest. The second option is that only the therapist views the patients’ scores on the measure. The third option, and the most unlikely, is that only the patient views their scores on the measure. The final option is that both the therapist and patient view their scores on the measure. The second and fourth options are the most commonly applied in the literature on routine outcome monitoring and feedback. As such, the feedback literature will be reviewed firstly by summarising the studies in which only the therapist received feedback and secondly by summarising the studies in which both the therapist and patient received feedback.

One of the initial studies to assess the effects of feedback to the therapist only, was conducted by Lambert, Whipple, et al. (2001). The authors of this important study randomly allocated patients from a university counselling centre to either a feedback or no feedback condition. Patients were invited to complete a measure assessing their symptoms and wellbeing (Outcome Questionnaire-45; OQ-45) (Lambert et al., 1996) before each session, and only the therapists received feedback on their patients’ scores. Therapists whose patients were in the feedback condition received a progress graph of their patients’ OQ-45 scores which included colour-coded feedback. White feedback indicated that the patient was functioning within the normal range and the therapist should consider termination. Green indicated that the patient was making change within the adequate range. Yellow indicated that the patient’s change was less than adequate, and suggested that the therapist should consider altering the treatment plan; and red indicated that the patient was not making the expected level of progress and suggested the therapist should carefully review the case and alter the treatment plan. The red and yellow feedback alerted therapists to patients who were not on track for a positive
outcome, whereas white and green feedback indicated that the patient was *on track* (Lambert, Whipple, et al., 2001).

Patients in the feedback condition, who were also not on track for a positive outcome, experienced significantly improved outcomes (in the period after they had received feedback) by termination, compared to not on track patients in the no feedback condition (Lambert, Whipple, et al., 2001). One aspect that was of interest was the observation that following feedback, the not on track patients received significantly more treatment sessions than not on track patients in the no feedback condition. However, this increase in the amount of treatment delivered was offset because the on track patients in the feedback condition attended fewer sessions than on track patients in the no feedback condition. The allocation of the treatment towards the not on track patients did not affect their outcomes. The authors argued convincingly that feedback therefore led to more efficient allocation of resources. This is because on track patients received adequate treatment within a shorter time frame, while not on track patients attended more sessions, thus reducing drop-out and improving their outcomes.

The results from Lambert, Whipple, et al. (2001) were replicated by the same research group a year later using a larger sample from the university counselling centre (Lambert, Whipple, Vermeersch, et al., 2002). It was found that feedback led to improved outcomes for not on track patients but did not have an impact on the amount of sessions patients attended (Lambert, Whipple, Vermeersch, et al., 2002). Connolly Gibbons et al. (2015) more recently assessed the effects of feedback to the therapist alone but did not distinguish between not on track and on track patients. The authors found that feedback improved outcomes overall for outpatients receiving treatment for depression.
In fact, the two initial studies conducted by Lambert and colleagues (Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002) included patients primarily diagnosed with mood or anxiety disorders. Therefore, Lambert and colleagues also investigated the effects of feedback to the therapist only in more specific patient groups. For example, Crits-Christoph et al. (2012) examined the effects of feedback to the therapist alone on the outcomes of patients treated at community-based substance abuse treatment programs. All patients were required to complete an adaptation of the OQ-45 which included drug and alcohol use items. Patients identified as not on track and in the feedback condition however, were required to complete a second questionnaire assessing problems that may be impeding treatment progress. Therapists were trained to then respond in a certain way based on what problems may be getting in the way of progress. It was found that feedback to the therapist alone led to reduced alcohol use, OQ-45 scores, and drug use for not on track patients. Another study investigated the effects of feedback to the therapist alone on patients with an eating disorder diagnosis in an inpatient setting (Simon et al., 2013). It was found that patients in the feedback condition achieved clinically significant change more frequently than patients in the no feedback condition. This suggests that feedback is beneficial in improving outcomes for patients with a variety of psychological disorders.

Another study that examined the effects of feedback to the therapist only, identified an unexpected condition in which therapists did not use the feedback (de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012). In other words, the first option of what can be done when routine outcome monitoring is applied in a clinical setting, in which neither the therapist or patient views the patients’ scores on the measure. The authors found that overall, there was no significant difference between the feedback and no feedback conditions in outcomes. There was also no significant difference in
outcomes between the feedback and no feedback conditions when not on track patients were assessed separately. However, feedback led to superior outcomes for not on track patients whose therapists indicated that they had used the feedback compared to not on track patients in the no feedback condition (de Jong et al., 2012).

Researchers have also examined the effects of feedback to the therapists on youths’ clinical outcomes. It was found that regardless of whether the youth was on track or not on track for improvement, youths whose therapists received feedback improved more rapidly than youths whose therapists did not (Bickman, Kelley, Breda, de Andrade, & Riemer, 2011). A more recent study conducted by Bickman et al. (2016) assessed the effects of feedback on youths attending two different outpatient clinics. The authors found that improvements to the youths’ outcomes as a result of feedback were found in only one (Clinic R) of the two (Clinic U) outpatient clinics. To investigate this further, Bickman and colleagues decided to assess any differences in the implementation of the feedback intervention between the clinics. It was found that Clinic U, the clinic that did not find an effect of feedback, had 50% lower scores of implementation than Clinic R. This means that half the amount of therapists at Clinic U were actually implementing the feedback intervention compared to Clinic R. In summary, feedback to the therapist alone, when the therapist uses the feedback, appears to consistently improve outcomes for patients with a variety of psychological disorders who are not on track for improvement.

Although the literature suggests that feedback to the therapist alone improved outcomes for not on track patients, some of these patients did not achieve clinically significant improvement at termination. Secondly, although on track patients needed fewer sessions when feedback was involved, these patients did not achieve improvements in outcomes as a result of feedback (Lambert, Whipple, et al., 2001;
Lambert, Whipple, Vermeersch, et al., 2002). In order to address the insufficient impact of feedback to the therapist only, Lambert and colleagues pursued two avenues of research, one of which is relevant to the present research at this stage. The relevant approach that Lambert and colleagues investigated was to determine whether feedback to both the therapist and patient would improve outcomes over and above providing feedback to the therapist alone.

Hawkins et al. (2004) assessed the effects of feedback to the therapist and patient at a hospital-based clinic by assigning patients to one of three treatment conditions: no feedback, therapist feedback, or therapist and patient feedback. Patients were assigned using a randomised block design so that therapists were equally represented across the three treatment conditions. It was found that patients in the therapist and patient feedback condition achieved significantly superior outcomes as compared to patients in the therapist feedback and no feedback conditions. Contrary to the previous studies, feedback to the therapist and patient was effective in improving outcomes for both on track and not on track patients, suggesting that feedback to both the therapist and patient has more global effects than feedback to the therapist alone (Hawkins et al., 2004).

A more recent study conducted by de Jong et al. (2014) also compared the effects of feedback to the therapist only and to the therapist and patient, to a no feedback condition. The authors compared feedback effects on both short- and long-term therapy. Feedback to both the therapists and patients who were not on track for improvement in short-term therapy was found to be the most beneficial intervention. A small positive effect was also found in long-term therapy when feedback was provided to both the therapist and patient (de Jong et al., 2014). The results from these two studies suggest that providing feedback to both the therapist and patient has benefits
over and above the provision of feedback to the therapist alone (de Jong et al., 2014; Hawkins et al., 2004).

Other studies have found that feedback improved outcomes for both not on track and on track patients. For example, Amble, Gude, Stubdal, Andersen, and Wampold (2015) found that patients from six psychiatric clinics in Norway, who received feedback, achieved significantly superior outcomes compared to patients in the no feedback condition. Another study assessed the effectiveness of feedback provided to the therapist and patient at an outpatient clinic in Sweden (Hansson, Rundberg, Österling, Öjehagen, & Berglund, 2013). The authors found that patients in the feedback condition experienced larger reductions in total OQ-45 scores. However, this reduction was not statistically significant and there were no differences between not on track and on track patients in terms of the effect of feedback on outcomes (Hansson et al., 2013). A number of other studies that did not necessarily distinguish between not on track and on track patients found an overall positive effect of feedback to the therapist and patient (Murphy, Rashleigh, & Timulak, 2012; Reese, Norsworthy, & Rowlands, 2009; Simon et al., 2013).

Researchers have also assessed the effects of feedback on patients participating in group treatment. For example, Davies, Burlingame, Johnson, Gleave, and Barlow (2008) assessed the effects of feedback to both the patient and therapist on the outcomes of patients participating in group treatment. Interestingly, the authors found that in this group treatment setting, feedback had little impact on patient outcomes. In fact, there was even a small negative effect on outcomes for patients who reported that their group was high in conflict. Newnham et al. (2010b) compared the outcomes of a historical cohort of patients who participated in group CBT and received no feedback to patients who participated in the same format of group CBT who received feedback as well as
their therapists on their wellbeing. This study was conducted in a private psychiatric hospital in Australia, thus the patients were either inpatients or day-patients. Contrary to Davies and colleagues’ study, the authors found that feedback to both the therapist and patient on wellbeing was successful in reducing depressive symptoms, but not wellbeing, in patients who were not on track for improvement (Newnham et al., 2010b).

In the same hospital setting, Dyer, Hooke, and Page (2016) found that providing feedback on both wellbeing and symptoms improved symptoms for patients participating in group CBT who were not on track for improvement. Further, Byrne, Hooke, Newnham, and Page (2012) found that feedback was associated with fewer readmissions for those patients who were on track for improvement. These findings were broadly consistent with previous studies (Hawkins et al., 2004; Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002).

In summary, providing feedback to both the therapist and patient appears to be superior in improving outcomes compared to the provision of feedback to the therapist alone, especially for patients who are not on track for improvement (de Jong et al., 2014; Dyer et al., 2016; Newnham et al., 2010b). However, some studies have found benefits to the outcomes of all patients, including on track patients (Amble et al., 2015; Byrne et al., 2012; Hawkins et al., 2004). Secondly, feedback appears to be, for the most part, effective in improving the outcomes of patients participating in both individual and group treatments (Hawkins et al., 2004; Newnham et al., 2010b).

Researchers have attempted to summarise the routine outcome monitoring and feedback literature using meta-analyses. For example, Shimokawa et al. (2010) conducted a meta-analysis aimed at synthesising the findings of the studies conducted by Lambert and colleagues, using the OQ-45 as the feedback tool. The authors concluded that overall, feedback interventions were effective at improving outcomes for
patients who were identified as being not on track for improvement to a greater extent than on track patients. It was also found that feedback interventions enhanced outcomes while reducing treatment failures among not on track patients. However, the patient and therapist feedback intervention was not effective in preventing treatment failure. The benefits of patient and therapist feedback however, were noted primarily within on track patients. On track patients who received progress feedback experienced more distress reduction and increased odds of experiencing clinically significant improvement compared to on track patients who received no feedback.

Another more recent study attempted to summarise the findings of the feedback literature by conducting a Cochrane Review meta-analysis on the effectiveness of feedback in improving treatment outcomes of common mental health disorders (Kendrick et al., 2016). Common mental health disorders were limited to depressive disorders (i.e., major depression and dysthymia), mixed anxiety and depression, and specific anxiety disorders (i.e., generalised anxiety disorder, phobias, obsessive-compulsive disorder, panic disorder, and post-traumatic stress disorder). The authors of this review concluded that feedback was not effective in improving treatment outcomes overall. However, some problems have been identified with the ability of this particular meta-analysis to support the stated conclusions, given its limitations. Firstly, Kendrick et al. (2016) framed their research question such that their meta-analysis was evaluating the effectiveness of feedback in improving outcomes for all patients. As on track patients are already expected to have a positive outcome, researchers expected and found that feedback was most effective in improving outcomes for patients who were not on track for improvement, not necessarily all patients (Shimokawa et al., 2010). Secondly, Lambert and colleagues produced a number of articles in which outcomes were compared to the same historical cohort of patients (Hawkins et al., 2004; Lambert,
Whipple, Bishop, et al., 2002; Lambert, Whipple, et al., 2001; Whipple et al., 2003). Therefore, only one of the Lambert and colleagues’ studies could be included in the Cochrane Review (Lambert, Whipple, et al., 2001). The omission of many of the Lambert and colleagues’ studies that found feedback to be effective in improving outcomes for not on track patients would have skewed the results in favour of feedback being ineffective. This highlights the importance of replicating feedback research outside of the Lambert and colleagues research group.

There are two main theories that seek to explain how feedback works. They are: the feedback intervention theory and the therapeutic assessment theory. The feedback intervention theory was derived from the industrial/organisational field of psychology research (Kluger & DeNisi, 1996) and suggests that feedback is effective because therapists upon receiving feedback, initiate some form of in-therapy action. Evidence for this theory comes from the study conducted by Hatfield et al. (2010). The authors found that when therapists became aware of the patient worsening or making no change, they were most likely to initiate in-therapy action, thus suggesting feedback led to therapist action. On the other hand, the therapeutic assessment theory suggests that the process of providing feedback to the patient is therapeutic for the patient in some way (Finn & Tonsager, 1992). Evidence for this theory comes from the studies that found that feedback to both the therapist and patient was superior to feedback to the therapist alone, as this suggests that feedback to the patient improved outcomes over and above feedback to the therapist only. It is unclear whether one of these theories is more feasible than the other or whether both theories explain why feedback works. The present research will therefore explore these theories in relation to therapists’ judgements of progress and outcomes with the provision of feedback.
Overall, the findings from the feedback literature suggest that feedback is consistently effective at improving outcomes in some way for patients who are not on track for improvement, but there is little consistent evidence that the benefits are present in patients who are on track (or reliably when these two groups of patients are combined). Further, there is limited research which shows that providing feedback to the therapist and patient is superior to the provision of feedback to the therapist alone. In the present thesis, feedback to both the therapist and patient was provided during treatment for two main reasons. Firstly, because some research has shown that feedback to the therapist and patient benefits both not on track and on track patients. Secondly, because collaboration between the therapist and patient has been found to positively affect treatment outcomes and the therapeutic alliance (A. T. Beck, 1979; Dattilio & Hanna, 2012; Tee & Kazantzis, 2011).

**Thesis Aims and Overview**

Given the positive effects of feedback and the potential for feedback to improve therapists’ ability to estimate current progress and predict final outcomes, the dearth of research in this area is surprising. The limited number of studies that have assessed therapists’ judgements of progress and outcomes did not allow therapists access to any routine outcome monitoring feedback (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010; Walfish et al., 2012). Interestingly, all four studies referenced the positive effects of feedback on outcomes, yet failed to assess whether feedback would aid therapists’ judgements of progress and outcomes. Therefore, the present research aimed to examine therapists’ estimates of progress and predictions of outcomes when feedback was available to both the therapist alone, and to the therapist and patient. The present research also compared a simpler empirical method of prediction, than the
methods used in previous research, to therapists’ predictions of outcomes and a more complex longitudinal empirical method.

Another novel aspect of the present research was that symptoms and wellbeing were analysed separately in each of the studies. There were three main reasons for analysing symptoms and wellbeing separately. Firstly, previous research in this area has primarily used a single measure that combines both symptoms and wellbeing (e.g., the OQ-45) (Lambert et al., 1996). This could be problematic as therapists may base their judgements on patient change more so on symptoms than wellbeing, or vice versa. Secondly, research suggests that different outcome measures, for the same patient, do not always tell the same story about that patient (Yasky, King, & O’Brien, 2015). For example, Newnham et al. (2010b) found that providing feedback on wellbeing alone did not significantly improve wellbeing. However, feedback on wellbeing improved patients’ symptoms, suggesting that although negatively correlated, symptoms and wellbeing are separate constructs. Thirdly, wellbeing has also been shown to improve less rapidly than symptoms, as a result of treatment. Therefore, it was thought necessary to analyse these related but different constructs separately. The instruments used to measure symptoms and wellbeing were the five-item Daily Index (DI-5) (Dyer, Hooke, & Page, 2014) and the World Health Organization’s wellbeing index (WHO-5) (Bech, Gudex, & Staehr Johansen, 1996). The DI-5 and WHO-5 were chosen as they are brief five-item measures that are appropriate for daily administration.

Daily administration was necessary for the present research as all of the patients were participants in a manualised CBT group program which runs for 10 consecutive working days (Manning, Hooke, Tannenbaum, Blythe, & Clarke, 1994; Page & Hooke, 2011). The present research was conducted in a day-patient group therapy setting, as the majority of research in this area was performed in university counselling services.
providing individual therapy (Hannan et al., 2005; Hatfield et al., 2010). The CBT group treatment program that patients participated in runs from 9am to 5pm over 10 consecutive working days with 6 to 8 patients in each group. Two therapists run the group sessions and the content aims to assist therapists in managing depression, anxiety, and stress; and involves cognitive restructuring, behavioural experiments, interpersonal skills, and relapse prevention.

The present thesis is comprised of three papers in preparation for journal submission as well as two chapters which serve to link each of the chapters together to form a conceptual whole.

- Chapter 2 asks the question “can therapists predict patient outcomes and estimate current progress when feedback is available to the therapist?” In this study, therapists had access to their patients’ DI-5 and WHO-5 scores on each day of the CBT group. Therapists were asked to predict outcomes and estimate current progress on day four of the CBT group, which was before therapists provided feedback to the patients on day five. As such, this study assessed the effect of feedback to the therapist alone on therapists’ estimates of current progress and predictions of outcomes.

- Chapter 3 serves as a linking segue between Chapters 2 and 4. This chapter touches on some issues with the clinical significance classifications, especially for patients who begin treatment within the functional range. This critical review led to some changes in the way clinical significance was categorised in Chapter 4. Further, in Chapter 2, it was unclear whether therapists were viewing the feedback graphs before making their predictions of patient outcomes and estimates of current progress.
Therefore, in Chapter 4 therapists’ were required to make their estimates of current progress and predictions of outcomes after they had provided feedback to the patient. This ensured that therapists had viewed patient graphs and meant that feedback was provided to both the therapist and patient before therapists made their judgements of progress and outcomes. Therefore, Chapter 4 asked the question; *can therapists predict patient outcomes and estimate current progress with the provision of feedback to both the therapist and patient?*

- Chapter 5 serves as a linking section between Chapters 4 and 6.

- Chapter 6 compares the clinical significance early change method used in the previous chapters and longitudinal (growth mixture modelling) early change method, at predicting negative outcomes. Chapter 6 therefore asks the question; *can early change, as defined by the growth mixture modelling method, predict no change and deterioration more effectively than the clinical significance early change method?*

The final chapter of this thesis (Chapter 7) consists of a general discussion of the effects of feedback on therapists’ estimates of current progress and predictions of outcomes. The chapter also provides a general discussion on therapists’ predictions of outcomes versus empirical methods’ predictions. Finally, recommendations for empirical methods of predicting deterioration are discussed, as well as future research directions.

The findings from the present research will clarify whether feedback can aid therapists’ estimates of current progress and predictions of final outcomes, especially for patients at risk for treatment failure (no change and deterioration). If feedback improves therapists’ judgements, then therapists will be able to initiate changes to their
treatment in order to improve patients’ outcomes. If therapists are still unable to identify negative progress and predict negative outcomes, then the final study will shed some light on which method may be more effective at predicting deterioration. With this information, therapists may be able to reduce the rates of negative outcomes in the future.
Chapter 2

Can Therapists Predict Patient Outcomes and Estimate Progress in Group Cognitive Behaviour Therapy when Feedback is Available to the Therapist?
CHAPTER 2

Abstract

Objectives: Research suggests that therapists are poor at predicting patient outcomes and feedback of routine outcome monitoring improves outcomes. Therefore, the aim of the present study was to bring these two lines of research together to determine whether providing patient progress feedback to therapists would improve therapists’ predictions of patient outcomes. Therapists’ predictions were also compared to an empirical method of prediction.

Method: Participants were 8 therapists and 200 voluntary day patients who participated in a ten-day CBT group. Therapists were required to predict patient outcomes and estimate patient progress firstly at post-assessment and secondly before feedback was given to the patients. Predictions and estimates were made using the clinical significance categories and outcomes were defined by both a symptom and wellbeing measure.

Results: Therapists overpredicted positive and underpredicted negative outcomes. Feedback did however lead therapists to correctly predict the proportion of recovered and deteriorated patients. Further, therapists correctly identified half of the patients who had deteriorated by the fourth session. However, the empirical method outperformed therapists at predicting outcomes.

Conclusions: Feedback to the therapist only partially improved therapists’ predictions of outcomes and estimates of progress. The results highlight the importance of encouraging therapists to use progress feedback to aid their treatment decisions.
Can Therapists Predict Patient Outcomes and Estimate Progress in Group Cognitive Behaviour Therapy when Feedback is Available to the Therapist?

A wealth of psychotherapy research suggests that on average, patients who receive Cognitive Behaviour Therapy (CBT) leave treatment with a positive outcome (Hofmann et al., 2012; Lambert, 2013a). However, somewhat neglected in the literature is the small proportion of patients who deteriorate, and the relatively large proportion who do not change after treatment (Lambert, 2013a). Research conducted in routine clinical settings suggests that rates of no reliable change can be as high as 60% and deterioration rates typically range from 5% to 10% of patients (Lambert, 2013b; Newnham & Page, 2007; Saxon et al., 2017). Despite these statistics suggesting non-optimal outcomes in nearly two-thirds of those treated, therapists both historically (Bergin & Lambert, 1971) and more recently (Walfish et al., 2012) predict that the majority of their patients (≈85%) will experience a positive outcome.

Hannan et al. (2005) for example, asked therapists to predict their patients’ outcomes in terms of their clinical significance category post-treatment. The four clinical significance categories are recovered, improved, unchanged, and deteriorated (Jacobson & Truax, 1991). The authors found that overall, therapists overpredicted positive outcomes and underpredicted no change and deterioration. More specifically, therapists predicted that approximately 90% of patients would achieve a positive outcome, when approximately 40% of patients actually achieved a positive outcome. Therapists also predicted that approximately 9% of patients would experience no change from therapy and 0.01% would deteriorate. In reality, these percentages were higher at approximately 50% for no change and 7.3% for deterioration.
Chapman et al. (2012) assessed therapists’ ability to predict patient outcomes in group therapy. Therapists were required to predict patients’ final outcomes from three categories (reliably improved, no significant change, or reliably worse) after the third group session. The authors found that, like in the Hannan et al. (2005) study, therapists overpredicted positive and underpredicted negative outcomes. Interestingly research has shown that predictions are poor even when they are made late in therapy (Breslin et al., 1997) and that therapist experience makes little if any difference to the accuracy of predictions (Grove & Meehl, 1996). These results suggest that therapists are overly optimistic in their predictions of outcomes such that they heavily underpredict negative outcomes, regardless of experience.

Since therapists were found to be poor at predicting future outcomes, researchers have also assessed therapists’ ability to estimate current patient progress, i.e., where is my patient at now? For example, Hannan et al. (2005) asked therapists: “Today this client is (choose one rating): recovered and ready for termination; improving as expected, but in need of continued treatment; making no progress or poor progress; or getting worse” (Hannan et al., 2005, p. 160). It was found that therapists still overestimated positive and underestimated negative progress. Hatfield et al. (2010) asked the question; “do we know when our patients get worse?” by reviewing therapists’ progress notes. The authors found that therapists identified only 21% of the deteriorated patients in their progress notes. In summary, the findings from the above studies suggest that therapists’ judgements of patient outcomes and progress are poor, such that they overestimate positive and underestimate negative outcomes and progress.

Researchers have been comparing clinical (therapists’) predictions to statistical (empirical) predictions for many years both in the field of psychotherapy and other areas of research (Grove, 2005; Meehl, 1954). It has consistently been found that
Chapter 2: Therapists’ Judgements with Therapist Feedback

Empirical methods of prediction either outperform or are as good as clinical predictions (Grove, 2005; Grove et al., 2000; Spengler, 2013). For example, Hannan et al. (2005) assessed whether an empirical method would be able to predict outcomes more accurately than therapists. As such, the authors compared therapists’ predictions to a lab-tested algorithm, i.e., an empirical method of prediction, which was developed using normative Outcome Questionnaire-45 (OQ-45) data (Lambert et al., 1996). It was found that the empirical method was superior to therapists at predicting patient outcomes. More specifically, the empirical method accurately predicted 36 (100%) of patients who went on to deteriorate, while therapists accurately predicted only 1 (2.8%) of the patients who deteriorated. However, it was found that the empirical method overpredicted deterioration such that 83 (≈70%) of the patients predicted to deteriorate did not in fact deteriorate. These patients were labelled false alarms; and upon further inspection of these false alarms, 74% made no reliable change. Therefore, the authors argued that the majority of patients predicted to deteriorate had a negative outcome (either unchanged or deteriorated), and would therefore be considered at risk for treatment failure. Of note however, is that Hannan et al. (2005) only reported the agreement for deterioration. Therefore, the agreement for the empirical method’s and therapists’ predictions with actual outcomes, for the no change and positive outcome categories, was unclear. Further, a limitation of this finding was that the measure used for the empirical method (OQ-45) was also used as the outcome comparator, which could have led to a bias favouring the empirical method over the therapists. In the present study, agreement was calculated for both positive and negative outcomes. An empirical method that can predict negative outcomes is a useful tool that could be used by therapists during psychotherapy. However, this poses the question; what should therapists do if a patient they are treating is predicted to have a negative outcome?
Research suggests that providing progress feedback of routine outcome monitoring data to therapists and patients improves outcomes in both outpatient and inpatient settings, particularly among those who are at risk for deterioration (Dyer et al., 2016; Hawkins et al., 2004; Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002; Newnham et al., 2010b; Shimokawa et al., 2010). For example, researchers have assessed whether providing progress feedback to therapists would improve outcomes for patients in an outpatient setting (Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002). Approximately half of the patients were randomly assigned to either a feedback or no feedback condition. Therapists of patients in the feedback condition received a progress graph with colour coded feedback regarding their patient’s score on a measure of symptoms and wellbeing, the OQ-45 (Lambert et al., 1996) after each session. The authors found that patients who were in the feedback condition, and were not on track for improvement, achieved significantly better outcomes at termination compared to not on track patients in the no feedback condition (Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002).

Other studies have found that feedback improved outcomes for both patients who were not on track for improvement and patients who were on track. For example, Amble et al. (2015) found that all patients from six psychiatric clinics in Norway who received feedback achieved significantly superior outcomes compared to patients in the no feedback condition. Further, some studies that did not necessarily distinguish between not on track and on track patients found an overall positive effect of feedback on outcomes in diverse clinical settings (e.g., community mental health, university outpatient clinics, and inpatient settings) (Bickman et al., 2011; Connolly Gibbons et al., 2015; Reese et al., 2009; Simon et al., 2013).
Hawkins et al. (2004) replicated the Lambert, Whipple, et al. (2001) study but added a third condition in which both the therapist and patient received progress feedback regarding patient OQ-45 scores. It was found that the therapist and patient feedback condition improved significantly more than the therapist feedback and control conditions. This suggests that providing feedback to both the therapist and patient improves outcomes over and above providing feedback to the therapist alone. Other studies have found similar results in individual and group therapy settings and in both short and long term therapy (de Jong et al., 2014; Newnham et al., 2010b).

Also of interest in the feedback literature is the question of how feedback actually affects change in patient outcomes. There are two main theories as to how feedback works. These are: the feedback intervention theory and the therapeutic assessment theory. The feedback intervention theory was derived from the industrial/organisational field of psychology research (Kluger & DeNisi, 1996) and posits that feedback works by encouraging therapists to initiate changes in treatment (Carlier et al., 2012). For example, upon receiving feedback of non-response, the therapist may refer the patient to another professional, change their treatment strategy, and/or provide further support and education to the patient on how to manage their presenting problem/s (Greenhalgh, Long, & Flynn, 2005). On the contrary, the therapeutic assessment theory suggests that providing the feedback is therapeutic for the patient in some way (Finn & Tonsager, 1992). More specifically, the therapeutic assessment theory posits that feedback promotes better communication between the patient and therapist and allows the patient to feel involved in treatment and work collaboratively with their therapist (Kendrick et al., 2016). Research also suggests that the majority of patients in feedback studies agree that it is important to monitor
Chapter 2: Therapists’ Judgements with Therapist Feedback

outcomes (Lutz, Rubel, et al., 2015; Strauss et al., 2015), suggesting that patients like receiving feedback.

Despite the developments in the feedback literature, there is still a small cohort of patients who do not achieve clinically significant improvement and others who even worsen over the course of therapy. If therapists could predict their patients’ outcomes, then they may be able to identify those patients veering off track for improvement, and guide them in the direction of positive outcomes, like the feedback intervention theory suggests. Unfortunately and as discussed, the results from the studies assessing therapists’ judgements of outcomes and progress suggest that therapists are poor at predicting those patients who will not achieve clinically significant improvement from therapy (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). However, also common to each of these studies is the fact that none of them included feedback, which has been shown to improve the outcomes of patients who are not on track for improvement (Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002; Newnham et al., 2010b). Hannan et al. (2005) recommended that in future research, therapists should be provided with routine outcome monitoring feedback to determine whether this improves their predictions of outcomes. Therefore, in the present study, therapists had access to patient progress graphs before making their predictions at the end of the fourth day of therapy, and before they gave this feedback to the patients. Previous research also grouped the recovered and improved clinical significance categories into one “positive outcome” category. Therefore, the present study analysed the recovered and improved groups separately to determine whether therapists’ judgements would differ for the two categories.

The aim of the present study was to determine whether results from previous research could be replicated in that therapists would overpredict positive and
underpredict negative outcomes. However, there were three main differences in the present study.

1. Feedback was added to the study to determine whether this would improve therapists’ predictions of outcomes and estimates of current progress.

2. The recovered and improved “positive outcome” clinical significance categories were analysed separately.

3. Therapists’ predictions and the empirical method of prediction were compared regarding their agreement with actual outcomes for both positive and negative outcomes.

Based on previous research, it was hypothesised that therapists would overestimate positive and underestimate negative progress and outcomes; and that the empirical method would be superior at predicting patient outcomes compared to therapists.

Method

Participants

Participants were 8 clinicians and 200 voluntary day-patients who were recruited from a private psychiatric hospital and took part in a ten-day CBT group program (Page & Hooke, 2011). The number of total cases (N) was 297 due to some occasions where two therapists predicted the outcomes of the same patient. Questionnaire data was obtained from patients attending the clinic from April 2015 to February 2016. Patients’ ages ranged from 15 to 77 years old with a mean of 39.4, and 57.5% were female. Patients were diagnosed by their treating psychiatrist using ICD-10 criteria (World Health Organization, 1992), and the main primary diagnoses were affective (57.5%), neurotic (27.5%) and substance abuse (7.5%) disorders. The University of Western Australia Human Research Ethics Committee approved the study protocol prior to
commencement, and patients provided informed consent as part of the routine admission procedure to the hospital.

Measures

The World Health Organization’s wellbeing index (WHO-5) (Bech et al., 1996) is a self-report five-item scale of positive wellbeing. The version used in this study (Newnham, Hooke, & Page, 2010a) was administered daily and items are rated on a 6-point Likert scale measuring frequency from 0 (“at no time”) to 5 (“all of the time”). A sample item from the WHO-5 is “Over the past 24 hours I have felt calm and relaxed”. Total scores range from 0 to 25 where a higher score indicates more positive wellbeing. The WHO-5 demonstrates good internal consistency in a psychiatric sample (Cronbach’s α = 0.89) and good convergent validity as shown by its strong correlation with other measures of psychological health (Newnham et al., 2010a). The WHO-5 has also been found to demonstrate sensitivity to change in a psychiatric setting (Newnham et al., 2010a) as well as in the present sample.

The five-item Daily Index (DI-5) (Dyer et al., 2014) is a self-report measure of five aspects of affective psychological distress: depression, anxiety, worthlessness, coping behaviours, and suicidal ideation. Items are rated on a 6-point Likert scale measuring frequency of the above symptoms over the past 24 hours from 0 (“at no time”) to 5 (“all of the time”). Scores range from 0 to 25 with higher scores indicating higher affective psychological distress. The DI-5 has demonstrated good internal consistency (Cronbach’s α = 0.86) and moderate test-retest reliability (r = 0.64) (Dyer et al., 2014). The DI-5 has also exhibited good convergent validity due to its strong correlation with other mental health measures (Dyer et al., 2014). The DI-5 has also been found to be sensitive to change as shown by the significant difference between
mean intake and discharge scores in a psychiatric setting (Dyer et al., 2014) as well as in the present sample.

Therapists were required to fill out the Therapist Prediction Questionnaire (TPQ), which was an extended adaptation from Hannan et al. (2005), at two time points. The first time point was on day one of the CBT group, which was following assessment but before treatment had begun. This time point will be referred to as post-assessment. The second time point was on day four, which was after therapists had access to feedback in the form of DI-5 and WHO-5 scores and progress graphs and before therapists provided this feedback to their patients. This time point will be referred to as post feedback. The item asking therapists to predict their patients’ outcome stated: “Using your clinical judgement, predict this patient’s outcome at the end of day 10 of the CBT group. This patient will leave treatment (circle one prediction): recovered, improved but not recovered, unchanged, or deteriorated.” The item asking therapists to estimate their patients’ current progress read: “Considering this patient’s pre-treatment functioning, rate this patient’s progress as of today’s session using your clinical judgement. Today this patient is (circle one response): recovered, improving as expected, but not recovered, making minimal or no improvement, or getting worse.”

Procedure

In the present study, patients took part in a manualised Cognitive Behaviour Therapy (CBT) group program (Page & Hooke, 2011). The CBT group program runs from 9am to 5pm over 10 consecutive working days and contains between 6 and 8 patients. The CBT group sessions are run by two therapists and focus on managing depression, anxiety, and stress, and involve cognitive restructuring, behavioural experiments, interpersonal skills, and relapse prevention. Patients were invited to complete both the DI-5 and WHO-5 at the beginning of each day and therapists had
access to this data via a computer throughout the whole course of the CBT group program. Following the assessment on day one (post-assessment), therapists were required to complete the prediction item from the TPQ. At post feedback (day four), therapists again completed the prediction item as well as the estimate item from the TPQ. This was completed after therapists had access to DI-5 and WHO-5 daily data and progress graphs. On day five, as part of routine clinical practice, therapists were required to provide their patients with feedback on their DI-5 and WHO-5 scores in the form of a patient progress graph and group discussion. The graphs present the raw scores for each measure over sessions in addition to an expected treatment response curve which was generated from archival data (see Figure 1 for example WHO-5 feedback graph). Thus, the graphs do not depict clinical significance categories. The example WHO-5 feedback graph presented in Figure 1 depicts a patient whose wellbeing initially improved over the first three sessions but got worse on the fourth session, such that their progress fell out of the expected treatment response curve.

Figure 1. Screen capture of a patient’s WHO-5 wellbeing feedback graph with the expected treatment response curve.
The feedback component of the session was semi-structured and involved the entire group. Therapists prompted their patients to discuss their progress by encouraging them to identify patterns of change on their graphs and share insights into these change patterns. The therapists also encouraged patients to reflect on what factors may have been responsible for improvement, and what factors may have led to lapses with a view to adapting treatment accordingly. The feedback component of the CBT group was also implemented on the final day of treatment. Although therapists were routinely required to provide progress feedback to their patients on day five, therapists were able to access their patients’ progress graphs throughout the whole course of treatment.

**Data Analysis**

**Clinical Significance.** In the present study, the Jacobson and Truax method of clinically significant change was used to determine actual patient progress and outcomes. This clinical significance method classifies patients based on two conditions. The first condition is whether the change was statistically meaningful and larger than the Reliable Change Index$^2$ (RCI) of the measure. The second condition is whether the patient finished treatment within a pre-determined functional population range on the measure. For the DI-5, the cut-off score for the functional population was 6.17. Therefore, scores below 6.17 were considered to be in the functional population range. For the WHO-5, the cut-off score for the functional population was 10.8. Therefore, scores above 10.8 were considered to be in the functional population range. The RCI expresses each patient’s change in pre- and post-scores on the measure in standard error units of measurement and signals that a reliable change has occurred when this value exceeds an increase or decrease of 1.96.

As discussed, these conditions produce a minimum of four outcome categories: recovered (reliable improvement and functional population range), improved (reliable
improvement but not in functional population range), unchanged (no reliable change), and deteriorated (reliable worsening). The clinical significance category for each individual patient was calculated at day four (for progress comparison) and day ten (for outcome comparison). Therapists’ predictions were then compared to patients’ clinical significance scores (actual progress and outcomes) at day four (for estimates of progress) and day ten (for predictions of outcomes).

**Comparisons.** The comparisons between therapists’ predictions of progress and actual patient progress, and therapists’ predictions of outcomes and actual patient outcomes were done using two methods. Firstly, the proportions of patients falling into each clinical significance category for therapists’ estimates and predictions were compared to the proportion in each category for actual patient progress and outcomes using two-proportion z-tests to test for significant differences. Secondly, therapists’ ability to predict outcomes was assessed by calculating the agreement between the actual outcomes and therapist predicted outcomes. This done by running cross-tabs in SPSS.

**Missing Values Analysis.** A missing values analysis revealed that missing cases were missing completely at random for both the DI-5 and WHO-5; Little’s MCAR test \( \chi^2 (494) = 521.98, p = .185 \). Therefore, expectation maximisation was conducted on the data for the purposes of further analysis.

**Results**

Overall descriptive statistics for the DI-5 and WHO-5 scores from day 1 to day 10 of the CBT groups are displayed in Table 1. In general, patients’ DI-5 scores decreased over the course of the CBT group, suggesting a reduction in symptoms consistent with recovery according to the Jacobson and Truax (1991) method of clinical significance categorisation. Patients’ WHO-5 scores increased over the course of the
Chapter 2: Therapists’ Judgements with Therapist Feedback

CBT group, suggesting an increase in positive wellbeing also consistent with recovery according to clinical significance categorisation (see Table 1).

Table 1

Means and standard deviations by day of treatment for the DI-5 and WHO-5.

<table>
<thead>
<tr>
<th></th>
<th>DI-5</th>
<th></th>
<th>WHO-5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Day 1</td>
<td>9.92</td>
<td>5.93</td>
<td>7.07</td>
<td>5.19</td>
</tr>
<tr>
<td>Day 2</td>
<td>7.78</td>
<td>5.44</td>
<td>7.69</td>
<td>5.27</td>
</tr>
<tr>
<td>Day 3</td>
<td>7.56</td>
<td>5.65</td>
<td>8.47</td>
<td>5.56</td>
</tr>
<tr>
<td>Day 4</td>
<td>7.16</td>
<td>5.69</td>
<td>8.68</td>
<td>5.72</td>
</tr>
<tr>
<td>Day 5</td>
<td>6.67</td>
<td>5.22</td>
<td>8.97</td>
<td>5.74</td>
</tr>
<tr>
<td>Day 6</td>
<td>6.83</td>
<td>5.15</td>
<td>10.06</td>
<td>5.90</td>
</tr>
<tr>
<td>Day 7</td>
<td>6.57</td>
<td>5.22</td>
<td>9.75</td>
<td>5.88</td>
</tr>
<tr>
<td>Day 8</td>
<td>5.91</td>
<td>4.92</td>
<td>10.56</td>
<td>6.06</td>
</tr>
<tr>
<td>Day 9</td>
<td>5.78</td>
<td>4.76</td>
<td>10.59</td>
<td>6.15</td>
</tr>
<tr>
<td>Day 10</td>
<td>5.51</td>
<td>4.75</td>
<td>11.00</td>
<td>6.32</td>
</tr>
</tbody>
</table>

Therapists’ Predictions of Outcomes

Post-assessment predictions. To illustrate the general pattern of results, the therapists’ predicted outcomes at post-assessment compared to actual patient outcomes, according to the DI-5, are presented in Figure 2. A “positive outcome” was defined as either recovered or improved, while the unchanged and deteriorated categories reflected the appropriate clinical significance groups.
Figure 2. Therapists’ predicted outcomes at post-assessment compared to actual patient outcomes for the DI-5.

Overall, therapists’ predictions at post-assessment were significantly different to actual patient outcomes on the DI-5, $\chi^2(6) = 15.28, p < .05$. Differences between therapists’ predictions at post-assessment and DI-5 actual outcomes for the positive, unchanged, and deteriorated categories were tested using the two-proportion $z$-test with Bonferroni adjustment (positive outcome: $Z = 10.05, p < .001$; unchanged: $Z = -9.32$, $p < .001$; deteriorated: $Z = -3.09, p < .01$). All of these tests were significant which suggests that therapists’ predictions at post-assessment were significantly different to actual patient outcomes for each category. More specifically, therapists overpredicted the proportion of positive outcomes and underpredicted the proportion of negative outcomes for the DI-5 (see Figure 2). Table 1 illustrates the pattern of results for both symptoms (DI-5) and wellbeing (WHO-5). The results suggest that the outcomes for the WHO-5 were very similar to that of the DI-5, such that therapists overpredicted positive and underpredicted negative outcomes at post-assessment for the WHO-5.
Previous research examined positive outcome as a single variable, whereas in the present study, positive outcome was separated into the recovered and improved clinical significance categories. For the recovered outcome, actual patient outcomes and therapists’ predictions at post-assessment were not significantly different at the \( p < .01 \) level for either the DI-5 (\( Z = -2.34, p = .02 \)) or WHO-5 (\( Z = -1.66, p = .10 \)). This suggests that therapists correctly predicted the proportion of recovered patients. However, therapists overpredicted the proportion of patients falling into the less successful improved outcome (see Table 2).

Table 2.

*Therapist predicted outcomes at post-assessment and post feedback compared to actual outcomes for the DI-5 and WHO-5.*

<table>
<thead>
<tr>
<th>CS Category</th>
<th>Therapist Predictions (%)</th>
<th>Actual Outcomes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Post-assessment</td>
<td>Post Feedback</td>
</tr>
<tr>
<td>Recovered</td>
<td>25.1</td>
<td>29.1</td>
</tr>
<tr>
<td>Improved</td>
<td>59.9</td>
<td>57.8</td>
</tr>
<tr>
<td>Unchanged</td>
<td>15.1</td>
<td>12</td>
</tr>
<tr>
<td>Deteriorated</td>
<td>0.0</td>
<td>1.1</td>
</tr>
</tbody>
</table>

**Post feedback predictions.** The results comparing actual outcomes with therapists’ post feedback predictions followed a similar pattern to therapists’ post-assessment predictions (see Table 1). The two-proportion \( z \)-tests were significant for the improved (DI-5: \( Z = 11.92, p < .01 \); WHO-5: \( Z = 13.82, p < .01 \)) and unchanged (DI-5: \( Z = -10.15, p < .01 \); WHO-5: \( Z = -12.01, p < .01 \)) categories, suggesting that therapists’ post feedback predictions were significantly different to actual outcomes for the improved and unchanged categories. Specifically, therapists overpredicted the improved
outcome and underpredicted the unchanged outcome (see Table 1). Whereas, the two-proportion z-test for the recovered outcome was not significant (DI-5: Z = -1.26, p = .21; WHO-5: Z = -0.58, p = .56), which means that therapists correctly predicted the proportion of patients who were classified as recovered by the end of treatment. For the deteriorated outcome, there was also no significant difference between actual outcomes and post feedback therapist predictions for either the DI-5 or WHO-5 (Z = -1.82, p = .07 for both). This means that post feedback, therapists correctly predicted the proportion of patients who would deteriorate. However, the 3 patients predicted to deteriorate did not actually deteriorate.

Therapists’ ability to predict patient outcomes was also assessed by calculating the agreement between the actual outcomes and therapist predicted outcomes. The agreement between the actual outcomes and therapist predicted outcomes was 27.9% for predictions made at post-assessment and 24% for post feedback therapist predictions. The agreement between actual outcomes for the WHO-5 and therapist predicted outcomes was 24% for post-assessment and post feedback predictions. This suggests that therapists were no better than chance (25%) at predicting patient outcomes in terms of agreement between therapists’ predictions and actual outcomes.

**Empirical Predictions of Outcomes**

The empirical method used in the present study was calculated by comparing the agreement between the post feedback DI-5 and WHO-5 scores and actual patient outcomes with the agreement between post feedback therapists’ predictions and actual outcomes. For these analyses, outcome was defined as positive (recovered and improved) or negative (unchanged or deteriorated). Agreement between post feedback DI-5 scores and DI-5 scores at the end of treatment was 74%; and agreement between post feedback WHO-5 scores and WHO-5 scores at the end of treatment was 70%. The
agreement between therapists' post feedback predictions and DI-5 scores at the end of treatment was 45.8% and agreement between therapists' post feedback predictions and WHO-5 scores at the end of treatment was 43.3%. This suggests that the empirical method was performing at a superior level to therapists at predicting actual patient outcomes.

**Therapists’ Estimates of Progress**

The DI-5 results for therapists’ estimates of post feedback progress (day four) compared to post feedback actual progress are presented in *Figure 3*, to illustrate the general pattern of results. Overall, therapists’ estimates of progress were significantly different to actual progress at day four for the DI-5, $\chi^2(9) = 42.27, p < .001$.

![Figure 3. Therapists’ estimates of progress post feedback compared to post feedback actual progress for the DI-5.](image)

The results showed that, similar to therapists’ predictions of outcomes, therapists tended to overestimate positive progress and underestimate negative progress (see *Figure 3*). This suggests that despite having spent more time with the patients and having access to patient motoring graphs, therapists were not able to correctly estimate
patient progress on day four (post feedback) of the CBT group. Table 3 illustrates the pattern of results for both symptoms (DI-5) and wellbeing (WHO-5), and again shows that actual progress on the WHO-5 was very similar to that of the DI-5.

Table 3.

*Therapists’ estimates of post feedback progress compared to actual patient progress at day four on the DI-5 and WHO-5.*

<table>
<thead>
<tr>
<th>Post Feedback Therapist Estimates (%)</th>
<th>Actual Progress (%)</th>
<th>DI-5</th>
<th>WHO-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovered</td>
<td>4</td>
<td>22.6</td>
<td>11.1</td>
</tr>
<tr>
<td>Improved</td>
<td>66.1</td>
<td>7.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Unchanged</td>
<td>27</td>
<td>64.3</td>
<td>81.8</td>
</tr>
<tr>
<td>Deteriorated</td>
<td>2.9</td>
<td>5.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

The deteriorated category was the only category for which there was no significant difference between therapists’ estimates of progress and actual progress post feedback (day four) for both the DI-5 (Z = -1.47, p = .14) and WHO-5 (Z = -0.31, p = .76). Interestingly, therapists correctly estimated half of the patients who were classified as deteriorated on either measure at the post feedback time point (day four). Of the patients therapists correctly estimated as deteriorated at the post feedback time point, 75% completed treatment with a positive outcome on at least one of the measures.

**Discussion**

The aim of the present study was to examine therapists’ ability to predict patient outcomes and estimate current progress, before and after receiving feedback regarding patient symptom and wellbeing scores. In light of previous research (Hannan et al., 2005), it was hypothesised that therapists would overestimate positive and
underestimate negative outcomes and progress; and the measures (empirical method) would be superior at predicting patient outcomes compared to therapists. As expected, therapists’ outcome predictions were significantly different to actual outcomes, such that therapists overpredicted positive and underpredicted negative outcomes both at post-assessment and post feedback. The results suggested that this pattern of optimism was also apparent when therapists were asked to estimate their patients’ current progress at the post feedback time point, thus supporting the hypothesis. In the present study, it was also found that the empirical method was superior to therapists at predicting patient outcomes, which supports the second hypothesis. However, when positive outcome was analysed as the separate recovered and improved categories, it was found that therapists correctly predicted the proportion of patients who would leave treatment recovered but overestimated the proportion of patients who would achieve an improved outcome. Finally, feedback had a positive effect on the accuracy of therapists’ predictions of outcomes and estimates of current progress in that therapists correctly predicted and estimated the proportion of deteriorated patients. However, feedback did not aid accuracy of therapists’ predictions of outcomes and estimates of current progress for all of the clinical significance categories.

The overall finding that therapists tend to be overly optimistic when making predictions about patient outcomes or estimates of patient progress was consistent with previous research (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). The findings were also consistent with long-standing evidence that statistical predictions (empirical methods) are frequently superior to clinical judgements (Dawes, Faust, & Meehl, 1989; Meehl, 1954). A possible explanation of the findings comes from a study conducted by Walfish et al. (2012). The authors found that therapists typically display self-assessment bias, which is the tendency to assess one’s own personal performance as
overly positive compared to others. It was found that 25% of therapists believed their skills to be at the 90\textsuperscript{th} percentile compared to other therapists, and none of the therapists believed they were below average (Walfish et al., 2012). Therefore, perhaps therapists were overly optimistic about their patients’ outcomes because they believed that any patient not responding to treatment would have been another therapists’ patient, and not a patient they were treating. Further, research suggests that therapists are just as likely to fall victim to cognitive biases when making judgements about people and their behaviour, as non-professionals are (Wilson, 1996). Another possibility, is that therapists are optimistic for the sake of the patient because therapist optimism is associated positive outcomes (M. Beck, Friedlander, & Escudero, 2006). Further, imagine if you were a patient and your therapist was predicting that you were going to fail. If the therapist believes you will not improve, then you might ask; “what is the point of putting in the effort that is required of therapy”? This could become a self-fulfilling prophecy even for the therapist, as they may “give up on” the patient who is not on track for improvement. Future research could explore this further using qualitative methods.

Despite the research which shows that empirical methods are superior to clinical judgements of outcomes, historically therapists typically monitor their patients’ progress informally rather than using formal monitoring and empirical outcome prediction methods (i.e., feedback) (Lutz et al., 2013). With regards to the present study, although available, it was unclear whether therapists were actually using the feedback (i.e., viewing the DI-5 and WHO-5 progress graphs) before they were required to feedback these results to the patients on day five. Therefore, future research should assess therapists’ predictions of outcomes after feedback has been given to both the therapist and patient.
Also unclear in the present study was whether therapists altered their treatment in any way after having received feedback. Previous research has shown that on approximately 70% of occasions that therapists received feedback of non-response to treatment, therapists adjusted their treatment in some way (Lutz, Rubel, et al., 2015). For example, they discussed the results with patients, adjusted their treatment strategies, or worked on enhancing the therapeutic alliance (Lutz, De Jong, & Rubel, 2015). The fact that therapists’ use of feedback (i.e., an implementation check) was not assessed was a limitation of the present study. Another possibility is that therapists viewed the patient progress graphs but trusted their clinical judgement over and above the feedback. Further, patient progress over the course of treatment is not linear (Lutz et al., 2013). Therefore, if the therapist views the graph of a patient who is deteriorating at the post feedback time point (day four), it is possible that this patient will achieve a positive outcome by the end of treatment, thus making predictions difficult even when feedback is available. Future research could apply a more sophisticated longitudinal modelling technique to investigate whether this aids therapists’ judgements. Interestingly and contrary to previous research, feedback did lead therapists to correctly predict the proportion of patients who would leave treatment deteriorated. However, the patients who therapists predicted would deteriorate, did not in fact deteriorate. Therefore, despite feedback possibly encouraging therapists to consider that some patients might deteriorate, therapists were unable to correctly predict who these patients would be.

A limitation of the present study was that a no feedback condition was not included. Feedback has been shown to consistently improve outcomes for patients at the hospital in which this data was taken from (Dyer et al., 2016; Newnham et al., 2010b). As such, feedback to the patient has become a mandatory component of the CBT groups. Therefore, it could be considered unethical to withdraw feedback from a
random sample of patients for the purposes of this study. The use of the feedback measures as the measures of actual outcomes could also be considered a limitation. Other outcome measures were not used in the present study to keep the comparisons simple. However, future research could include other measures of outcome such as the Depression Anxiety Stress scale (DASS-21) (Lovibond & Lovibond, 1995). Another limitation of the present research was the small number of therapists who participated (8 therapists) and that therapist effects were not accounted for. Further, the nesting of cases within therapists may have led to some biases in the results. For example Therapist A in the group may have been influenced by Therapist B, or vice versa, when completing the TPQ. Future research could address this by accounting for therapist effects and the nesting of cases within therapists in the model.

Despite these limitations, as compared to previous studies and post-assessment predictions in the present study, the inclusion of feedback improved therapists’ estimates of current patient progress. More specifically, therapists correctly identified 50% of deteriorated patients in comparison to previous research in which therapists identified only 21% of deteriorated patients (Hatfield et al., 2010). Interestingly, 3 of the 4 patients who therapists estimated to be deteriorated at the post feedback time point, ended treatment with a positive outcome. This result may provide support for the feedback intervention theory (Kluger & DeNisi, 1996) as it suggests that when therapists noticed that the patient was deteriorated at post feedback, they altered some aspect of their treatment of that patient to turn their outcome into a positive one. These results highlight the importance of encouraging therapists to use feedback in the form of progress monitoring to aid their decisions in treatment.

In summary, the results suggest that therapists tended to overpredict positive and underpredict negative outcomes. This same pattern was observed when therapists were
asked to estimate current patient progress, and these results were broadly consistent with previous research (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). This study was the first, to our knowledge, to ask therapists to predict patient outcomes and estimate current progress when feedback was available to the therapist. Feedback appeared to improve therapists’ predictions of outcomes and estimates of progress in some way. For example, at the post feedback time point (after therapists had access to patient DI-5 and WHO-5 graphs), therapists correctly predicted the proportion of recovered and deteriorated patients. Therapists also correctly identified 50% of deteriorated patients at the post feedback time point, which is an improvement from previous studies (Hatfield et al., 2010). Further, the majority of the correctly identified deteriorated patients went on to achieve a positive outcome. This suggests that upon identifying deterioration, therapists may have intervened to help these patients achieve positive outcomes by the end of treatment, and this is consistent with the feedback intervention theory (Kluger & DeNisi, 1996). The results of the present study highlight the importance of encouraging therapists to use patient progress feedback to aid in their treatment decisions. Future research could assess therapists’ predictions of outcomes after feedback has been provided to both the therapist and patient and endeavour to further explore the nature of therapist feedback use.
Chapter 3

The Addition of Feedback to the Patient and Clinical Significance Categorisation Issues
CHAPTER 3

The Addition of Feedback to the Patient and Clinical Significance Categorisation Issues

The results presented in Chapter 2 suggest that therapists overpredicted positive and underpredicted negative outcomes; and the same patterns were observed for therapists’ estimates of current progress. These results were consistent with previous research (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010), but were found in the study presented in Chapter 2 despite the availability of feedback to the therapist. Although feedback did improve therapists’ predictions of outcomes and estimates of progress to some extent, therapists’ ability to predict and identify negative outcomes and progress was, arguably, insufficient. This is problematic because if patients who are not progressing as expected are left unnoticed, then they may “slip through the cracks”, potentially leading to negative outcomes.

One limitation of the study presented in Chapter 2 was that although therapists had access to their patients’ progress graphs before the post feedback time point, there was no way of testing whether therapists were actually viewing their patients’ feedback scores and graphs before making their predictions of outcomes and estimates of progress. Further, even if therapists had viewed their patients’ progress graphs before making their estimates and predictions, changes in their abilities could only be attributed to feedback to the therapist alone. There is some evidence to suggest that feedback to both the therapist and patient is superior to feedback to the therapist alone in improving outcomes for patients at risk for a negative outcome (de Jong et al., 2014; Hawkins et al., 2004). Therefore, it would be fair to speculate that feedback to both the therapist and patient could improve therapists’ predictions of outcomes and estimates of progress over and above feedback to the therapist alone.
As a result of this limitation, it would be premature to conclude that feedback was insufficient at improving therapists’ ability to judge future outcomes and current progress to the extent necessary to prevent negative outcomes. To address this problem, therapists (in Chapter 4) were invited to make their predictions of future outcomes and estimates of current progress at the end of the fifth day of the ten-day CBT group program. The fifth day was chosen because therapists are scheduled to provide progress feedback to their patients on this day, as part of routine clinical practice (Newnham et al., 2010a). Therefore, we can be more confident that therapists will have viewed and interpreted their patients’ feedback graphs as well as provided this as feedback to their patients before making their predictions and estimates.

Another change made to Chapter 4 was the way in which we chose to analyse positive outcomes. In Chapter 2, we examined positive outcome separately, using the Jacobson and Truax (1991) recovered and improved clinical significance categories, to determine whether this would return different results to previous studies (Chapman et al., 2012; Hannan et al., 2005). Analysing positive outcomes separately made some differences to the results, such that therapists correctly predicted the proportion of recovered patients. However, for Chapter 4 we decided to analyse the recovered and improved categories together as one “positive outcome” category for three main reasons. Firstly, because the present thesis focuses more on negative outcomes. Secondly, because the improved category form only a small proportion of the overall patients (≈7%); and thirdly to make our results more easily comparable to previous studies (Chapman et al., 2012; Hannan et al., 2005).

Another limitation of the first study was the way in which the Jacobson and Truax (1991) clinical significance method classified patients who began treatment within the functional population range on the DI-5 and/or WHO-5. Previous research
has frequently removed patients who begin treatment within the functional population range on the measure of interest (Bauer, Lambert, & Nielsen, 2004; McGlinchey, Atkins, & Jacobson, 2002; Seggar, Lambert, & Hansen, 2002). However, when more than one measure is used, it would be possible for patients to fall within the dysfunctional population range on one measure but the functional population range on another. The measure in which the patient falls within the functional population range simply may not capture their presenting problems. Moreover, in a CBT group program with a diagnostically heterogeneous group of patients, it is not surprising that patients will occasionally not have problems that are captured by the routine outcome monitoring instrument. That is, there may be a target of treatment that is not specifically asked about in the scales. For example, the DI-5 and/or WHO-5 may not capture substance abuse issues, obsessions, hoarding, interpersonal issues, etc.

Some researchers have acknowledged a sample of patients who begin treatment within the functional population range. For example, Tingey, Lambert, Burlingame, and Hansen (1996) suggested the inclusion of a fifth clinical significance category for patients who begin treatment within the functional population range and make a reliable positive change. The authors named this category, reliable improvement (Tingey et al., 1996). Nevertheless, what about patients who begin and end treatment within the functional population range but do not make a reliable change? According to the Jacobson-Truax method of clinical significance categorisation, these patients would technically be classified as unchanged (Jacobson & Truax, 1991). However, from a service perspective, someone who leaves treatment within the functional population range would be classified as a positive outcome. For example, imagine you are measuring flu symptoms using a questionnaire at a hospital, and you ask a patient presenting with a broken leg to complete the questionnaire on admission and discharge.
On admission, the patient with the broken leg reported no flu symptoms and on discharge they again reported no flu symptoms. This could be deemed a positive outcome because the patient has not obtained a problem they did not have upon entering hospital.

Another possibility is that the patient began treatment in the middle of the functional population range and as such would not have room (i.e., ceiling/floor effects) to have a reliable (statistically significant) change in a positive direction. According to the Jacobson and Truax (1991) method of clinical significance classification, these patients would be classified as unchanged, which is associated with a negative outcome. Further, patient change is not necessarily linear and as such, some patients may experience ups and downs during therapy but end within the functional population range (Lutz et al., 2013). Again logically, this outcome would be perceived as a positive one because in a hospital setting, one of the primary aims for treatment would be to get the patient well enough to be able to function in everyday life again (Malhi et al., 2015; McGorry, 2017).

To address these issues, in Chapter 4 we have conducted the analyses in two ways. Firstly, rather than remove all of the patients who began treatment within the functional population range on the DI-5 and/or WHO-5, we included these patients and classified those who ended treatment within the functional range as a positive outcome. Secondly, we re-ran all of the analyses without the patients beginning treatment within the functional population range on the DI-5 and/or WHO-5 to determine whether this would make a difference to the results. Chapter 4 uses a different sample to that of Chapter 2 as therapists were required to estimate progress and predict outcomes at a different time point. Therefore, we made some adjustments to the Therapist Prediction Questionnaire to reflect this, and reduce the time it took for therapists to complete the
questionnaire. As such, Chapter 4 asks the question: can therapists predict outcomes and estimate progress when feedback is given to both the therapist and patient, and when the clinical significance categorisation is modified to account for patients beginning treatment within the functional population range?
Chapter 4

Therapists’ Ability to Judge Patient Outcomes and Progress with the Provision of Feedback to both the Therapist and Patient
CHAPTER 4

Abstract

Objective: The aim of the present study was to determine whether feedback to both the therapist and patient would improve therapists’ predictions of patient outcomes and estimates of patient progress. Therapists’ predictions were also compared to an empirical method of prediction.

Method: Participants were 9 therapists and 141 voluntary day patients who participated in a ten-day CBT group program. Therapists were required to predict patient outcomes and estimate patient progress using clinical significance categories. Therapists’ predictions and estimates were made after they had received routine outcome monitoring feedback on their patients’ symptoms and wellbeing, and had fed this back to their patients on day five.

Results: Therapists correctly predicted and estimated the proportion of patients in each clinical significance category at the end of treatment and by day five of treatment, suggesting that feedback led to improved clinical judgements. However, agreement between therapists’ predictions and actual patient outcomes was unsatisfactory for negative outcomes. Finally, the empirical method was not necessarily superior to therapists at predicting outcomes.

Conclusions: The results suggest that feedback to both the therapist and patient led to some improvements in therapists’ predictions of outcomes and estimates of progress. However, therapists still require help to identify which patients are at risk for a negative outcome. Implications of the results and suggestions for future research are discussed.
Therapists’ Ability to Judge Patient Outcomes and Progress with the Provision of Feedback to both the Therapist and Patient

Although Cognitive Behaviour Therapy (CBT) is effective in treating the majority of patients (Hofmann et al., 2012; Lambert, 2013b; Westbrook & Kirk, 2005), the more recent patient-focused research approach has drawn attention to the patients who leave treatment unchanged or worse off. Research conducted in routine clinical settings has found that rates of no reliable change can be as high as 60%, while deterioration rates typically range between 5% and 10% (Lambert, 2013b; Newnham & Page, 2007; Saxon et al., 2017). Despite statistics suggesting that nearly two-thirds of their patients may have a negative outcome, therapists typically predict that their patients will achieve positive outcomes (Chapman et al., 2012; Hannan et al., 2005).

Chapman et al. (2012) for example, assessed therapists’ ability to predict patient outcomes in a group therapy setting. Therapists were required to predict whether their patient would leave treatment reliably improved, having made no significant change, or reliably worse. The authors found that therapists overpredicted positive outcomes and underpredicted no change and deterioration (reliably worse) (Chapman et al., 2012). Research in this area has also found that therapist experience made little or no difference to outcome prediction accuracy (Grove & Meehl, 1996), and that therapist predictions made later in therapy were also poor (Breslin et al., 1997).

Hannan et al. (2005) asked therapists to predict their patients’ outcomes using the clinical significance categories: recovered, improved, unchanged, or deteriorated (Jacobson et al., 1984; Jacobson & Truax, 1991), and their clinical judgement. Consistent with Chapman et al. (2012), the results showed that therapists significantly overpredicted positive outcome (recovered and improved) and underpredicted no change and deterioration. However, unlike in the Chapman et al. (2012) study,
therapists were told that the deterioration rate at the clinic had consistently remained around 8% before making their predictions. Despite having this information, therapists predicted that only 3 patients (0.01%) would deteriorate, and only one of those patients had actually deteriorated at the end of treatment.

As therapists’ predictions of outcomes were found to be poor, Hannan et al. (2005) compared therapists’ predictions to an empirical method based on normative OQ-45 data (Lambert et al., 1996). It was found that the empirical method correctly identified all 36 (100%) patients who left treatment deteriorated and 86% of these patients were identified by the third treatment session. Although the empirical method identified all patients who deteriorated, it also predicted that another 83 patients would deteriorate who did not, thus overpredicting deterioration. Upon further examination of these patients however, 74% made no change by the end of treatment. Therefore, the authors argued that the empirical method was still effective because even the false alarms (i.e., patients predicted to leave treatment with a particular outcome but did not) had poorer outcomes than those patients who were not predicted to deteriorate (Hannan et al., 2005). In summary, therapists’ ability to predict their patients’ outcomes appeared to be quite poor, and an empirical method of prediction, although imperfect, was superior to therapists at predicting deterioration. As a result of these somewhat disappointing clinical predictions, rather than predicting final outcomes, researchers decided to examine whether therapists were able to estimate their patients’ current progress; i.e., where is my patient at after today’s session?

Hannan et al. (2005) asked therapists to estimate their patients’ progress as of today’s session based on the clinical significance categories (recovered, improved, unchanged, deteriorated) (Jacobson & Truax, 1991) and using their clinical judgement. The authors found that therapists still overestimated the proportion of patients who had
made positive progress and significantly underestimated the proportion of patients who had made no change or deteriorated. In a similar vein, Hatfield et al. (2010) assessed therapists’ ability to detect deterioration by reviewing therapists’ progress notes. The authors found that therapists mentioned deterioration in the progress notes of only 21% of deteriorated patients. Therefore, 79% of deteriorating patients were left unnoticed according to the therapists’ progress notes.

Therapists’ inability to identify patient deterioration is worrying because research suggests that when therapists become aware of patient deterioration, *in-therapy action* was their most likely response (Hatfield et al., 2010). In-therapy action can involve discussing the deterioration with the patient, focusing on precipitating events, modifying treatment, among other actions. However, if therapists are unable to identify that their patient has made no change or is deteriorating, then they will not know to initiate any in-therapy action or to change the treatment direction. Therapists’ optimism when it comes to predicting patient outcomes and estimating current progress has been attributed to the self-assessment bias, which is the tendency to overestimate one’s own abilities. For example, Walfish et al. (2012) found that all therapists in their study rated themselves as above average in terms of effectiveness. Therefore, no therapists rated themselves as below average. Routine outcome monitoring and feedback was therefore introduced to psychotherapy due to the inadequacies of clinical judgements of outcomes and progress.

Routine outcome monitoring and feedback involves the repeated administration of a reliable and valid measure and the provision of this data as feedback to the therapist. For example, Lambert and colleagues provided feedback to therapists based on their patients’ symptoms and wellbeing as measured by the OQ-45 (Lambert et al., 1996; Lambert, Whipple, et al., 2001; Lambert, Whipple, Vermeersch, et al., 2002).
Feedback was found to improve outcomes, especially for patients at risk for treatment failure (no change or deterioration). These findings have been replicated in a number of other studies (Bickman et al., 2011; Crits-Christoph et al., 2012; de Jong et al., 2012; Newnham et al., 2010b; Shimokawa et al., 2010; Simon et al., 2013).

Collaboration between the therapist and patient positively affects treatment outcomes and has been recognised as an important contributor to the therapeutic alliance (A. T. Beck, 1979; Dattilio & Hanna, 2012; Tee & Kazantzis, 2011). Therefore, researchers decided to examine whether providing progress feedback of routine outcome monitoring to both the therapist and patient would benefit outcomes. For example, Hawkins et al. (2004) assigned patients to three treatment conditions: one in which no feedback was given, another in which only the therapist was given feedback, and another condition in which the therapist and patient received feedback. The authors found that patients in the therapist and patient feedback condition, achieved significantly superior outcomes compared to patients in the other two conditions (Hawkins et al., 2004). Another study conducted by Newnham et al. (2010b) also found that feedback to both the therapist and patient reduced depressive symptoms in patients at risk for treatment failure.

Despite these promising results, it was unclear how feedback led to improved outcomes. The feedback intervention theory, which was derived from the industrial/organisational field of psychology research (Kluger & DeNisi, 1996), suggests that feedback is effective because therapists initiate some form of in-therapy action. As discussed previously, Hatfield et al. (2010) found support for this theory because when therapists became aware of deterioration or no change, they were most likely to initiate in-therapy action, thus suggesting feedback led to therapist action. The therapeutic assessment theory on the other hand, suggests that the process of providing
feedback to the patient is therapeutic for the patient in some way (Finn & Tonsager, 1992). Lacking in the literature that assesses therapists’ judgements of outcomes and progress is the use of feedback. In the present study, therapists were asked to predict outcomes and estimate progress after they had given feedback to their patients. Therefore, the aim of the present study was to determine the degree to which feedback would improve therapists’ predictions of outcomes and estimates of current progress.

Method

Participants

Nine therapists and 141 voluntary day-patients recruited from a private psychiatric hospital participated in the present study. Patients took part in a ten-day group CBT program run by two therapists (Page & Hooke, 2011). The total number of cases (N) was 209 due to some occasions (68 patients) in which both therapists predicted the outcomes of the same patient. Questionnaire data was obtained from patients attending the clinic from August 2016 to April 2017. Patient age ranged from 15 to 69 years old (M = 40.78), and 51.8% were female. Patients were diagnosed by their treating psychiatrist using ICD-10 criteria (World Health Organization, 1992), and the main primary diagnoses were affective (47.5%), neurotic (35.5%), and substance abuse (6.4%) disorders. The University of Western Australia Human Research Ethics Committee approved the study protocol prior to commencement, and patients provided informed consent as part of the routine admission procedure to the hospital.

Measures

In the present study, two measures were used to monitor progress and assess outcomes: the DI-5 (Dyer et al., 2014) and WHO-5 (Bech et al., 1996), and we decided to analyse these separately for two main reasons. Firstly, because research suggests that the outcomes for the same patient on different measures may not necessarily point to the
same outcomes (Yasky et al., 2015). For example, Newnham et al. (2010b) found that while providing feedback on wellbeing did not significantly improve wellbeing, feedback improved aspects of psychological distress, suggesting that although negatively correlated, they are separate constructs. The second reason was because previous research has analysed symptoms and wellbeing together, rather than separately, by using measures such as the OQ-45 (Lambert et al., 1996).

The Daily Index (DI-5) (Dyer et al., 2014) is a five-item self-report measure. The DI-5 measures five aspects of psychological distress: depression, anxiety, worthlessness, coping behaviours, and suicidal ideation. Frequency of symptoms over the past 24 hours are recorded using a 6-point Likert scale from 0 “at no time” to 5 “all of the time”. Scores range from 0 to 25 with higher scores indicating more severe psychological distress/symptoms. The DI-5 demonstrated moderate test-retest reliability ($r = 0.64$) (Dyer et al., 2014) and good internal consistency (Cronbach’s $\alpha = 0.86$). The DI-5 also exhibited good convergent validity due to its strong correlation with other mental health measures (Dyer et al., 2014). The DI-5 has also been found to be sensitive to change as shown by the significant difference between mean intake and discharge scores in a psychiatric setting (Dyer et al., 2014) and the present sample.

The World Health Organization’s wellbeing index (WHO-5) (Bech et al., 1996) is a measure of positive wellbeing. The version used in this study was administered daily and the five-items were rated on a 6-point Likert scale measuring frequency from 0 “at no time” to 5 “all of the time”. A sample item from the WHO-5 is “Over the past 24 hours I have felt cheerful and in good spirits”. Total scores range from 0 to 25 where a higher score indicates more positive wellbeing. The WHO-5 demonstrated good internal consistency in a psychiatric sample (Cronbach’s $\alpha = 0.89$) and good convergent validity as shown by its strong correlation with other measures of psychological health.
(Newnham et al., 2010a). The WHO-5 has also been found to demonstrate sensitivity to change in a psychiatric setting (Newnham et al., 2010a) as well as in the present sample.

Therapists were required to complete the Therapist Prediction Questionnaire (TPQ), which was an extended adaptation from (Hannan et al., 2005). Therapists were first asked: “Using your clinical judgement, predict this patient’s outcome at the end of Day 10 of the CBT group. This patient will leave treatment (circle one prediction): recovered, improved but not recovered, unchanged, or deteriorated. The final item required therapists to estimate their patient’s current progress. Therapists were asked: “Considering this patient’s post-assessment functioning, rate this patient’s progress as of today’s session using your clinical judgement. Today this patient is (circle one response): recovered, improving as expected, but not recovered, making minimal or no improvement, or getting worse.”

Procedure

All patients in the present study took part in a manualised Cognitive Behaviour Therapy (CBT) group program (Manning et al., 1994; Page & Hooke, 2011). The CBT group programs are run by two therapists from 9am to 5pm over 10 consecutive working days with 6 to 8 patients per group. The CBT group sessions focus on managing depression, anxiety, and stress; and involve cognitive restructuring, behavioural experiments, interpersonal skills, and relapse prevention. At the beginning of each day of the CBT group program, patients were invited to complete the DI-5 and WHO-5. Therapists had access to this questionnaire data via a computer throughout the whole course of the CBT group program.

As part of routine clinical practice, on day five of the CBT group program, therapists were required to provide their patients with feedback on their DI-5 and WHO-5 scores in the form of a patient progress graph and group discussion. The graphs reflect
the raw scores for each measure over sessions one to five in addition to an expected treatment response curve which was generated using archival data. Thus, the graphs do not depict clinical significance categories (see Figure 1 for example WHO-5 feedback graph). The example WHO-5 feedback graph presented in Figure 1 depicts a patient whose wellbeing initially improved over the first three sessions but got worse on the fourth session, such that their progress fell out of the expected treatment response curve.

Figure 1. Screen capture of a patient’s WHO-5 wellbeing feedback graph with the expected treatment response curve.

The feedback component of the session was semi-structured and involved the entire group. Therapists prompted their patients to discuss their progress by encouraging them to identify patterns of change on their graphs and share insights into these change patterns. The therapists also encouraged patients to reflect on what factors may have been responsible for improvement, and what factors may have led to lapses with a view to adapting treatment accordingly. The feedback component of the CBT group was also implemented on the final day of treatment. At the end of day five, and once feedback had been given to the patients, the therapists were required to complete the TPQ.
Therapists were instructed to take into account the patients’ DI-5 and WHO-5 scores when making their predictions of outcomes and estimates of progress.

**Data Analysis**

**Clinical Significance.** In the present study, the Jacobson and Truax method of clinically significant change was used to determine actual patient progress and outcomes. This clinical significance method classifies patients based on two conditions. The first condition is whether the change was statistically meaningful and larger than the Reliable Change Index\(^2\) (RCI) of the measure. The second condition is whether the patient finished treatment within a pre-determined functional population range on the measure. For the DI-5, the cut-off score for the functional population was 6.17. Therefore, scores below 6.17 were considered to be in the functional population range. For the WHO-5, the cut-off score for the functional population was 10.8. Therefore, scores above 10.8 were considered to be in the functional population range. The RCI expresses each patient’s change in pre- and post-scores on the measure in standard error units of measurement and signals that a reliable change has occurred when this value exceeds an increase or decrease of 1.96.

As discussed, these conditions produce a minimum of four outcome categories: recovered (reliable improvement and functional population range), improved (reliable improvement but not in functional population range), unchanged (no reliable change), and deteriorated (reliable worsening). The clinical significance category for each individual patient was calculated at day four (for progress comparison) and day ten (for outcome comparison). Therapists’ predictions were then compared to patients’ clinical significance scores (actual progress and outcomes) at day four (for estimates of progress) and day ten (for predictions of outcomes).
Comparisons. The comparisons between therapists’ estimates of progress and actual patient progress, and therapists’ predictions of outcomes and actual patient outcomes were done using two methods. Firstly, the proportions of patients falling into each clinical significance category for therapists’ estimates and predictions were compared to the proportion in each category for actual patient progress and outcomes using two-proportion z-tests to test for significant differences. Secondly, therapists’ ability to estimate and predict progress and outcomes was assessed by calculating the agreement between the actual progress and outcomes and therapist estimated progress and outcomes. This done by running cross-tabs in SPSS.

Missing Values Analysis. A missing values analysis revealed that missing cases were missing completely at random; Little’s MCAR test: $\chi^2 (315) = 340.43, p = .155$. Therefore, expectation maximisation was conducted on the data for the purposes of further analysis.

Results

Overall descriptive statistics for the DI-5 and WHO-5 scores from day 1 to day 10 of the CBT groups are displayed in Table 1. In general, patients’ DI-5 scores decreased over the course of the CBT group, suggesting a reduction in symptoms consistent with recovery according to the Jacobson and Truax (1991) method of clinical significance categorisation. Patients’ WHO-5 scores increased over the course of the CBT group, suggesting an increase in positive wellbeing also consistent with recovery according to clinical significance categorisation (see Table 1).
Table 4.1

Means and standard deviations by day of treatment for the DI-5 and WHO-5

<table>
<thead>
<tr>
<th></th>
<th>DI-5</th>
<th>WHO-5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Day 1</td>
<td>10.03</td>
<td>6.34</td>
</tr>
<tr>
<td>Day 2</td>
<td>8.20</td>
<td>6.08</td>
</tr>
<tr>
<td>Day 3</td>
<td>8.48</td>
<td>5.75</td>
</tr>
<tr>
<td>Day 4</td>
<td>7.63</td>
<td>5.35</td>
</tr>
<tr>
<td>Day 5</td>
<td>6.85</td>
<td>5.49</td>
</tr>
<tr>
<td>Day 6</td>
<td>6.59</td>
<td>5.14</td>
</tr>
<tr>
<td>Day 7</td>
<td>6.78</td>
<td>5.22</td>
</tr>
<tr>
<td>Day 8</td>
<td>6.99</td>
<td>5.76</td>
</tr>
<tr>
<td>Day 9</td>
<td>5.63</td>
<td>5.20</td>
</tr>
<tr>
<td>Day 10</td>
<td>5.36</td>
<td>4.71</td>
</tr>
</tbody>
</table>

Patients who began treatment within the functional range on either of the measures were included for the first set of analyses. Further, patients within the functional range at the second time point measurement, despite their trajectory of change, were classified as a positive outcome for predictions or as positive progress for estimates. Therapists’ predictions of outcomes, estimates of progress, and the measures’ (DI-5 and WHO-5) predictions of outcomes (the empirical method of prediction) were measured in two ways:

1. Proportions: The proportion of patients estimated/predicted to be in each clinical significance category vs. the actual proportion of patients in each clinical significance category.
2. Agreement: Whether patients predicted/estimated to have a particular outcome actually had that predicted/estimated outcome.

**Therapists’ vs. the DI-5’s Empirical Predictions of Patient Outcomes**

**Proportions.** Three steps were taken in order to compare therapists’ predictions of outcomes, the DI-5’s predictions of outcomes, and actual DI-5 outcomes in terms of proportions in each clinical significance category. Firstly, the Jacobson-Truax method was used to calculate clinical significance categories for final outcomes (day 10) on the DI-5 (Jacobson & Truax, 1991). Patients achieving either a recovered or improved outcome were classified as having a positive outcome. Secondly, the DI-5’s predictions of outcomes were determined by using the day five calculations of clinical significance categories and comparing these to the clinical significance categories for the DI-5 at day 10 (final outcome). In other words, we assessed whether a patient’s day five DI-5 score could predict their final DI-5 score. Thirdly, frequencies by clinical significance category for therapists’ predictions of outcomes, the DI-5’s predictions of outcomes, and actual DI-5 outcomes were run in order to determine the therapists’ and DI-5’s predictions of outcomes (see Figure 2).
Figure 2. Comparison of therapist predicted, DI-5 predicted, and actual outcomes for the DI-5.

Differences between predictions and actual outcomes were tested using the two-proportion z-test with Bonferroni adjustment. The results showed that therapists predicted 82.8% (N = 173) of patients would achieve a positive outcome when 82.3% (N = 172) actually achieved a positive outcome on the DI-5. There was no significant difference between therapist predictions of outcomes and actual outcomes for the positive outcome category, $Z = 0.13, p = .90$. Therapists predicted that 15.8% (N = 33) of patients would leave treatment unchanged while 15.3% (N = 32) actually finished treatment unchanged; this difference was also not significant, $Z = 0.14, p = .90$. Finally, there was no significant difference between the proportion of patients therapists predicted to deteriorate (1.4%; N = 3) and the actual proportion of patients who deteriorated (2.4%; N = 5), $Z = -0.71, p = .48$.

In contrast, the DI-5 predicted that 71.3% (N = 149) of patients would leave treatment with a positive outcome, which was significantly less than the actual
proportion of patients who achieved a positive outcome, \( Z = -2.66, p < .01 \). There was also a significant difference between the DI-5’s prediction and actual outcomes for the unchanged category such that significantly less patients left treatment unchanged than the DI-5 predicted (24.4%; \( N = 51 \), \( Z = 2.33, p < .05 \). However, there was no significant difference between the proportion of patients the DI-5 predicted to deteriorate (4.3%; \( N = 9 \)) and the actual proportion who deteriorated, \( Z = 1.09, p = .28 \). In summary, therapists were more accurate than the DI-5 at predicting the proportion of patients who would leave treatment with a positive and unchanged outcome. However, both therapists and the DI-5 accurately predicted the proportion of patients who would deteriorate.

These results tell us how accurate therapists were at predicting the proportion of patients who would fall into each of the clinical significance categories at the end of treatment. However, the above results do not tell us the agreement between those predictions and actual outcomes. In other words, the results do not tell us whether a patient predicted to end treatment with a specific outcome actually ended treatment with that specific outcome. Therefore, the next analyses will address the agreement between therapists’ predictions of outcomes and actual outcomes for the DI-5.

**Agreement.** It was found that therapists correctly predicted 145 of the 172 patients (84.3%) who achieved a positive outcome, while the DI-5 correctly predicted 78.5% (\( N = 135 \)). This difference was not significant, \( Z = 1.39, p = .16 \). Therapists correctly predicted 5 of the 32 patients (15.6%) who left treatment unchanged, while the DI-5 correctly predicted 43.8% (\( N = 14 \)). This difference was statistically significant, \( Z = -2.46, p < .05 \). Finally therapists did not correctly predict any of the five patients who left treatment deteriorated, whereas the DI-5 correctly predicted one patient (20%). However, this difference was not significant, \( Z = -1.05, p = .29 \). In summary, the DI-5
was no better than therapists in terms of prediction and actual outcome agreement, except for the unchanged category.

Therapists' vs. the WHO-5’s Empirical Predictions of Patient Outcomes

Proportions. Frequencies were run in order to determine the proportion of patients in each clinical significance category as determined by the therapists’ predictions, the WHO-5’s predictions, and the actual outcomes. The WHO-5’s prediction of outcomes was determined by calculating the clinical significance categories at day five. Figure 3 presents therapist and WHO-5 predictions, as well as the actual outcomes for the WHO-5.

Figure 3. Comparison of therapist predicted, WHO-5 predicted, and actual outcomes for the WHO-5.

Differences were again tested using the two-proportion z-test with Bonferroni adjustment. It was found that actual positive outcome (61.2%; N = 128) for the WHO-5 was significantly different to therapists’ predictions such that therapists overpredicted positive outcome, $Z = 4.90, p < .01$. Therapists significantly underpredicted unchanged outcome which was 36.4% (N = 76) for the WHO-5, $Z = -4.79, p < .01$. However, there
was no significant difference between the proportion of patients therapists predicted to deteriorate and the actual proportion of patients who deteriorated (2.4%; N = 5), Z = -0.71, p = .48. The WHO-5 on the other hand, significantly underpredicted positive outcome (41.1%; N = 86), Z = -4.11, p < .01, and overpredicted unchanged outcome (55.5%; N = 116), Z = 3.93, p < .01. There was no significant difference between the proportion of patients the WHO-5 predicted to deteriorate (3.3%; N = 7) and the actual proportion who deteriorated, Z = 0.59, p = .56. In summary, therapists and the WHO-5 were not accurate at predicting the proportion of patients who would leave treatment in the positive and unchanged categories when comparing predictions to WHO-5 outcomes. However, both therapists and the WHO-5 correctly predicted the proportion of patients who would leave treatment in the deteriorated category for the WHO-5.

Agreement. Therapists were found to correctly predict 113 of the 128 patients (88.3%) who achieved a positive outcome, whereas the WHO-5 correctly predicted 60.2% (N = 77). This difference was significant, Z = 5.14, p < .01. Therapists correctly predicted 16 of the 76 patients (21%) who left treatment unchanged, while the WHO-5 correctly predicted 85.5% (N = 65). This difference was significant, Z = -7.97, p < .01. Finally, therapists did not correctly predict any of the five patients who left treatment deteriorated, whereas the WHO-5 correctly predicted one patient (20%). However, this difference was not significant, Z = -1.05, p = .29. In summary, in terms of agreement between predictions and actual outcomes, therapists were more accurate at predicting
positive outcome than the WHO-5, whereas the WHO-5 was more accurate at predicting unchanged outcome. There was no difference in agreement between therapists’ predictions and WHO-5 predictions for the deteriorated outcome.

**Therapists’ Estimates of Patient Progress**

**Proportions.** Two steps were taken in order to compare therapists’ estimates of patient progress (i.e., their patient’s progress as of today’s [day five] session) with actual DI-5 and WHO-5 progress in terms of proportions. Firstly, clinical significance categories at day five were calculated for both the DI-5 and WHO-5 using the Jacobson-Truax method (see Jacobson et al. (1984) and Jacobson and Truax (1991) for the relevant formulae). Consistent with previous research (Hannan et al., 2005), *recovered* or *improved* progress was defined as *positive progress*, while the unchanged and deteriorated categories reflected the appropriate clinical significance categories. Secondly, frequencies by clinical significance category for therapists’ estimates of progress and DI-5 and WHO-5 progress were run in order to determine therapists’ estimates of progress (see Figure 4).
Figure 4. Therapist estimates of progress compared to actual DI-5 and WHO-5 progress.

It was found that there was no significant difference between therapists’ estimates of progress and actual DI-5 progress for the positive progress, $Z = -1.47, p = .14$; unchanged, $Z = 1.95, p = .05$, and deteriorated, $Z = -1.09, p = .28$ categories. This suggests that therapists accurately estimated the proportion of patients in each clinical significance category at day five for the DI-5. For the WHO-5 on the other hand, there was a significant difference between therapist estimates and WHO-5 progress for the positive progress category, $Z = 4.61, p < .01$ and unchanged category, $Z = -4.63, p < .01$, such that therapists overestimated positive progress and underestimated negative progress. However, there was no significant difference between estimated deterioration and actual WHO-5 deterioration, $Z = 0, p = 1$. This suggests that therapists did not accurately estimate the proportion of patients who made positive progress or remained unchanged by day five for the WHO-5. However, therapists were able to accurately estimate the proportion of patients who had deteriorated on the WHO-5 by day five.
These results tell us how accurate therapists were at estimating the proportion of patients who would fall into each of the clinical significance categories at day five. However, the above results do not tell us the agreement between those estimates and actual progress. For example, whether a patient estimated to have made positive progress by day five actually made positive progress by day five. The next analyses will address the agreement between therapists’ estimates of progress and actual progress for the DI-5 and WHO-5, respectively.

**Agreement.** For the DI-5, it was found that therapists correctly estimated 71.8% of the patients who had made positive progress, 45.1% of the patients who were unchanged at day five and, 22.2% of the patients who had deteriorated by day five (see Table 1). For the WHO-5, therapists accurately estimated 81.8% of the patients who made positive progress, 44% of the patients who were unchanged at day five, and none of the patients who were deteriorated by day five (see Table 1). These results suggest that most of the patients who therapists estimated to have made positive progress by day five had in fact made positive progress for both measures. However, therapists accurately estimated the progress of just under half of the patients who were unchanged by day five and only two of the deteriorated patients.
Table 1.

Agreement between therapists’ estimates of progress and actual DI-5 and WHO-5 progress.

<table>
<thead>
<tr>
<th>Therapist Estimate</th>
<th>Therapist DI-5 Estimate</th>
<th>WHO-5 Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td>N</td>
<td>N (%)</td>
</tr>
<tr>
<td>Positive Progress</td>
<td>107 (71.8%) of 149</td>
<td>72 (81.8%) of 88</td>
</tr>
<tr>
<td>Unchanged</td>
<td>23 (45.1%) of 51</td>
<td>51 (44%) of 116</td>
</tr>
<tr>
<td>Deteriorated</td>
<td>2 (22.2%) of 9</td>
<td>0 (0%) of 5</td>
</tr>
</tbody>
</table>

Post-assessment Functional Patients Removed

As previous research has typically removed patients who began treatment within the functional range on the measure/s of interest, all of the analyses were re-run with the exclusion of these patients. Seventy-six cases met this criteria and were removed thus leaving a total of 133 cases. It was found that excluding these patients made no difference to the pattern of results and z-tests, except for one analysis in relation to the DI-5. Therapists significantly outperformed the DI-5 at predicting positive outcome in terms of agreement, Z = 2.37, p < .05.

Discussion

The aim of the present study was to assess therapists’ ability to predict patient outcomes and estimate current progress, and whether feedback to the therapist and patient would assist therapists’ accuracy. The findings overall suggest that feedback assisted therapists’ predictions of outcomes and estimates of progress. Therapists were able to correctly predict the proportion of patients who would fall into the positive outcome, unchanged, and deteriorated clinical significance categories for the DI-5 (symptoms). This finding is inconsistent with previous research which found that
therapists overpredicted the proportion of patients who would have a positive outcome and underpredicted the proportion who would have a negative (unchanged or deteriorated) outcome (Chapman et al., 2012; Hannan et al., 2005). The same pattern of results were found for therapists’ estimates of current progress for symptoms, such that therapists accurately estimated the proportion of patients who had made positive progress, remained unchanged, or deteriorated by day five. This finding is also inconsistent with previous research as Hannan et al. (2005) found that therapists still overpredicted positive and underpredicted negative progress when asked to estimate their patients’ current status.

The findings for the WHO-5 (wellbeing) on the other hand, were generally consistent with previous research (Chapman et al., 2012; Hannan et al., 2005), as therapists overpredicted positive outcomes and underpredicted unchanged outcomes. However, therapists correctly predicted the proportion of patients whose wellbeing had deteriorated by the end of treatment. Again, the same pattern of results were found when therapists were asked to estimate their patients’ current progress. Specifically, therapists overestimated positive progress, underestimated unchanged progress, but correctly estimated the proportion of patients who had deteriorated by day five for wellbeing.

Previous research has focused on the fact that therapists do not seem to even consider that any of their patients will deteriorate. For example, Hannan et al. (2005) found that therapists predicted that only 0.01% of patients would deteriorate when 7.3% actually deteriorated. Interestingly, in the present study therapists correctly predicted the proportion of patients who would deteriorate for both symptoms and wellbeing. Thus, one could argue that allowing the therapists access to feedback, and the process of therapists voicing this feedback to their patients, assisted with their judgements.

Although therapists sometimes predicted the wrong patients would deteriorate, one
could argue that the feedback at least made therapists more open to the possibility of
deterioration.

Also of interest is that in the present study, therapists’ predictions of outcomes
and estimates of progress were more aligned with patients’ symptoms rather than their
wellbeing. This is consistent with research suggesting that therapists base their
judgements of patient progress and outcomes on the patient’s symptoms more so than
other factors (Hatfield et al., 2010). Previous research used an outcome measure (OQ-
45) which measures symptoms and wellbeing together rather than separately, which
could have been why therapists’ predictions of outcomes and estimates of progress were
poor in these studies (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010).
Had the authors analysed symptoms and wellbeing separately, they may have found a
similar pattern of results to the present study. This finding therefore highlights the
importance of monitoring symptoms and wellbeing separately, and supports research
suggesting that although negatively correlated, symptoms and wellbeing are separate
constructs (Newnham et al., 2010b).

The finding that therapists were able to accurately predict and estimate the
proportion of patients, on at least one measure, who had a positive outcome, remained
unchanged, or deteriorated is a promising improvement upon previous research.
However, we also needed to examine whether therapists could not only accurately
predict/estimate the proportion of patients in each clinical significance category, but
whether patients predicted/estimated to have a particular outcome actually achieved that
outcome. This concept was termed agreement between therapist predictions/estimates
and actual outcomes/progress in the present study. Agreement between therapists’
predictions of outcomes and actual outcomes on both measures was excellent for the
positive outcome category. Specifically, the majority of patients who therapists
predicted to have a positive outcome, did in fact achieve a positive outcome. However, therapists correctly predicted only 15.6% and 21% of the unchanged patients for the DI-5 and WHO-5, respectively; and of the patients therapists predicted to deteriorate, none left treatment deteriorated. Upon further examination however, all of the patients who therapists predicted would deteriorate left treatment with an unchanged outcome on at least one of the measures. An unchanged outcome would still be considered a negative outcome, so this suggests that therapists were alert to the fact that these patients would not improve.

When therapists were asked to estimate their patients’ current progress, the results for agreement between therapists’ estimates and actual progress were slightly different compared to predictions. Agreement between therapists’ estimates of progress and actual progress on both measures was again excellent for the positive progress category. Specifically, the majority of patients who therapists estimated to have made positive progress had in fact made positive progress. For the unchanged category, therapists correctly estimated the progress of approximately half of the unchanged patients. This is an improvement from agreement between therapists’ predictions of outcomes and actual outcomes for the unchanged category, and suggests that therapists do notice when patients have not made any progress at the time, but cannot accurately predict who will end treatment unchanged. For the deteriorated category, therapists correctly identified two (22.2%) of the patients whose symptoms had deteriorated, but none of the patients whose wellbeing had deteriorated by day five. This finding is consistent with Hatfield et al. (2010) who found that therapists correctly identified 21% of deteriorated patients in their progress notes.

Although improvements from previous research were found for therapists estimates of current progress, one might wonder why therapists could not correctly
estimate every patient’s progress because they had the information available to them in
the form of a graph. It is important to note however, that the graphs did not denote
clinical significance categories, and therapists were instructed to take into account their
patients’ symptom AND wellbeing scores when making their estimates of current
progress. As it is possible for a patient to be improving on one measure and
deteriorating on the other, this could have led to some confusion in the therapists.
Future research could ask therapists to predict outcomes and estimate progress on a
measure assessing symptoms alone, to determine whether agreement between
predictions and actual outcomes is improved.

Further, it is possible that social desirability, the tendency to respond in a more
socially acceptable way (Anchor, Vojtisek, & Berger, 1972; Crowne & Marlowe, 1960),
could have been a factor at play here. Perhaps patients who were not progressing well
downplayed their deterioration or no change in the session to please their therapists.
This could have led to more confusion in the therapists such that they wondered
whether to trust the patient’s explanation in the group, their DI-5 scores, or their WHO-
5 scores. Future research could also measure social desirability and see whether this
impacts feedback discussions in the session.

Upon further examination of the agreement results, the five patients who
therapists estimated as deteriorated, were all unchanged or deteriorated on at least one
of the measures by day five. Again, this suggests that therapists would have been alert
to these patients’ lack of progress at day five, which in turn could have triggered the
therapist to initiate some kind of in-therapy action in order to improve outcomes for
these patients. In fact, two of these patients went on to achieve a positive outcome by
the end of treatment, which could be attributed to the feedback. It is unclear however,
what exactly led these two patients to achieve a positive outcome. The positive
outcomes of these patients could have been due to factors outside of the therapists’
control. For example, the non-linear trajectory of progress (Lutz et al., 2013). However,
the feedback intervention theory would suggest that the therapist initiated some kind of
in-therapy action which led to the improved outcome (Kluger & DeNisi, 1996). The
therapeutic assessment theory on the other hand, would suggest that the process of the
patient receiving feedback was therapeutic for the patient in some way (Finn &
Tonsager, 1992). Hatfield et al. (2010) found that when therapists became aware of
patient deterioration, their most likely response was in-therapy action. Future research
could expand on the study by asking the patients who were at risk for treatment failure,
as well as the therapists, what it was that turned their outcome into a positive one
following feedback. It is quite possible that both theories of how feedback works come
into play.

The two theories that aim to explain why feedback works to improve patient
outcomes could also help to explain why feedback improved therapists’ judgements of
progress and outcomes. The feedback intervention theory would suggest that therapists’
judgements of progress and outcomes improved because therapists had access to extra
information (Kluger & DeNisi, 1996). In other words, therapists had access to
observational data and information communicated by therapists in the course of therapy
which they could use in conjunction with the DI-5 and WHO-5 scores. The therapeutic
assessment theory however, would suggest that perhaps the process of sharing the
feedback with the patient opened up communication from the patient to the therapist
about how they were progressing in treatment (Finn & Tonsager, 1992).

In the present study, we also compared therapists’ predictions of outcomes to the
measures’ (DI-5 and WHO-5) predictions, an empirical method. It was found that the
empirical method was no more accurate than therapists at predicting the proportion of
patients who would deteriorate. In fact therapists were more accurate at predicting the proportion of patients who would achieve a positive outcome or remain unchanged for symptoms. This is inconsistent with previous research which found that the empirical method was far superior to therapists at predicting patient outcomes (Hannan et al., 2005). However, in terms of agreement with actual outcomes, the empirical method outperformed therapists at accurately predicting which patients would leave treatment unchanged. This however, could be due to the fact that the empirical method significantly overpredicted the proportion of patients who would leave treatment unchanged, thus making it easier for the empirical method to accurately detect these patients. This finding is somewhat consistent with the findings of Hannan et al. (2005) who found that the empirical method they used overpredicted deterioration. However, in the present study, the majority of false alarms (i.e., patients predicted to leave treatment unchanged but did not) achieved a positive outcome. Therefore, any extra resources allocated to these false alarm patients may be wasted. Further, therapists significantly outperformed the empirical method in terms of agreement between predictions and actual outcomes for the positive outcome category, and there was no difference for deterioration.

Overall, the empirical method was not superior to therapists at predicting outcomes. However, the empirical method used in the present study was a simplistic one, thus it could be argued that a more sophisticated empirical method of prediction would outperform therapists in their predictions. Future research could compare a more sophisticated empirical method of prediction with therapists’ predictions while feedback is available. Another limitation of the present study was the sample size of both the patients and therapists. As the deteriorated clinical significance category was only a small proportion of the overall sample, it was more difficult to get an accurate
representation of how well therapists could predict deterioration. Further, a no feedback comparison group was not included in the present study. Finally, the use of the therapist providing feedback on day five as a proxy for an implementation check was a limitation as there was no definitive way of knowing whether the therapists did in fact provide feedback to patients on that day even though they are required to. Future research could ask the patients whether they were given feedback on their DI-5 and WHO-5 scores on fifth day of treatment to ensure feedback was given.

Despite these limitations, the present study suggests that therapists do have the ability to accurately predict the proportion of patients who will go on to achieve a positive outcome, remain unchanged, or deteriorate in their symptoms when feedback is available. Further, therapists accurately estimated the proportion of patients who had made positive progress, remained unchanged, or deteriorated by day five for symptoms, and with the provision of feedback. This is a promising improvement upon previous research and suggests that with the provision of feedback, therapists can accurately predict patient outcomes and estimate current progress in terms of the proportion of patients in each clinical significance category. However, even with feedback, agreement between therapists’ predictions/estimates and actual outcomes/progress was unsatisfactory for the unchanged and deteriorated categories. This highlights the importance of therapists paying attention to routine outcome monitoring data in order to accurately identify patients at risk for treatment failure, thus giving them the opportunity to prevent negative outcomes. The results also present further benefits of feedback to both the therapist and patient, as it appeared to aid therapists’ awareness of patient no change and deterioration which could in turn improve outcomes overall.
Chapter 5

A Longitudinal Empirical Method of Predicting Patient Outcomes
The results from Chapter 4 suggest that feedback improved therapists’ predictions of outcomes and estimates of current progress to some extent. With the provision of feedback to both the therapist and patient, therapists correctly predicted and estimated the proportion of patients in each clinical significance category. This is a promising result as it suggests that therapists were more realistic in their predictions of outcomes and estimates of progress compared to previous studies (Chapman et al., 2012; Hannan et al., 2005). Further, therapists were more aware that some patients leave treatment having made no change or worse off, such that they did not overestimate positive and underestimate negative outcomes and progress. Despite these promising results however, agreement between therapists’ predictions/estimates and actual patient outcomes/progress was insufficient, especially for negative outcomes. In other words, therapists were unable to correctly identify which patients had made negative progress midway through therapy, and which patients would ultimately make no change or deteriorate. This is problematic because it means that therapists were not necessarily aware of who of their patients were making no change or deteriorating (as measured by the self-report indices). As a result, therapists would not know which patients to target any in-therapy action with.

The empirical method of prediction however, was found to be superior to therapists at predicting negative outcomes in terms of agreement between predictions and actual outcomes. This suggests that the empirical method could more accurately predict which patients would leave treatment unchanged or deteriorated compared to therapists. However, the empirical method was still imperfect at predicting negative outcomes in that there were a number of false alarms identified for the unchanged
category. In other words, a large proportion of the patients that the empirical method predicted to make no change, actually ended treatment with a positive outcome. Therefore, how can we help therapists to predict negative outcomes and identify negative progress as it occurs, so that they can prevent those negative outcomes from occurring? In the final study (Chapter 6) we compare the empirical method used in Chapters 2 and 4 to a more sophisticated and complex empirical method of predicting negative outcomes, to determine whether this method will be superior and therefore more helpful to therapists during treatment. The empirical method used in Chapters 2 and 4 was calculated by comparing mid-treatment clinical significance categorisations with clinical significance categorisations at the end of treatment, for both the DI-5 and WHO-5. That is, this empirical method assessed whether early change, as defined by clinical significance, could predict final outcomes (Jacobson & Truax, 1991) This empirical method will now be referred to as the clinical significance early change method.

The empirical method that will be used in Chapter 6 is a longitudinal modelling method named *Growth Mixture Modelling* (GMM) (Duncan & Duncan, 2004; Li, Duncan, Duncan, & Acock, 2001; B. Muthén & Shedden, 1999). GMM allows researchers to detect unobserved groups of individuals based on their entire trajectory of change across treatment (Duncan, Duncan, & Strycker, 2013). As a result, GMM is ideal for exploring inter-individual (between-person) differences in intra-individual (within-person) change over time (Duncan & Duncan, 2004; Duncan et al., 2013; McArdle & Bell, 2000). Researchers have found that early change as defined by GMM can successfully predict final outcomes (Haas, Hill, Lambert, & Morrell, 2002; Lutz et al., 2014; Lutz, Stulz, & Köck, 2009; Stulz, Lutz, Leach, Lucock, & Barkham, 2007). Early change as defined by GMM will therefore be used in Chapter 6 as a comparison
to the early change clinical significance method used in Chapters 2 and 4 (Jacobson & Truax, 1991). This method will be termed; the GMM early change method.

Previous research varies in its definition of early change. For example, Haas et al. (2002) defined early change as the first three sessions of treatment, Lutz et al. (2014) defined early change as the first five sessions, while Lutz et al. (2009) used as many as the first eight weeks (sessions) of treatment to define early change. GMM requires at least three time points to be conducted. Therefore, in Chapter 6 we needed to define early change as at least the first three days of treatment. The CBT group treatment used in the present thesis involves sessions held on 10 consecutive working days. Therefore, there is arguably limited time between session one and three for significant changes to occur. Further, Chapter 4 assessed therapists’ predictions of outcomes on day five (the fifth session), after therapists had given feedback to their patients. Therefore, to make for an easier comparison, we chose to assess early change as the first five days of treatment in Chapter 6.

Early change is defined differently between studies and there is no real consensus within the literature for what constitutes as early change (Lutz et al., 2014; Rubel et al., 2015). In the present study, early change was defined using the first five days of DI-5 and WHO-5 scores for several reasons. Firstly, although the fifth data point is obtained on the fifth day of treatment, this data point only reflects symptoms and wellbeing for the first four days of treatment. This is because patients complete the measures in the morning of the fifth day of treatment and the measures ask for responses based on the previous 24 hours. Therefore the first five data points for the DI-5 and WHO-5 would still be considered to reflect the early stages of treatment (the first four days). Secondly, we wanted to be able to compare the empirical methods’ predictions of outcomes with the therapists’ predictions in Chapter 4. Chapter 6 will
therefore ask the question, can early change defined by GMM over the first five days of treatment predict no change and deterioration better than the clinical significance early change method?
Chapter 6

A Comparison Between the Clinical Significance and Growth Mixture Modelling Early Change Methods at Predicting Negative Outcomes
Abstract

**Objective:** Routine outcome monitoring benefits treatment by identifying potential no change and deterioration. The present study compared early change as defined by two empirical methods at predicting negative outcomes on self-report symptom and wellbeing measures to determine if using longitudinal modelling could improve existing approaches.

**Method:** 1,467 voluntary day patients participated in a ten-day group Cognitive Behaviour Therapy (CBT) program and completed the symptom and wellbeing measures daily. Early change, as defined by (a) the clinical significance method and (b) longitudinal modelling, were compared on each measure.

**Results:** Early change, as defined by the simpler clinical significance method, was found to be superior at predicting negative outcomes than longitudinal modelling. The longitudinal modelling method failed to detect a group of deteriorated patients, and agreement between the early change methods and final unchanged outcome was higher for the clinical significance method.

**Conclusions:** Therapists could use the clinical significance early change method during treatment to alert them of patients at risk for negative outcomes, which in turn could allow therapists to prevent those negative outcomes from occurring.
A Comparison Between the Clinical Significance and Growth Mixture Modelling

Early Change Methods at Predicting Negative Outcomes

Cognitive Behaviour Therapy (CBT) is effective in treating patients with a variety of different psychological disorders (Hofmann et al., 2012; Lambert, 2013a), and the majority of patients achieve a positive outcome (Hofmann et al., 2012; Westbrook & Kirk, 2005). However, a proportion of patients do not make meaningful progress and a smaller number leave treatment worse off than when they entered. Although rates of negative outcomes vary from study to study, research in routine clinical settings finds that up to 60% of patients make no change and around 5% to 10% of patients deteriorate (Lambert, 2013b; Saxon et al., 2017).

Arguably the most popular method used for defining patient outcome, both in research and clinical practice, is the Jacobson-Truax clinical significance method (Jacobson et al., 1984; Jacobson & Truax, 1991; Ogles et al., 2001). The clinical significance method classifies patient outcomes based on whether or not the patient finished treatment within a pre-determined functional population range on the measure of interest; and whether their change was statistically significant (Ronk et al., 2013). Combining these two classifications produces four clinical significance categories for patients who begin treatment within the dysfunctional population range: recovered, improved, unchanged, or deteriorated. In a real-world clinical setting however, some patients begin treatment within the functional population range for a variety of potential reasons as outlined in Chapter 3. For example, in a diagnostically heterogeneous sample, the measure/s may not capture the patient’s presenting problems. Further, when more than one measure is used, it is possible for patients to fall within the dysfunctional population range on one measure but the functional population range on another. If patients beginning in the functional population range also leave treatment within the
functional population range, then they would still be classified as a positive/recovered outcome from a service perspective.

When this clinical significance method is used, therapists can identify the proportion of patients who have made no change or deteriorated by the end of treatment. However, this information is ineffective for preventing no change and deterioration in practice, as therapists become aware only after treatment has ended. If therapists could detect when their patients are at risk for deterioration or no change, then they could respond in real-time. However, research suggests that therapists are poor at estimating patients’ current progress and predicting their final outcomes (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). For example, Hannan et al. (2005) asked therapists to rate patient progress as of today’s session using the four clinical significance categories (Jacobson & Truax, 1991). Therapists overestimated the frequency of improvement and underestimated the rate of no change and deterioration. Similarly, Hatfield et al. (2010), after reviewing patients’ progress notes, found that therapists detected only 21% of deteriorated patients. Although Chapters 2 and 4 found some improvements to therapists’ estimates of patient progress with the inclusion of feedback, therapists still struggled to accurately identify negative progress compared to positive progress.

Despite therapists’ limited ability to estimate current progress, researchers have also investigated therapists’ ability to predict their patients’ final outcomes. For example, Hannan et al. (2005) asked therapists to predict their patients’ outcomes on the OQ-45 (Lambert et al., 1996), based on their clinical judgement and experience alone. The therapists were asked to decide for each patient whether: “This client will (choose one prediction) recover, improve but not recover, make no progress in treatment, or get worse” (Hannan et al., 2005, p. 160). Therapists predicted that only 0.01% of patients
would deteriorate; which was 730 times less than the actual deterioration rate across the study (7.3%). Chapman et al. (2012) examined therapists’ predictions of patient outcomes in a group therapy setting. Following the third group session, therapists in this study were asked to predict their patients’ outcomes as either reliably improved, no significant change, or reliably worse. Consistent with Hannan et al. (2005), therapists overpredicted positive outcomes and underpredicted negative outcomes.

Due to the shortcomings of therapists’ estimates of current patient progress and predictions of outcomes, researchers have sought other ways of identifying negative change and predicting negative outcomes. For example, several researchers have compared two methods of identifying patients at risk for negative outcomes: a rationally-derived method based on clinical judgements using early response to treatment, dose-response relationship, and clinically significant change information; and an empirically-derived method based on expected recovery curves with tolerance intervals generated from historical therapy data from various settings (Hannan et al., 2005; Lambert, Whipple, Bishop, et al., 2002; Lutz et al., 2006; Spielmans et al., 2006). Although both methods were effective at identifying patients at risk for negative outcomes, the empirical (statistically-based) method consistently outperformed the rational (clinical judgement-based) method. While comparing these two methods, the researchers observed that patients who finished treatment with a negative outcome showed early deterioration. Lambert (2010) therefore suggested that it would be useful to examine the trajectories that unchanged and deteriorated patients produce early in treatment.

Fortunately, an empirically based longitudinal statistical analysis that allows researchers to group patients based on their trajectories of change over the course of treatment has become available, the Growth Mixture Modelling (GMM) method.
GMM allows for researchers to identify unobserved groups of individuals based on their trajectories of change across time (Duncan et al., 2013). More specifically, GMM identifies unobserved subgroups, describes the longitudinal change within each unobserved subgroup, and reveals differences in change among the unobserved sub-groups (Ram & Grimm, 2009). Therefore, GMM is appropriate for exploring inter-individual (between-person) differences in intra-individual (within-person) change over time (Duncan & Duncan, 2004; Duncan et al., 2013; McArdle & Bell, 2000).

Researchers have used GMM to determine whether early change trajectories in therapy can predict final outcomes. For example, Haas et al. (2002) found that rapid positive response to treatment, within the first three sessions, led to superior outcomes at termination and follow-up in a university counselling centre in the United States. Stulz et al. (2007) examined the trajectories of early change in various outpatient clinics in the United Kingdom, using GMM as well. The GMM identified a group of patients who began treatment highly impaired and made rapid improvements within the first six sessions. Over 90% of this group of early positive responders went on to achieve improvements at discharge, suggesting that early rapid improvement predicted positive outcomes. Another study that used GMM to assess early change in self-rated depression across the first eight weeks of treatment, identified three groups of patients (Lutz et al., 2009). Of note was the group of patients who began treatment with moderate to severe depression and showed rapid early improvement. All of the patients in this group (100%) showed reliable improvement at 16 weeks, which again suggests that early response to treatment predicts positive outcomes (Lutz et al., 2009).

More recently, Lutz et al. (2014) used GMM, over the first five sessions of treatment, to identify different patterns of early change in patients with panic disorder at
four university counselling services in the United States. Four groups of patients were identified by the GMM: a rapidly improving group, initially high-symptom and slowly improving group, initially low-symptom and slowly improving group, and an early deteriorating group. This was one of the few studies in the literature to identify a group of early deteriorating patients. This group of early deteriorating patients showed the least amount of improvement at termination compared to the other three groups. Specifically, none of the patients in the early deteriorating group had made reliable improvement by the end of treatment. This suggests that, like early positive response predicts positive outcomes, early negative response also predicts negative outcomes (Lutz et al., 2014). Lutz et al. (2013) assessed sudden gains and sudden losses in patients across the whole trajectory of treatment using the reliable change index (Jacobson & Truax, 1991). The authors found that sudden gains occurred most commonly at the beginning of therapy whereas it was more difficult to pinpoint when sudden losses were more likely to occur. Patients who experienced sudden losses were found to have significantly lower effect sizes in a number of different outcome measurements. The authors argued that it is therefore crucial to identify sudden losses, i.e., deterioration, as early as possible in treatment to prevent potential treatment failure (Lutz et al., 2013).

Other researchers have compared different methods of identifying early change to predict final outcomes. For example, Rubel et al. (2015), compared the GMM early change method to clinical significance early change methods in identifying early positive response to treatment (first three sessions). The authors found that there were strengths and weaknesses associated with both methods. The clinical significance early change methods were more sensitive in predicting final outcomes in that they were able to identify a larger number of early positive responders than the GMM early change
method. However, the GMM early change method was more specific in predicting treatment success in that the patients identified as early positive responders were likely to be true early positive responders. These results suggest that there are advantages and disadvantages to both the clinical significance and GMM early change empirical methods at predicting positive outcomes.

The early change literature has focused primarily on identifying patients who show early positive response to treatment. The problem with focusing on early positive change alone, is that only patients who are likely to achieve positive immediate and long-term outcomes are identified (Haas et al., 2002). This information could be useful in determining which patients may be ready for early discharge or termination of treatment. However, focusing on early positive change does not help us identify patients who are unlikely to respond well to treatment (i.e., make no change or deteriorate). Therefore, mental health service providers and therapists would also be interested in identifying early negative response to treatment as it has been shown to predict negative outcomes (Lutz et al., 2014). If early negative response is predictive of negative outcomes, then therapists can guide these patients back on track for a positive outcome. Little research has examined early negative change trajectories and their ability to predict negative outcomes (unchanged and deteriorated). Furthermore, despite the clinical significance early change method predicting some negative outcomes it does not predict all, and has the tendency to identify a higher rate of false alarms. Therefore, it would be useful to determine whether the GMM method can predict no change and deteriorators more effectively. GMM data could then be used in real time to identify patients at risk for negative outcomes so that therapists can make in-therapy action to guide these patients onto a positive outcome trajectory. The present study therefore
aimed to compare and contrast early change as defined by the clinical significance and GMM methods, at predicting patient no change and deterioration.

Method

Participants

Participants were 1,467 voluntary day patients recruited from a private psychiatric clinic. Questionnaire data was obtained from patients attending the clinic from March 2009 to November 2013. Patients’ ages ranged from 17 to 82 years old with a mean of 39.21, and 59.6% were female. Patients were diagnosed by their treating psychiatrist using ICD-10 criteria (World Health Organization, 1992), and the main primary diagnoses were affective (62.1%), neurotic (27.9%), and substance abuse (4.9%) disorders. The University of Western Australia Human Research Ethics Committee approved the study protocol prior to commencement, and patients provided informed consent as part of the routine admission procedure to the hospital.

Measures

In the present study, patients completed two brief measures daily; the five-item Daily Index (DI-5; a measure of psychological distress) (Dyer et al., 2014) and the World Health Organization’s wellbeing index (WHO-5; a measure of positive wellbeing) (Bech et al., 1996). The decision to analyse these two concepts separately in the present study was based on two reasons. Firstly, research suggests that different outcomes for the same patient may not necessarily tell us the same story about that patient’s outcomes (Yasky et al., 2015). For example, Newnham et al. (2010b) found that while providing feedback on wellbeing did not significantly improve wellbeing, feedback improved aspects of psychological distress, suggesting that although negatively correlated, they are separate constructs. Secondly, research suggests that providing feedback to therapists and patients on their symptoms (DI-5) and wellbeing
(WHO-5) reduces deterioration rates over and above providing feedback on wellbeing alone (Dyer et al., 2016). Therefore, it was useful to determine whether the early change trajectories as defined by the clinical significance and GMM methods differed for the DI-5 and WHO-5 in the present study.

The five-item Daily Index (DI-5) (Dyer et al., 2014) is a self-report measure designed to track five aspects of affective psychological distress, including depression, anxiety, worthlessness, coping behaviours, and suicidal ideation. The DI-5 is administered daily using a 6-point Likert scale where participants indicate frequency of the above symptoms over the past 24 hours from 0 (“at no time”) to 5 (“all of the time”). Scores range from 0 to 25 such that a higher score is indicative of higher affective psychological distress. The DI-5 has been found to be an appropriate measure of symptomatology for use in a psychiatric hospital setting as it correlated strongly with other mental health measures (Dyer et al., 2014). It also demonstrated high internal consistency (Cronbach’s $\alpha = 0.86$) and acceptable test-retest reliability ($r = 0.64$) for a state measure (Dyer et al., 2014). The DI-5 has also been found to be sensitive to change as shown by the significant difference between mean intake and discharge scores in a psychiatric setting (Dyer et al., 2014) and in the present sample.

The World Health Organization’s wellbeing index (WHO-5) (Bech et al., 1996) is a self-report measure of positive wellbeing. The index consists of five items rated on a 6-point Likert scale measuring frequency from 0 (“at no time”) to 5 (“all of the time”). Participants are asked to rate each item based on the previous 24 hours (adaptation by Newnham et al., 2010a) and scores range from 0 to 25 such that a higher score indicates more positive wellbeing. The WHO-5 exhibits high internal consistency in a psychiatric setting (Cronbach’s $\alpha = 0.89$) and high convergent validity as shown by its strong correlation with other measures of psychological health (Newnham et al., 2010a).
Chapter 6: Clinical Significance vs. Growth Mixture Modelling

The WHO-5 has also been found to demonstrate sensitivity to change in a psychiatric setting (Newnham et al., 2010a) and in the present sample.

**Procedure**

As part of routine clinical practice, participants took part in a Cognitive Behaviour Therapy (CBT) group program which ran from 9am to 5pm over 10 consecutive working days (Page & Hooke, 2011). Group sessions focused on managing depression, anxiety, and stress, cognitive restructuring, behavioural experiments, interpersonal skills, and coping with setbacks. At the beginning of each day, patients were invited to complete both the DI-5 and WHO-5. The therapists had access to this data via computer in the form of two graphs. The two graphs depicted raw scores for the DI-5 and WHO-5, respectively, plus an expected treatment response curve which was generated for both measures based on archival data.

On day five, the therapist provided the patients with feedback on their DI-5 and WHO-5 scores in the form of a progress graph and group discussion during the session. The graphs present the raw scores for each measure over sessions in addition to an expected treatment response curve which was generated from archival data. Thus, the graphs do not depict clinical significance categories (see *Figure 1* for example WHO-5 feedback graph). The example WHO-5 feedback graph presented in *Figure 1* depicts a patient whose wellbeing initially improved over the first three sessions but got worse on the fourth session, such that their progress fell out of the expected treatment response curve.
Chapter 6: Clinical Significance vs. Growth Mixture Modelling

Figure 1. Screen capture of a patient’s WHO-5 wellbeing feedback graph with the expected treatment response curve.

The feedback component of the session was semi-structured and involved the entire group. Therapists prompted their patients to discuss their progress by encouraging them to identify patterns of change on their graphs and share insights into these change patterns. The therapists also encouraged patients to reflect on what factors may have been responsible for improvement, and what factors may have led to lapses with a view to adapting treatment accordingly. The feedback component of the CBT group was also implemented on the final day of treatment.

Data Analysis

Early Change. As mentioned previously, early change is defined differently between studies and there is no real consensus within the literature for what constitutes as early change. In the present study, early change was defined using the first five days of DI-5 and WHO-5 scores. Although the fifth data point is obtained on the fifth day of treatment, this data point only reflects symptoms and wellbeing for the first four days of treatment. This is because patients complete the measures in the morning of the fifth day of treatment and the measures ask for responses based on the previous 24 hours.
Therefore the first five data points for the DI-5 and WHO-5 would still be considered to reflect the early stages of treatment (the first four days).

**Clinical Significance.** Early change was initially assessed using Jacobson and Truax’s method of clinically significant change. The clinical significance method classifies patient outcomes based on two conditions. The first condition is whether the change was statistically significant and larger than the Reliable Change Index\(^2\) (RCI) of the measure. The second condition is whether the patient finished treatment within a pre-determined functional population range on the measure. For the DI-5, the cut-off score for the functional population was 6.17. Therefore, scores below 6.17 were considered to be in the functional population range. For the WHO-5, the cut-off score for the functional population was 10.8. Therefore, scores above 10.8 were considered to be in the functional population range. The RCI expresses each patient’s change in pre- and post-scores on the measure in standard error units of measurement and signals that a reliable change has occurred when this value exceeds an increase or decrease of 1.96.

As discussed, these conditions produce a minimum of four outcome categories: recovered (reliable improvement and functional population range), improved (reliable improvement but not in functional population range), unchanged (no reliable change), and deteriorated (reliable worsening). In the present study, the clinical significance early change method was used to categorize each individual patients’ change on both the DI-5 and WHO-5. This method of classifying early change will be referred to as the clinical significance early change method. In order to determine how well the clinical significance early change method was at predicting final outcomes, agreement between day five and day ten clinical significance categories for each patient was then calculated and reported as a percentage.
GMM. Early change was also assessed using GMM. Firstly, a log-linear relationship was assumed between the amount of treatment and outcome as is widely accepted in the psychotherapy literature (Kopta & Lowry, 2002). Typical patterns of early change on the DI-5 and WHO-5 scores over the first five sessions were then identified using GMM. The GMM method enables the identification of unobserved groups of patients with shared trajectories of change across time. For the present GMMs, within-class variances were fixed to zero but the intercept variances were freely estimated. In the present study, the GMMs were conducted using the Mplus Version 7.3 software (L. Muthén & Muthén, 2014). This method of classifying early change will be referred to as the clinical significance early change method.

In order to determine the GMM groups’ ability to predict final outcomes, the corresponding clinical significance categories were calculated using day one and day five DI-5 and WHO-5 scores for each GMM group (Jacobson & Truax, 1991). Agreement between day five groups (based on clinical significance category) and day ten clinical significance categories was then calculated and reported as a percentage.

Missing Values Analysis. A missing values analysis revealed that missing cases were missing completely at random for the WHO-5; Little’s MCAR test: $\chi^2 (1378) = 1341.18, p = .76$. However, for the DI-5, the missing values analysis revealed that missing cases were not missing completely at random; Little’s MCAR test: $\chi^2 (1378) = 1538.40, p < .05$. As for at least one measure, missing cases were missing completely at random we decided to conduct expectation maximization on the data for the purposes of further analysis.

Results

Overall descriptive statistics for the DI-5 and WHO-5 scores from day one to day ten of the CBT groups are displayed in Table 1. In general, patients’ DI-5 scores
decreased over the course of the CBT group, suggesting a reduction in symptoms consistent with recovery according to the Jacobson and Truax (1991) method of clinical significance categorisation. Patients’ WHO-5 scores increased over the course of the CBT group, suggesting an increase in positive wellbeing also consistent with recovery according to clinical significance categorisation.

Table 1

*Means and standard deviations by day of treatment for the DI-5 and WHO-5*

<table>
<thead>
<tr>
<th></th>
<th>DI-5</th>
<th>WHO-5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Day 1</td>
<td>8.70</td>
<td>5.61</td>
</tr>
<tr>
<td>Day 2</td>
<td>7.25</td>
<td>5.31</td>
</tr>
<tr>
<td>Day 3</td>
<td>6.71</td>
<td>5.11</td>
</tr>
<tr>
<td>Day 4</td>
<td>6.25</td>
<td>5.19</td>
</tr>
<tr>
<td>Day 5</td>
<td>5.98</td>
<td>5.07</td>
</tr>
<tr>
<td>Day 6</td>
<td>5.93</td>
<td>5.13</td>
</tr>
<tr>
<td>Day 7</td>
<td>5.69</td>
<td>5.02</td>
</tr>
<tr>
<td>Day 8</td>
<td>5.63</td>
<td>5.10</td>
</tr>
<tr>
<td>Day 9</td>
<td>5.10</td>
<td>4.84</td>
</tr>
<tr>
<td>Day 10</td>
<td>4.53</td>
<td>4.59</td>
</tr>
</tbody>
</table>

**DI-5 Symptom Index**

**Clinical significance early change method.** Early change for the DI-5 was first assessed using the Jacobson and Truax (1991) method of defining clinically significant change. Early change was defined as the first five days of treatment after feedback had been given to the therapist and patient. Therefore, day one and day five DI-5 scores were used to calculate early change as defined by the clinical significance method. The
reliable change index and cut-off score between the functional and dysfunctional populations for the DI-5 were taken from Dyer et al. (2014). Note that 6.17 is the functional range cut-off score for the DI-5, such that any score equal to or below 6.17 indicates that the patient is within the functional range. The DI-5 clinical significance category trajectories across the first five days of treatment are presented in Figure 2. The majority of patients were recovered (64.35%) at day five according to the clinical significance early change method. The improved category made up 8.18% of the sample, while 23.31% were classified as unchanged, and 4.16% had deteriorated by day five.

![Graph showing patient trajectories of change by clinical significance category at day five for the DI-5.](image)

Figure 2. Patient trajectories of change by clinical significance category at day five for the DI-5.

**Growth mixture modelling early change method.** Early change on the DI-5 was assessed for a second time using Growth Mixture Modelling (GMM) across the first five days of treatment. Table 2 displays the model fit indices for the 2, 3, and 4 DI-
Chapter 6: Clinical Significance vs. Growth Mixture Modelling

5 log-linear solutions. The Lo-Mendell-Rubin likelihood ratio test of model fit (BLRT) was used to determine the best solution (Lo, Mendell, & Rubin, 2001).

Table 2

<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Linear</td>
<td>-22574.94</td>
<td>-20542.24</td>
<td>-19755.18</td>
</tr>
<tr>
<td>BIC</td>
<td>41157.40</td>
<td>39605.14</td>
<td>39090.07</td>
</tr>
<tr>
<td>Entropy</td>
<td>.93</td>
<td>.89</td>
<td>.88</td>
</tr>
<tr>
<td>Posterior</td>
<td>.99, .96</td>
<td>.96, .92</td>
<td>.94, .91, .91, .98, .95</td>
</tr>
<tr>
<td>Probabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR-LRT</td>
<td><em>p &lt; .01</em></td>
<td><em>p &lt; .01</em></td>
<td><em>p = .41</em></td>
</tr>
<tr>
<td>BLRT</td>
<td><em>p &lt; .01</em></td>
<td><em>p &lt; .01</em></td>
<td><em>p = .40</em></td>
</tr>
</tbody>
</table>

The results showed that the addition of a fourth class did not significantly improve model fit (three classes vs. four classes: *p = .40*). As a result, the three class model was the best solution for the present sample and was used for further analyses.

Group 1 comprised of 744 patients (50.71%) who began treatment with a relatively low and unimpaired average DI-5 score and improved over the first five days of CBT (see Figure 3). Group 1 was therefore labelled the “recovered” group. Group 2 comprised of 528 patients (35.99%) who began treatment with an impaired average DI-5 score and made reliable change over the first five days of treatment (see Figure 3). However, their mean score at day five was not within the functional range for the DI-5. Therefore, Group 2 was labelled the “improving” group. Group 3 comprised of 195 patients (13.29%) who began treatment with an impaired average DI-5 score and made no reliable change over the first five days of treatment (see Figure 3). Group 3 were
therefore labelled the "unchanged" group. No deteriorating group was identified using GMM early change method, even when four classes were considered.

Figure 3. Estimated mean DI-5 change trajectories over the first five days of CBT for a three-class growth mixture model solution.

Predicting final outcomes. Using the Jacobson and Truax (1991) method, the majority of patients left treatment recovered (75.39%), while 9.13% improved, 12.54% remained unchanged, and 2.93% deteriorated (see Figure 4). The next question to address was how well the two methods of categorising early change were at predicting final outcomes for the DI-5.
Figure 4. Trajectories of change on the DI-5 by clinical significance at final outcome.

It was found that the clinical significance early change method accurately predicted 58.7% of the unchanged patients and 22.22% of the deteriorated patients. In contrast, the GMM only accurately predicted 32.07% of unchanged patients, and failed to predict any deteriorated patients at all.

WHO-5 Wellbeing Index

Clinical significance early change method. Early change for the WHO-5 was first assessed using the Jacobson and Truax (1991) method of defining clinically significant change. Like for the DI-5, early change was defined as the first five days of treatment thus day one and day five DI-5 scores were used to calculate clinical significance defined early change. See Table 3 for day one and day five descriptive statistics for the WHO-5 by clinical significance category at day five.
Table 3

Descriptive statistics for day one and day five WHO-5 scores by clinical significance category at day five.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Day 1</th>
<th></th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Recovered</td>
<td>494</td>
<td>10.53</td>
<td>5.72</td>
<td>15.79</td>
</tr>
<tr>
<td>Improved</td>
<td>55</td>
<td>2.42</td>
<td>1.20</td>
<td>9.04</td>
</tr>
<tr>
<td>Unchanged</td>
<td>868</td>
<td>4.39</td>
<td>2.78</td>
<td>4.94</td>
</tr>
<tr>
<td>Deteriorated</td>
<td>50</td>
<td>12.90</td>
<td>3.69</td>
<td>4.12</td>
</tr>
</tbody>
</table>

The reliable change index and cut-off score between the functional and dysfunctional populations for the WHO-5 were taken from Newnham et al. (2010a).

Note that 10.80 is the functional range cut-off score for the WHO-5, such that any score equal to or above 10.80 indicates that the patient is within the functional range. The WHO-5 clinical significance category trajectories across the first five days of treatment are presented in Figure 5. The majority of patients were unchanged (59.17%) at day five according to the clinical significance method. The recovered category made up 33.67% of the sample, while 3.75% were classified as improved, and 3.4% had deteriorated by day five.
Figure 5. Patient trajectories of change by clinical significance category at day five for the WHO-5.

Growth mixture modelling early change method. Early change on the WHO-5 was assessed for a second time using the Growth Mixture Modelling (GMM) method across the first five days of treatment. Table 4 displays the model fit indices for the 2, 3, and 4 WHO-5 log-linear solutions. The Lo-Mendell-Rubin likelihood ratio test of model fit (BLRT) was again used to determine the best solution (Lo et al., 2001).
Table 4

*GMM Model Fit Indices for the WHO-5 Wellbeing Index (N = 1467)*

<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Linear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td>-22950.76</td>
<td>-20835.93</td>
<td>-20082.33</td>
</tr>
<tr>
<td>Value</td>
<td>41744.76</td>
<td>40259.44</td>
<td>39795.29</td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>.93</td>
<td>.90</td>
<td>.85</td>
</tr>
<tr>
<td>Posterior</td>
<td>.98, .97</td>
<td>.97, .97, .93</td>
<td>.92, .93, .88, .94</td>
</tr>
<tr>
<td>Probabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR-LRT</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
<td>p = .18</td>
</tr>
<tr>
<td>BLRT</td>
<td>p &lt; .01</td>
<td>p &lt; .01</td>
<td>p = .17</td>
</tr>
</tbody>
</table>

The results showed that the addition of a fourth class did not significantly improve model fit (four classes vs. three classes, p = .17). Therefore, the three class model was the best solution for the present sample and was used for further analyses.

Group 1 comprised of 209 patients (14.25%) who began treatment with a relatively high and unimpaired average WHO-5 score and improved over the first five days of CBT (see Figure 6). Group 1 was therefore labelled the “recovered high start” group. Group 2 comprised of 465 patients (31.70%) who began treatment with an impaired average WHO-5 score and made reliable change and finished treatment within the functional population range for the measure (see Figure 6). Group 2 were therefore labelled the “recovered low start” group. Group 3 comprised of 793 patients (54.06%) who began treatment with an impaired average WHO-5 score and made no reliable change over the first five days of treatment (see Figure 6). Group 3 were therefore labelled the “unchanged” group. Like for the DI-5, no deteriorating group was identified using the GMM early change method, even when four classes were considered.
Figure 6. Estimated mean WHO-5 change trajectories over the first five days of CBT for a three-class growth mixture model solution.

**Predicting final outcomes.** Using the Jacobson and Truax (1991) method, the majority of patients left treatment either recovered (47.99%) or unchanged (43.63%), while only 5.79% improved, and 2.59% deteriorated (see Figure 7). The next question to address was how well the two methods of categorising early change were at predicting final outcomes for the WHO-5.
Figure 7. Trajectories of change on the WHO-5 by clinical significance at final outcome.

It was found that the clinical significance early change method accurately predicted 86.72% of the unchanged patients and 36.84% of the deteriorated patients. Unlike for the DI-5, the GMM early change method accurately predicted almost the same amount of unchanged patients as the clinical significance early change method (81.41%). However, like for the DI-5, the GMM early change method failed to predict any deteriorated patients at all.

Discussion

Routine outcome monitoring benefits treatment by identifying potential no change and deterioration. However, the current methods of predicting negative outcomes are imperfect. Thus, the aim of the present study was to compare the clinical significance early change method and the GMM early change method at predicting negative outcomes on self-report symptom and wellbeing measures. Firstly, it was found that by the end of treatment, the majority of patients (84.52%) had achieved a
positive outcome (either recovered or improved) in terms of their symptoms, and just over half of the patients achieved a positive outcome (53.78%) with regards to their wellbeing. This supports the notion that symptoms and wellbeing are separate constructs even though they are negatively correlated, which is consistent with Newnham et al. (2010b). This result is also consistent with research that suggests wellbeing takes longer to change compared to symptoms, because a smaller proportion of patients achieved a positive outcome for wellbeing, following the intensive but brief ten-day CBT group program (Newnham et al., 2010b).

Overall, it was found that the simpler clinical significance early change method was superior to the GMM early change method at predicting negative outcomes. The GMM early change method showed that patients could be grouped into three classes over the first five days of treatment for both the DI-5 (recovered, improving, and unchanged) and WHO-5 (recovered high start, recovered low start, and unchanged). Interestingly, the GMM early change method did not detect a group of early deteriorators for either symptoms or wellbeing. This suggests that the early change trajectory of deteriorated patients did not significantly deviate from those of unchanged or recovered patients, probably due to the high heterogeneity of these large groups of patients. This finding was inconsistent with the findings of Lutz et al. (2014) who found a group of early deteriorating patients. Although the present study had a larger sample size, deterioration rates tend to be higher in the United States (Hansen et al., 2002; Lambert, 2013a), which was where Lutz and colleagues conducted their research, compared to Australia (Newnham et al., 2007; Newnham & Page, 2007). Therefore, the sample size may not have been large enough for the GMM to detect a small group of early deteriorating patients.
The clinical significance early change method was able to identify a larger number of deteriorated patients than the GMM early change method, which identified no deteriorated patients. This finding is consistent with Rubel et al. (2015) who found that the clinical significance early change method was more sensitive than the GMM early change method in identifying early positive responders to treatment. Rubel et al. (2015) found that the GMM early change method was more specific in identifying early positive responders. In other words, the patients identified by the GMM as early positive responders were more likely to be true early positive responders. In the present study, the GMM early change method did not detect any deteriorated patients and agreement between the early change methods and final unchanged outcome was higher for the clinical significance early change method. This suggests that as well as being more sensitive than the GMM early change method, the clinical significance early change method was also more specific in predicting negative outcomes in the present study. Of note however, was that the deteriorated group of patients at final outcome are generally a very small proportion of the overall sample. Further, the sample size, although large in the present study, was towards the lower end of the spectrum for the purposes of GMM. This was a limitation of the present study, so it is possible that with a larger sample size, the GMM early change method may have predicted the deteriorated patients with more specificity than the clinical significance early change method.

Upon further examination of the GMM figures for both symptoms and wellbeing, of note was that the slopes for the three groups were quite similar, whereas the intercepts differed. This suggests that the amount and type of change patients experienced depended on their post-assessment symptom and wellbeing severity. More specifically, patients who began treatment with more impaired scores achieved less
improvement over the first five days of treatment than those who began treatment with moderate scores or scores within or nearing the functional range. This is consistent with research which suggests that more severe patients tend to have inferior outcomes in CBT (Lambert, 2013a). These graphs however, would not help therapists to target specific individuals at risk for a negative outcome in real time, especially in hospital settings where many patients begin treatment within the more impaired range (Lambert, 2013a). This group of patients would simply be too large to specifically target with an intervention aimed at reducing the risk for negative outcomes (e.g., clinical support tools) (Harmon, Hawkins, Lambert, Slade, & Whipple, 2005). In addition, the GMM results were quite similar to therapists’ estimates of progress and predictions of final treatment outcomes.

As discussed in the introduction, therapists’ ability to judge their patients’ current progress is poor such that they overestimate positive and underestimate negative progress (Hannan et al., 2005; Hatfield et al., 2010; Walfish et al., 2012). The same pattern is observed when therapists are asked to predict their patients’ final outcomes (Chapman et al., 2012; Hannan et al., 2005). Specifically, therapists rarely acknowledge that some of their patients may deteriorate over the course of treatment. Interestingly, the GMM early change method, in this study, was also unable to detect patients who were deteriorating and suggested that most patients generally improved early in treatment. As a result, using the GMM early change method in real time to predict negative outcomes would not only be more time consuming for therapists, but would not be very useful to them. This is because the GMM failed to identify an “at risk” group of patients, and therapists may be thinking in a similar way to the GMM early change method; i.e., most patients improve.
It is however necessary to point out another limitation of the present study, which was the fact that one of the early change methods (the clinical significance early change method) compared in this study was also the method used to assess treatment outcome. This could have led to some bias to the disadvantage of the GMM early change method. However, the reason the clinical significance method was chosen to define final outcomes is because it is widely used in clinical research and practice (Ronk et al., 2013). Further, using clinical significance calculations at day five of treatment would be fairly simple for therapists to calculate during treatment and could potentially help therapists make treatment decisions. Although the clinical significance early change method was not perfect at predicting early deterioration, it was good at predicting early no change. One could also argue that any extra information that the therapist can use to help them make decisions in treatment would be useful and that targeting potentially the wrong clients would probably not do any harm.

The clinical significance early change method would be fairly simple to use because clinical significance calculations only require two data points, and a system has already been developed to make these calculations for the therapists (Dyer et al., 2016). This system uses an excel spreadsheet in which therapists can enter their patients’ DI-5 and WHO-5 scores. The excel spreadsheet then automatically calculates the clinical significance category using the two data points, and depicts this on a graph showing the boundaries for each clinical significance category based on the patient’s post-assessment score. Therapists could use this system at day five of treatment to calculate early change as defined by the clinical significance method which will determine whether their patient is at risk for a negative outcome (i.e., unchanged or deteriorating at day five). The therapist could then initiate some kind of in-therapy action (e.g., clinical support tools) (Harmon et al., 2005) to prevent the potential negative outcome.
In conclusion, the simpler of the two empirical methods, the clinical significance early change method, was superior to the GMM early change method at predicting negative outcomes. In light of this finding, therapists could use the system developed by Dyer et al. (2016) during treatment to calculate early change as defined by the clinical significance method. This will alert therapists to patients who have deteriorated or remained unchanged by day five, thus allowing therapists to potentially prevent those negative outcomes. Although the clinical significance early change method was superior to the GMM early change method at predicting negative outcomes, the clinical significance early change method was not perfect. Therefore, future research could endeavour to develop a superior empirical method to predict negative outcomes in order to assist therapists in preventing no change and deterioration.
Chapter 7

General Discussion
CHAPTER 7

General Discussion

Routine outcome monitoring and feedback was developed to compensate for therapists’ inability to perfectly judge patient progress and outcomes. However, no research has investigated whether the provision of feedback has served its purpose from the therapists’ perspective. In other words, does providing feedback to the therapist and patient improve therapists’ judgements of patient progress and outcomes? Thus, the present thesis aimed to examine therapists’ estimates of current progress and predictions of final outcomes with the provision of feedback to both the therapist alone, and to the therapist and patient. Our secondary aim was to compare therapists’ predictions of outcomes to empirical methods of prediction.

The study presented in Chapter 2 found that therapists overestimated positive and underestimated negative progress and outcomes overall. However, therapists correctly identified half of the deteriorated patients following feedback, and correctly predicted the proportion of patients who left treatment recovered or deteriorated. Further, the empirical method was found to be superior to therapists at predicting final outcomes. The study presented in Chapter 4 provided further benefits of feedback on therapists’ judgements of progress and outcomes. The findings also suggested that the empirical method of prediction was not necessarily superior to therapists. Finally, the study presented in Chapter 6 found that early change as defined by growth mixture modelling was not superior to the simpler clinical significance early change method at predicting outcomes. These findings and their implications are discussed in more detail below. Potential limitations of the present research, as well as suggestions for future research, are also discussed as relevant throughout. Finally, the implications of these
findings with reference to clinical practice, as well as limitations and future research directions more broadly, are discussed.

**Therapists’ Judgements with the Provision of Feedback**

One of the main findings presented in Chapter 2 suggested that when feedback was provided to the therapist, therapists correctly estimated and predicted the proportion of patients who left treatment either recovered or deteriorated. Conversely, therapists overestimated the improved and underestimated the unchanged clinical significance categories. Previous research suggests that therapists consistently overestimated positive (recovered and improved) and underestimated negative (unchanged and deteriorated) progress and outcomes (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010; Walfish et al., 2012). Therefore, the finding from Chapter 2 does show improvements in therapists’ ability to judge progress and outcomes compared to previous research. That is, therapists accurately estimated the amount of patients who would fall into one of the positive (recovered) and one of the negative (deteriorated) clinical significance categories.

Previous research examined positive outcome as one category rather than examining the recovered and improved outcomes separately (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). Therefore, in the study presented in Chapter 4, we combined the recovered and improved categories into one positive outcome category, to allow our results to be more easily comparable to previous research. Further, in Chapter 4, feedback was provided to both the therapist and patient before therapists were asked to make their estimates of progress and predictions of outcomes. Interestingly, Chapter 4 demonstrated that when therapists and patients were provided with feedback, therapists correctly estimated and predicted the proportion of patients in all categories (positive outcome, unchanged, and deteriorated) for symptoms. This
finding is inconsistent with previous research and somewhat inconsistent with the findings from Chapter 2. However, the results from the studies presented in Chapters 2 and 4 generally show improvements to therapists’ estimates of progress and predictions of outcomes, as compared to previous research. (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010).

One possible explanation for the improvements to therapists’ judgements of progress and outcomes found in Chapters 2 and 4 could be attributed to the positive effects of feedback generally (Shimokawa et al., 2010). In other words, feedback of routine outcome monitoring scores to the therapist could have meant that therapists were more aware that deterioration and no change were occurring, and that not all patients would end treatment with a positive outcome. This finding is consistent with one of the theories that attempts to explain why feedback improves outcomes in the first place; the feedback intervention theory (Kluger & DeNisi, 1996). The feedback intervention theory argues that the reason feedback improves outcomes for patients at risk for a negative outcome is because it helps therapists to identify those patients during treatment, and take action towards preventing those negative outcomes (Kluger & DeNisi, 1996; Macdonald & Mellor-Clark, 2015). Thus, in the present research, perhaps the process of therapists being provided with feedback improved their judgements of progress and outcomes. Another finding from our research also supports this theory.

Taking the results from the studies presented in Chapters 2 and 4, approximately half of the patients that therapists accurately identified as deteriorated midway through therapy, went on to achieve a positive outcome. One could therefore speculate that upon recognising deterioration, the therapists took some kind of action towards turning those patients’ outcomes into positive ones. This finding would also be consistent with
Hatfield et al. (2010) who found that when therapists were asked what they do when they notice deterioration, the most likely response was *in-therapy action*. In-therapy action involves strategies such as discussing the deterioration with the patient, increasing the frequency of sessions, and assessing motivation to change (Hatfield et al., 2010). Like in Hatfield et al. (2010), future research could ask the therapists who correctly identified negative progress what they did differently after becoming aware of no change or deterioration.

However, as mentioned above, only half of the correctly identified deteriorating patients achieved positive outcomes by the end of treatment which means the other half ended treatment with a negative outcome. In order to gain more insight into what it was that led some of the deteriorating patients to a positive outcome and the others to a negative outcome, future research could also ask the *patients* what led to their changes. Because regardless of whether the patient change was positive or negative, not all patient change can be attributed to the therapist (Lambert, 2013b), and some research suggests that therapists only account for 5% of treatment outcome (King, Orr, Poulsen, Giacomantonio, & Haden, 2017). For example, some patients may be on a negative trajectory when they commence treatment, which the therapist cannot stop. Other patients may experience a stressful life event or symptom changes due to a change in medications. Further, some patients may be prevented from taking their own lives as a result of effective therapy, even if they do not show overall progress. On the other hand, the patient may experience a pleasant event that lifts their mood, or a new medication could begin to ameliorate their symptoms (Lambert, 2013b; Lutz, Rubel, et al., 2015; Saxon et al., 2017). However, if the therapist is deemed responsible, especially for negative change, what is it that therapists should be doing to prevent negative outcomes? As outlined throughout this thesis, providing this information as feedback to
therapists and patients has the potential to improve outcomes, but usually not to the point where no patients experience negative outcomes. As such, researchers have designed programs for therapists to apply when patients are identified as not making any progress or deteriorating, and have examined its effectiveness in preventing treatment failure (Harmon et al., 2005; Harmon et al., 2007).

Whipple et al. (2003) for example, designed clinical support tools to be implemented by therapists once they were given feedback that their patient was not responding to treatment. The clinical support tools were designed to direct therapists’ attention, when dealing with non-responders, to several theoretical factors known to influence psychotherapy outcome. These were the therapeutic alliance, client motivation, and the social support network. Treatment strategies were designed for each of these factors, and the therapist was required to implement all or just one of the strategies, depending on whether it was identified as a problem by the patients’ self-report. The authors found that patients who were not on track for improvement, whose therapists received feedback and implemented the clinical support tools, improved significantly more than not on track patients whose therapists received feedback only (Whipple et al., 2003). These findings have been replicated (Harmon et al., 2005; Probst et al., 2013; Probst, Lambert, Loew, Dahlbender, & Tritt, 2015; Simon et al., 2013).

Future research could complement and build upon these studies to ensure that the strategies can be tailored to different clinical settings, and for patients presenting with a broader variety of psychiatric disorders, to reduce rates of negative outcomes across the board.

Another key finding of the present research was that therapists’ judgements of progress and outcomes in Chapter 4 were superior to therapists’ judgements in Chapter 2. Moreover in Chapter 2, therapists overestimated improved and underestimated
unchanged patients, whereas in Chapter 4, therapists correctly judged the proportion of patients for all clinical significance categories. There are a number of possible explanations for the apparent differences. Firstly, the improvements in therapists' judgements found in Chapter 4 compared to the findings in Chapter 2 could be attributed to the added benefits of feedback to both the therapist and patient. Although not consistently found (Shimokawa et al., 2010), some previous research suggests that feedback to both the therapist and patient improves outcomes over and above feedback to the therapist alone (de Jong et al., 2014; Hawkins et al., 2004). One theory to explain this is the therapeutic assessment theory which suggests that the process of providing feedback to the patient is therapeutic for the patient in some way (Finn & Tonsager, 1992). On the contrary, perhaps the collaborative process of providing feedback to both the therapist and patient makes therapists more aware of how their patient is truly progressing. The latter explanation would be more in line with the present research and the feedback intervention theory, in that accurate identification of potential negative outcomes by the therapists, leads to improved outcomes as a result of in-therapy action (Kluger & DeNisi, 1996). Of course, the therapeutic assessment theory was not assessed in the present research. Therefore, future research using more qualitative methods that involves both the therapist and patient could help to clarify whether the therapeutic assessment or feedback intervention theory is more tenable; or whether both theories ring true.

Another explanation for the improvements in therapists’ judgements between Chapters 2 and 4 could be the change in the clinical significance categorisation. As discussed in Chapter 3, previous research frequently dealt with patients beginning treatment in the functional population range, on the relevant measure, by removing them (Bauer et al., 2004; McGlinchey et al., 2002; Seggar et al., 2002). However, from a
mental health service perspective, all patients’ treatment experiences are important. Further, as the present research used two measures to assess treatment progress and outcomes, it is possible for patients to begin treatment within the functional population range on one measure but the dysfunctional range on the other. This could be due to a number of reasons. For example, one of the measures may not capture the patient’s presenting problems. Therefore, in Chapter 4, we analysed the results both including patients beginning treatment within the functional range, and with these patients removed, which did not make any difference to the results.

Once we had conducted the analyses, with the inclusion of patients beginning treatment within the functional population range, we found that one of the classifications did not make clinical sense. Specifically, if patients who began treatment within the functional population range were still within the functional range by the end of treatment, they were being classified as unchanged by the Jacobson and Truax clinical significance method (Jacobson & Truax, 1991). This seemed counterintuitive as the unchanged clinical significance category implies a negative outcome, or at least an outcome that suggests there was no measurable impact, positive or negative on the patient. Ending treatment within the functional population range however, implies a positive outcome. The reason these patients were being classified as unchanged was because they had not made a statistically significance change in a positive direction. However, it would be difficult for patients beginning treatment within the functional population range to make a statistically significant change in a positive direction due to ceiling/floor effects (depending on the measure). As a result of this reasoning, we classified these patients as a positive outcome in Chapter 4. This change in the clinical significance categorisation meant that, in Chapter 4, more patients were classified as a positive outcome for both measures compared to Chapter 2. This served the therapists
well as they typically overestimate positive progress and outcomes. However, we believe that this change in clinical significance categorisation was necessary to reflect true outcomes. Future research could survey those patient who begin and end treatment within the functional population range to determine whether they deemed their treatment a positive and helpful experience. This would ensure the validity of categorising these patients as a positive outcome, despite not having made a statistically significant change.

The change in clinical significance categorisation from Chapter 2 to Chapter 4 (i.e., classifying all patients who began and ended treatment within the functional population range as a positive outcome) also meant that the outcomes for symptoms and wellbeing were quite different. Specifically, more patients were classified as a positive outcome for the symptom measure (DI-5). Interestingly in Chapter 4, therapists’ predictions and estimates of outcomes and progress were more in line with patients’ symptoms rather than their wellbeing (WHO-5). Previous research has examined patient outcomes using a single measure that assesses both symptoms and wellbeing (e.g., the OQ-45) (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). Therefore, it was useful to examine symptoms and wellbeing separately, as the outcomes differed for these two constructs in Chapter 4. One possible reason that therapists’ judgements were poor in previous research could be because therapists were basing their judgements primarily on their patients’ symptoms, and not their wellbeing. As a result, therapists’ judgements would not necessarily match up with the outcome measure as it assessed symptoms and wellbeing in a single questionnaire, rather than separately. Previous research also supports the idea that therapists’ base their judgements of patient progress on the patient’s symptoms more so than other factors. For example, Hatfield et al. (2010) found that when assessing how their patient is progressing, therapists said that
they focus most on patient symptoms. Therapists may base their judgements more so on symptoms than other factors because symptoms are often a primary focus during treatment (Manning et al., 1994; Page & Hooke, 2011). Future research could replicate the methods of Chapter 4 and interview therapists to determine whether their judgements of progress and outcomes are based more so on their patients’ symptoms, rather than their wellbeing.

Although Chapters 2 and 4 found that feedback generally improved therapists’ judgements, therapists still struggled to identify which patients had currently made no change or deteriorated, or would leave treatment unchanged or deteriorated. Specifically, therapists were good at judging the proportion of patients who would fall into each clinical significance, but the agreement between therapists’ judgements and actual outcomes and progress was not as accurate. In other words, some of the patients who therapists predicted to have a negative outcome, did not, and some of the patients therapists predicted would have a positive outcome actually had a negative outcome. This finding was consistent with previous research because therapists failed to: a) correctly identify which patients were deteriorating or unchanged midway through therapy, and b) predict which patients would end treatment deteriorated or unchanged (Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010).

The finding that therapists generally struggle to identify negative progress and outcomes accurately has been explained by the self-assessment bias: the tendency to overestimate one’s own abilities (Walfish et al., 2012). In studies with a large number of therapists it would be easier for therapists to attribute any negative outcomes to other therapists. However, in the present research, there was a total of 13 therapists across the two studies presented in Chapters 2 and 4, and the therapists worked in pairs. Therefore, it would be more difficult to attribute negative progress and outcomes to other
therapists. The fact that the present research had a small group of therapists who
frequently worked together could be why they had a better idea of the proportion of
patients who had made negative progress or left treatment with a negative outcome.
However, it does not explain why agreement between therapists’ judgements and actual
progress and outcomes was imperfect.

Previous research has suggested that therapists may have a tendency to repress
or ignore any recognition of negative progress in patients, or they may attribute the
causes of negative change to the patient rather than to themselves (Linden, 2013). As
discussed previously, not all patient non-response to treatment can be attributed to the
therapist (Lambert, 2013b; Lutz, Rubel, et al., 2015; Saxon et al., 2017), thus perhaps
therapists hold this belief too highly. Further, the present research was conducted in a
group setting. Therefore, one could speculate that it would be easier for patients
experiencing deterioration or no change to slip through the cracks, unnoticed by the
therapist. This could be because patients whose symptoms are not changing or are
worsening do not communicate this to the therapist due to the group therapy setting
(Roback, 2000; Schneibel et al., 2017). However, these patients’ no change or
deterioration would be reflected in their self-report measures. Thus, the possibility that
patients may be less honest about how they are progressing through group therapy,
highlights the importance of therapists paying attention to routine outcome monitoring
scores so that they are aware of unchanged or deteriorating patients.

Another possibility for the non-satisfactory agreement between therapists’
estimates of progress and actual progress could be due to the use of feedback. In the
present research, an implementation check was not available through the technology
used for the feedback intervention. Previous research suggests that when an
implementation check is used, therapists who used the feedback saw greater
improvements in the outcomes of their patients compared to therapists who did not use the feedback (Bickman et al., 2016; de Jong et al., 2012). If an implementation check were used in the present research, it is possible that therapists who used the feedback could have made superior estimates of progress and predictions of outcomes compared to therapists who did not use the feedback.

Research suggests that therapists are sometimes hesitant to implement feedback for a number of reasons. Some reasons for this include the extra time that it takes to use feedback, a perception that outcome assessment is different or more important than other assessment, and the level of acceptance of the feedback by the therapist (Boswell, Kraus, Miller, & Lambert, 2015; Riemer, Rosof-Williams, & Bickman, 2005). For example, if the feedback received by the therapist differs too greatly from their own perception of the patients’ progress, therapists may be more likely to reject the feedback, favouring their own views over the feedback message. Future research assessing therapists’ judgements of progress and outcomes with the provision of feedback could ensure that an implementation check is in place and could compare the judgements of therapists who used the feedback compared to those who did not.

**Empirical Methods of Prediction**

As a result of the shortfalls associated with therapists’ ability to predict which patients would have a negative outcome, we compared therapists’ predictions to an empirical method of prediction. In Chapter 2 it was found that the empirical method was superior to therapists at predicting final outcomes overall in terms of agreement between the predictions and actual outcomes. This finding of superiority of the empirical method was consistent with previous research (Grove, 2005; Hannan et al., 2005; Meehl, 1954). However in Chapter 4, we looked again into the comparison between the empirical method’s and therapists’ predictions of outcome, but this time examining the difference
in accuracy for each clinical significance category. It was found that the empirical method underpredicted the proportion of patients who would leave treatment with a positive outcome and overpredicted the proportion of patients who would leave treatment unchanged. This was similar to the pattern found when comparing therapists’ predictions with the patients’ wellbeing results. However, therapists accurately predicted the proportion of patients in each clinical significance category when comparing their predictions to the patients’ symptom results. As such, therapists were superior to the empirical method at predicting the correct proportion of patients ending treatment in each clinical significance category. This suggests that with the provision of feedback to the therapist and patient, therapists become more aware of the amount of patients who will either achieve a positive outcome, remain unchanged, or deteriorate; and therapists can do this better than an empirical method.

When examining the agreement between predictions and actual patient outcomes however, therapists were superior at predicting which patients would achieve a positive outcome, while the empirical method was superior at predicting which patients would remain unchanged. Previous research has generally not compared therapists and empirical methods at predicting positive outcomes. Interestingly, therapists outperformed the empirical method at predicting positive outcomes. It is possible that therapists were more accurate at predicting positive outcomes because they are more likely to attend to information that is consistent with their self-assessment bias and belief that the majority of their patients will improve (Walfish et al., 2012). Future research could investigate this possibility further by measuring both accuracy of therapists’ predictions of outcomes, and therapists’ self-assessment bias, using questions similar to those addressed in the Walfish et al. (2012) study. Another explanation for therapists’ improved judgements could have been because therapists had access to both
observational data and information communicated by clients in the course of the therapy which they could use in conjunction with the DI-5 and WHO-5 scores. In other words, more information could have resulted in more reliable outcome prediction.

The skill of accurately predicting positive outcomes could be useful to therapists during treatment, as they would know which patients are not at risk for treatment failure, and can therefore continue on with treatment as usual. While the empirical method overpredicted the amount of unchanged patients, it was sensitive in predicting which patients would leave treatment unchanged. The finding that the empirical method identifies some false alarms is not necessarily a problem. Moreover, although some patients predicted to remain unchanged may achieve a positive outcome, bringing therapists’ attention to these patients probably will not do any damage. However, bringing therapists’ attention to patients who are at risk for treatment failure could prevent that negative outcome. Previous researchers have also found that empirical methods tended to overpredict negative outcomes and their argument was similar (Hannan et al., 2005). However, Hannan et al. (2005) assessed the deterioration clinical significance category only.

In the present research, both the empirical method and therapists were poor at predicting which patients would deteriorate. Specifically, the empirical method correctly predicted only one patient, while the therapists correctly predicted no patients. However, only a small proportion of patients (≈4%) deteriorated on either measure across the first two studies. Due to the overall sample size, this equated to less than 20 cases of the 297 cases in Chapter 2 and the 209 cases in Chapter 4. As such, it was more difficult for both the empirical method and therapists to predict deteriorated patients. However, a precise empirical method should be able to identify those patients who will go on to deteriorate, even if they are a small proportion of the population. Thus,
although the empirical method was able to identify most patients who would leave
treatment unchanged, it was unable to predict patients who would deteriorate. As a
result of this shortcoming, we compared the empirical method used in Chapters 2 and 4
(the clinical significance early change method) to a more sophisticated longitudinal
method in Chapter 6. This more sophisticated method was based on early change as
defined by Growth Mixture Modelling (GMM).

Contrary to what was expected, the GMM early change method was inferior to
the simpler clinical significance early change method at predicting outcomes. In fact the
GMM early change method failed to accurately predict any deteriorated patients, much
like therapists. Further, the clinical significance early change method accurately
predicted more patients than the GMM early change method did for each of the other
clinical significance categories as well.

Clinical Recommendations and Considerations

The findings from the first two studies suggest that feedback broadly improved
therapists’ judgements of progress and outcomes as compared to previous research
(Chapman et al., 2012; Hannan et al., 2005; Hatfield et al., 2010). The findings also
provided some support for the feedback intervention theory, which suggests that
feedback works because therapists take action upon recognising patient no change or
deterioration (Kluger & DeNisi, 1996). As such, therapists could be encouraged to
check in with their patients’ routine outcome monitoring scores as regularly as possible
throughout treatment, as is a requirement of evidence-based practice in psychology, so
that they are aware of how their patients are progressing. Further, the findings from all
three studies suggest that the clinical significance early change method of predicting
outcomes was superior to both therapists and a more complex longitudinal method
(GMM) at predicting negative outcomes. Therefore, therapists could be encouraged to
calculate clinical significance at different times during treatment to determine how their patient is progressing. Methods such as the OQ analyst can effectively do this for therapists if their patients are routinely completing the OQ-45 outcome measure (www.OQMeasures.com). However, a more recent tool developed by Dyer et al. (2016) which calculates clinical significance for the DI-5 and WHO-5 can be used if the therapist wishes to monitor symptoms and wellbeing separately for reasons discussed in this thesis. Consequently, therapists can make in-therapy action once they have identified patients who are not responding to treatment. What type of in-therapy action the therapist should make, is up for debate and future research. As discussed, a program such as the clinical support tools (Harmon et al., 2005; Harmon et al., 2007), could be implemented so therapists have a clear step-by-step guide of what to do once they become aware of patient non-response. However, future research is needed to confirm the effectiveness of this program among different psychological disorders and clinical settings.

The findings from Chapter 4 suggested that feedback to both the therapist and patient improved therapists’ judgements even further than when feedback was provided to the therapist alone. This highlights the importance of the collaborative nature of therapy and suggests that we should encourage therapists to be transparent with their patients by sharing routine outcome monitoring information with them. This will likely facilitate the collaborative process of therapy and encourage patients to be active agents in their treatment. One point to consider however, is that different patients may respond differently to certain types of feedback. For example, Sng (2014) found that factors such as self-esteem impact how patients respond to feedback, especially if the patient is not progressing well. Therefore, to avoid potential damaging effects of feedback to the patient, future research is needed in order to tailor feedback to the individual patient.
General Limitations and Future Research

The primary limitation of the present research was the fact that the deteriorated group of patients formed a small proportion of the overall sample. Of course from a service perspective, a small proportion of deteriorated patients is a positive as it means that very few patients left treatment worse off than when they entered. From a research perspective however, this made it difficult for any analyses involving the deteriorated group to have enough statistical power to find significant differences. Further, the deterioration group was probably not large enough for the GMM to detect it. We did manage to somewhat overcome this limitation by focusing on both unchanged patients as well as deteriorated patients. However, to examine deterioration more closely, future research could use a larger sample size so that the deterioration group has greater statistical power when making comparisons.

Another limitation of the present research was the absence of a “no feedback” control condition in the first two studies (Chapters 2 and 4). As such, we could only compare our findings with the findings from previous research which did not include feedback at all. The main reason for not including a no feedback control condition was because in the setting that this research was conducted, feedback has become a mandatory part of routine clinical practice for some years. This was due to the research showing the positive effects of feedback on patient outcomes, especially patients at risk for treatment failure (Shimokawa et al., 2010). As such, it could have been viewed as unethical to withdraw feedback from a random sample of patients for the purposes of this research. However, future research could replicate the present research in a setting where feedback has not yet been implemented to enable a no feedback comparison group.
A further limitation of the present research was the use of only two outcome measures (DI-5 and WHO-5). Although these measures have been found to be reliable and valid, they are both very brief measures (5 items). Therefore, it is possible that using additional measures for comparison could have provided different or richer results on outcomes. Another limitation for the therapists’ judgements studies (Chapter 2 and 4) was that these measures were used as both the predictor and criterion, which could have led to a bias favouring the empirical method over the therapists. Further, only self-report measures were used which could be seen as a limitation.

The nesting of cases within therapists (in Chapter 2 and 4) may have led to some biases in the results. For example Therapist A in the group may have been influenced by Therapist B, or vice versa, when completing the TPQ. Future research could address this by accounting for therapist effects and the nesting of cases within therapists in the model. Finally, access to implementation checks regarding whether the therapists had viewed the feedback graphs was not available. Therefore, future research could implement technology that allows metadata to be extracted in order to determine whether the therapist viewed the feedback or not.

The present research was conducted in a naturalistic setting. While this means that external validity was high, we cannot claim high internal validity. Further, the findings from all three studies showed that although superior to the GMM early change method and therapists, the clinical significance early change method was imperfect. Therefore, future research is needed to develop and assess better methods of identifying negative progress in order to prevent negative outcomes. Finally, the present research was the first to coalesce both the feedback literature and therapists’ judgements of progress and outcomes literature; and we found that feedback improved therapists’ judgements overall. Future research could therefore explore how feedback improves therapists’
judgements of progress and outcomes, and what can be done to make therapists’ judgements even better, to further reduce the rates of negative outcomes.

**General Conclusions**

As well as improving outcomes for patients at risk for treatment failure, feedback appears to help therapists to be more in tune with how their patients are progressing and what outcomes their patients will leave treatment with. Further, feedback of routine outcome monitoring data to the therapist will allow the therapist to calculate early change clinical significance, using the tools developed by Dyer et al. (2016). Although the clinical significance prediction method is imperfect, it will help therapists to determine whether their patient is at least at risk for no change or deterioration. As such, therapists must be encouraged to access any routine outcome monitoring feedback about their patients regularly throughout treatment. Firstly, because this will help them to assess how their patient is progressing during treatment. Secondly, if their patient has not made any changes or has deteriorated, accessing routine outcome monitoring feedback will warn the therapist that they need to take in-therapy action to prevent potential negative outcomes.
References


CLIENTs' EXPERIENCES OF THE THERAPEUTIC ALLIANCE: A
DISCOVERY-ORIENTED INVESTIGATION. Journal of Marital and Family
Therapy, 32(3), 355-368.

Handbook of psychotherapy and behavior change, 1, 217-270.

S., & Hoagwood, K. (2016). Implementing a measurement feedback system: A
tale of two sites. Administration and Policy in Mental Health and Mental Health
Services Research, 43(3), 410-425.

of routine feedback to clinicians on mental health outcomes of youths: results of
a randomized trial. Psychiatric Services.

routine outcome monitoring in clinical practice: Benefits, challenges, and

Toward a stepped care approach to treating problem drinkers: the predictive
utility of within-treatment variables and therapist prognostic ratings. Addiction,
92(11), 1479-1489.

progress monitoring on subsequent readmission to psychiatric care: A six-month


References


References


Sng, A. A. (2014). *Developing a theoretical framework for studying the effects of feedback in psychotherapy and clinical practice*. (PhD Doctoral Dissertation), University of Western Australia.


World Health Organization.