A chance for change : building an outcome monitoring feedback system for outpatient mental health care

Jong, K. de

Chapter 7

General discussion

Main objectives and conclusions

The principal aim of this thesis was to develop an outcome monitoring feedback model for Dutch outpatient mental health care in the Netherlands and to test whether providing feedback to therapists and patients can improve treatment outcomes. First, the psychometric properties of the Dutch translation of the Outcome Questionnaire (OQ-45; Lambert, et al., 2004) were tested and normative samples were collected (Chapter 2). The next step was designing a study in which feedback was provided to therapists. During the design phase of the study the power to detect an effect was attempted to be optimized, by looking for the right proportion of patients within therapists and by anticipating missing data (Chapter 3). Meanwhile, data were collected on patient progress in ten outpatient centers in three different mental health care institutions in both rural and residential areas in the Netherlands. These data were used to predict the functioning of patients at the end of treatment and the speed of recovery (rate of change). The data were analyzed using contemporary statistical techniques, that are flexible in handling missing data (Chapter 4). Finally, two feedback studies were conducted. The first study was a multicenter study in an outpatient mental health care setting and followed patients in their treatment for one year. Feedback was provided to the therapist only and in addition to patient outcomes, therapist characteristics on relevant traits related to feedback effectiveness were studied (see Chapter 5). The second study was conducted in both public outpatient centers and in private practices and contained both short and longer term therapies, up until two years. This study had three treatment conditions, a control group, a group with feedback to therapists alone and a group in which both therapists and patients received feedback about the patient's progress (see Chapter 6).

Cross-cultural validation of the OQ-45

The cross-cultural validation of the OQ-45 results showed that the American and Dutch versions of the OQ-45 are similar when it comes to reliability and validity estimates, but differences in factor structure and normative scores were found. The three-domain structure of the instrument, for which there was no strong evidence in the original version, was slightly better in the Dutch population, but still not satisfactory. Further analyses on the residual correlation matrix, which consisted of the variance that was unexplained by the three factor solution, resulted in two additional factors. The first factor consisted of four items that were in the social role domain. The unexplained variance in this domain was most likely caused by the poor performance of item 14 ('I work/study too much'), a problematic item in the American OQ-45 as well (see Mueller, Lambert, & Burlingame, 1998). The second factor, named Anxiety and Somatic Distress,

was considered a useful addition to the existing scales. Reliability and validity estimates for the ASD factor were promising. This factor might be especially interesting for use by care providers that specialize in anxiety or psychosomatic disorders.

Comparison of normative scores between the American and Dutch populations showed that the Dutch community and clinical samples scored somewhat below their American equivalents. These differences resulted in a cutoff score for the Dutch population (55) of 8 points below the American cutoff point (63). Sensitivity and specificity values were very similar to those of the original version. The reliable change indices were equal (14 points). A marked difference between the American and Dutch normative scores is that in the Dutch population gender differences were found in both the clinical and the community sample. Men had more problems in the social role domain, whereas women showed higher levels of symptom distress as well as anxiety and somatic distress.

The reliability of the subscales and the total scale was adequate in most of the samples. An exception was the internal consistency of the social role domain, which was too low in all three samples, but substantially better when the clinical and community sample were combined. Sensitivity to change was very good and the OQ-45 could effectively discriminate between functional and dysfunctional populations. The concurrent validity showed proper values for the Symptom Distress and Anxiety and Somatic Distress subscales, but less support for the Interpersonal Relations and Social Role subscales. Overall, the Dutch version of the Outcome Questionnaire had sufficient to good psychometric properties.

Power in three-level multilevel models with therapist effects

Multilevel analysis has become increasingly popular for the analysis of longitudinal data in psychotherapy research. Thus far, limited attention had been paid to power analysis in these models. Chapter 3 demonstrates the effects of intraclass correlation, level of randomization, sample size, covariates and drop-out on the power to detect an effect, using data from a routine outcome monitoring study as the basis for simulation studies. A three-level multilevel model was postulated, with therapists at the highest level (level 3), patients within therapists at the middle level (level 2) and measurements within patients at the lowest level (level 1). Results demonstrated that randomization at the patient level was more effective, in terms of power, than randomization at the therapist level. Increasing the number of patients within therapists was shown to be the best way to improve power when randomization took place at the patient level. In the case of randomization at the therapist level, including more therapists was more effective. Increasing the number of measurements per patient did not have a strong effect on power in both randomization designs. In our example, adding gender as a

covariate did not influence the power much. However, our covariate did not have a strong effect and other, more significant covariates may have different effects on power. Drop-out from the study or treatment also did not affect power substantially, although it did reduce power to some extent, especially when drop-out was concentrated at the beginning of the study. Besides power, it is necessary to have appropriate sample sizes at each level to ensure accurate estimation of parameters and standard errors. In some cases this may require larger sample sizes than are necessary for sufficient power. In addition, in order to effectively distinguish between the slope variances at the patient and therapist level, there needs to be a sufficient number of patients per therapist.

Results indicate that in three-level multilevel models larger sample sizes are required than are common in general linear model approaches. This is especially the case in naturalistic data, in which the proportion of variance explained by therapist variance in outcomes is usually larger than in randomized controlled trials (Crits-Christoph, et al., 1991). The larger the portion of variance that is explained by the therapist level (referred to as the intraclass-correlation), the larger the sample size needed. Providing feedback to therapists on their patients' progress may reduce variance in outcomes between therapists, since therapists that have more negative outcomes are provided with the opportunity to adapt their treatments based on the feedback.

Risk factors for negative outcomes

Since one of the main objectives of feedback is to prevent negative outcomes, it would be useful to know which factors are associated with negative outcomes. In Chapter 4 we aimed to predict the risk for negative treatment outcomes at the end of treatment using Classification And Regression Trees (CART) and for rate of change using multilevel modeling. A common problem in finding predictors of outcomes is that naturalistic databases are used and those usually have missing data. Both CART and multilevel analysis are flexible in handling missing data: CART for missing values on the predictor variables and multilevel models for missing values on the dependent variable. Multiple imputation was used to impute missing data in predictor variables in the multilevel analysis.

Fifty-one per cent of the patients in our sample (n = 1540) improved and had scores on the OQ-45 outside the clinical range at the end of treatment. In the CART analyses we found that patients with relatively low pre-treatment scores for symptom distress, and patients with high education have a better chance of favorable outcomes. An extended model, with more nodes (branches) to the regression tree, showed the complexity of the relation between predictors and outcome and showed how pre-treatment expectancies, social role problems and GAF scores and the working alliance at the beginning of treatment (Task subscale) interacted in different ways to predict

negative outcomes at the end of treatment. The multilevel analyses showed that initial severity, the working alliance (Task or Goal subscale) and GAF score were significant predictors for the rate of change in patients. In the complete case sample, having a mood or adjustment disorder as main diagnosis had a positive relationship with the rate of change, whereas in the imputed sample previous treatment, having comorbid Axis I disorders and having a personality disorder as main diagnosis had a negative relationship with the rate of change. The model based on the multiply imputed data was considered the most reliable model, and further analyses were computed only for this model. The CART models and multilevel models differed in their sensitivity to detect negative outcomes. The first CART model had high sensitivity, but low specificity, whereas the multilevel model had high specificity and low sensitivity. The multilevel model was good at picking up deterioration, but not at identifying the no change group. The extended CART model had the best balance between sensitivity and specificity.

Effect of feedback

The effect of feedback on outcome was investigated in two randomized clinical trials. In the first study (Chapter 5), we aimed to research the efficacy of 'simple' (no warning signals or expected recovery curves) feedback compared to no feedback. The largest effects of feedback have been found in models that have expected treatment recovery curves and warning systems for patients that are deviating too much from the expected course. However, most outcome monitoring feedback systems do not have these features and the effectiveness of those systems has been studied insufficiently (Marshall, Haywood, & Fitzpatrick, 2006). Patients (n = 413) were randomly assigned to a no-feedback control group or the feedback condition. Patients that were not progressing well in therapy, so called 'not on track' cases (NOT), were identified post-hoc based on experiencing reliable deterioration in the course of treatment and feedback was expected to be especially effective for them. Contrary to our expectations, for the full sample of therapists no additional beneficial effect of feedback was found and there was no significant interaction between feedback and patients being not on track (NOT). However, in NOT cases a positive significant effect was found when therapists indicated that they used the feedback.

In the second study (Chapter 6) we aimed to demonstrate the additional effect of feedback to therapists and patients. Patients were randomly assigned to three conditions: no feedback, feedback to the therapist alone and feedback to both patient and therapist. Feedback was provided without expected recovery curves, but therapists did get feedback messages that suggested that a patient had deteriorated or not changed. As anticipated, feedback to both therapists and patients was most

effective. Subgroup analyses were performed for short-term (less than 35 weeks of therapy) and long-term (35 weeks of therapy or longer) treatment. The benefits were strongest for NOT cases in short-term therapies. Feedback provided to the therapist alone was also effective for NOT patients in short-term therapies. In long-term therapies only feedback to therapist and patient was effective. In general, feedback influenced the rate of change, but did not significantly improve end state functioning.

Therapist effects

Characteristics of the therapists and the way in which they use feedback may play a central role in the effectiveness of feedback. After all, if therapists do not use feedback constructively, it is unlikely that outcomes will improve. Several characteristics of the therapists that might influence the effectiveness of feedback were studied (Chapter 5). Feedback is more likely to be accepted if it comes from a source that has credibility and has personal relevance to the receiver (Claiborn & Goodyear, 2005). This concept is referred to as perceived validity. Another factor that seems important in acceptance is feedback orientation (Herold & Fedor, 2003; Herold, Parsons, & Rensvold, 1996). External feedback propensity reflects the preference for externally mediated feedback as well as greater faith in such information than in what one can self-generate, whereas internal feedback propensity reflects preference for internally generated feedback as well as the tendency to reconcile differences between internal and external feedback in the direction of internally generated information (Herold & Fedor, 2003). Self-efficacy is another characteristic that influences the feedback process. It refers to a person's beliefs concerning his or her ability to successfully perform a given task or behavior (Bandura, 1977). The commitment to use the feedback in therapy might also be an important factor. Australian research showed that 44% of therapists thought outcome monitoring was a waste of time (Aoun, Pennebaker, & Janca, 2002) and two-third of the therapists was not willing to use the monitoring feedback, not even if it would lead to demonstrably better outcomes (Walter, Cleary, & Rey, 1998).

The results of the study showed that therapists variables moderated the effectiveness of feedback. Therapists with a high internal feedback propensity, who are more likely to trust their own opinion than feedback from external sources, had patients with a slower rate of change than therapists with a low internal feedback propensity, whereas therapists who were more committed to use the feedback at the beginning of the study had patients who progressed faster. Both findings occurred regardless of whether therapists actually received feedback, which suggests that therapists with an open attitude towards getting feedback reach faster progress with their patients. Strangely though, when therapists with a high commitment to use the feedback actually received feedback, this slowed down the rate of change in

their patients. There also was a positive effect of self-efficacy. Patients in the feedback condition who had therapists with higher self-efficacy progressed quicker in therapy than patients of therapist with lower self-efficacy or patients of therapists that did not receive feedback. No effect was found for external feedback propensity and perceived validity. Therapists were more likely to use the feedback if they were more committed to use the feedback at the start of the study and if they were female.

Theoretical and methodological considerations

Symptoms versus functioning

There has been discussion on whether symptom reduction is the best outcome measure to study the effect of psychosocial interventions. Gladis, Gosch, Dishuk and Crits-Christoph (1999) state that a fundamental problem with a symptom-focused approach is that it is based on a narrow, outdated notion of health and disease. Although symptom relief is a major goal of treatment efforts, there are many reasons to expand outcome assessment to include other aspects of clinical progress. The OQ-45, which was used as the outcome measure in the studies in this thesis, measures symptoms as well as social functioning in work and interpersonal relationships and is broader than symptoms alone. However, analyses were on the level of the total score and outcomes were not assessed separately on the different domains of the OQ-45. Using an outcome instrument that assesses multiple domains has its implication on the effect sizes found. Outcome measures that are tailored to specific complaints (for instance the BDI for depression) usually have the highest effect sizes (Lee, Jones, Goodman, & Heyman, 2005). Moreover, since therapy length is under pressure by insurers and is often limited to a certain number of sessions, and studies show that change on symptoms usually occurs before social functioning improves (Howard, Lueger, Maling, & Martinovich, 1993; Stulz & Lutz, 2007) it is likely that a clinically significant improvement in social functioning may not have taken place yet by the end of therapy. A recent (internal) analysis of data collected as part of routine outcome monitoring in our outpatient psychotherapy centers supported this hypothesis and showed that of the patients entering treatment with problems on interpersonal and work functioning, around 70% leaves treatment unchanged on these domains (De Jong & Mooij, unpublished).

Self-report versus other perspectives

Gold and Stricker (2011) state that: "exclusive reliance on self-observation and self-report by the patient may be an ineffective and unreliable method for assessing

psychotherapy. " (p. 1098). Patients may exaggerate or diminish their complaints on a questionnaire in a need to please or rebel against the therapist or by cause of another motivational conflict (Gold & Stricker, 2011). This effect might be more pronounced if patients receive feedback themselves. We saw this phenomenon in a study we recently performed in a inpatient long-term psychotherapy setting for patients with personality disorder. Preliminary data-analysis showed that in the first few weeks, the feedback to patients and therapists group had on average higher dysfunctioning than the feedback to therapists only and no feedback control group. This effect faded after a few weeks (De Jong, Segaar, Busschbach, & Timman, in preparation). An alternative example is that patients sometimes may score higher close to discharge, if they are anxious to end therapy, or lower if they want to end it. Another issue is that patients may not always have insight in their own thoughts and feelings. People often attempt to block out unwanted thoughts and feelings and mental processes in the mind are for a large part implicit and inaccessible (Wilson & Dunn, 2004). Although patient reports have their limitations, the patient perspective in assessing level of symptoms and functioning remains valuable, but should be evaluated in the light of its limitations.
Other perspectives may provide valuable information on the patients' functioning as well. A common perspective in outcome monitoring is that of the therapists, especially in patients that are not capable of completing self-report instruments or in patient populations that are prone to biased reports, such as addiction and forensic populations. Where patients may experience response bias, therapists may be subject to observer bias. For instance, therapists were found to have inaccurate ideas about how symptoms interrelated and this affected their judgments (e.g. Lewis, 1991). The study by Hannan et al. (2005) that was discussed in Chapter 1, demonstrated that therapists tend to overestimate treatment success and underestimate treatment failure. However, making a prediction over time may require different mental processes than assessing functioning of the patients at a certain moment in time. Recent studies on fast and frugal heuristics in decision making demonstrate that with the help of simple decision rules, clinical judgment can outperform statistical predictions (e.g. Katsikopoulos, Pachur, Machery, & Wallin, 2008). Comparisons of measurements from the therapist and patient perspective attest that correlations are usually low and that measurements from the therapist perspective shows more positive effects (Trauer, 2010). It seems that both perspectives have their advantages and disadvantage and provide a different type of information.

Alternative perspectives might be the patients' family and friends, employer or society. Kazdin suggests to include measures in the outcome evaluation that are generally accepted to be of critical importance in everyday life (e.g. days missed from work or amount of arrests) for the population under examination (Kazdin, 2003). Ideally, multiple perspectives are chosen, but collecting information from multiple

sources, especially family members or friends, can be time consuming and costly. Therefore, most outcome measurement systems tend to focus on either the therapist or the patient.

Definition of treatment success and failure

In Chapter 1 we stated that the definition of treatment success and failure is complicated and each definition has its drawbacks. Or, as one reviewer put it: "Defining negative outcomes is a tricky business, that might affect the results.". Although we tried to stay close to existing literature on the topic in our definition of negative outcomes, we did deviate from it to some extent as well. For instance, we did not consider people functioning in the normal range as negative outcomes, even if they had not changed in functioning through the course of therapy. Others have considered the full 'no change' group as negative outcomes (e.g. Lutz, et al., 2006). Whether functioning in the normal range is a good criterion for treatment success remains the question. Some people in the normal population may experience marked symptomatology, but do not seek treatment for it.

An alternative criterion for treatment success is that a patient is mentally healthy when finishing treatment. The World Health Organization (WHO) defines mental health as "A state of well-being in which the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community"(WHO, 2001). In outcome research mental health is seldom used as a criterion for treatment success. Traditionally, researchers tended to focus on patients being no longer being mentally *ill*. One reason that mental health is so seldom used as criterion for successful treatment may be that the concept of mental health is complicated to measure and there is no generally accepted theory of what mental health should encompass. Some models of mental health are available (e.g. Jacobson & Greenley, 2001; e.g.Taylor & Brown, 1988), but these have not been adopted on a larger scale.

Cultural equivalence

The cultural equivalence between the Dutch and American versions of the OQ-45 were discussed in Chapter 2. There seemed to be both differences and similarities between the instruments. The most striking difference was found in the normative samples: Dutch respondents tended to report lower scores in dysfunctioning than American respondents. The measurements reported in Chapter 2 were all prior to the start of treatment for the patient group, but the discussion of cultural equivalence can be extended to the topic of change patterns and the use of feedback. Almost all data on expected treatment recovery

curves and feedback has been collected in the United States. Little is known about the cultural equivalence of change patterns and even less about cultural differences in how therapists respond to feedback. Do patients from different countries change in the same way, or are there differences to be expected? Although Dutch respondents score lower than American respondents, outcomes seem similar in the US and the Netherlands. The care system is quite different as well, which makes it hard to disentangle the language and cultural factor on the one hand and the characteristics of the care system and the patients within it on the other hand. The same holds for therapists: the training for therapists is different in the US than in the Netherlands, which may result in different attitudes towards feedback. More research in this area might provide better insight in this issue. In addition, our own sample of patients has its limitations as well. The vast majority of our patient and therapist groups were Caucasian and not much is known on how other cultural subgroups in the Netherlands are responding to therapy.

Implications for clinical practice

Providing outcome monitoring feedback to whom?

Feedback is a complicated process and many factors may influence its effectiveness. Our results show that providing feedback on outcomes in clinical practice can be effective, but is not equally effective in all cases. Results from Chapter 6 suggest that feedback to therapists and patients both might be more effective than feedback to therapists alone. Outcome monitoring feedback may also not be equally effective in all treatment settings: the largest effects were found in short-term therapies (up to 35 weeks). A recent (yet unpublished) study in crisis care patients showed that patients that got feedback were worse off than the no feedback control group (J.J.M. Dekker, personal communication). Getting feedback that you have severe problems and that these are not improving can be demoralizing for patients. The same may be true for patients that are not expected to improve in symptoms, like patients with severe mental disorders or patients that are seeking therapy for personal growth and life issues rather than reducing symptoms. Feedback is probably most effective in improving outcomes if progress is possible but not achieved at that moment. In other patient groups it may be helpful in promoting patient-therapist communication (Carlier, et al., 2010), but may not be helpful in improving outcomes.

Besides feedback to patients and therapists, feedback on outcomes can also be provided at higher levels of the care organization, in which case it contains information on outcomes at the group level. No research has yet been performed on the effect of this type of feedback on outcomes or implementation, but in our experience it can be very stimulating for teams and organizations.

Characteristics of the feedback

Important characteristics of the feedback are the timing and frequency of measuring outcomes and providing feedback and the valence and content of the feedback. Research shows that providing feedback immediately is more effective than delayed feedback (Slade, Lambert, Harmon, Smart, & Bailey, 2008). Deciding on the frequency of measuring outcomes and providing feedback is more complicated. No research has been done on the subject, but common sense tells us that the more frequently outcome is measured, the more likely it is that we are able to detect negative progress early on and the more effective the feedback may be. However, studies (including ours) also show that feedback is most effective for patients that are not progressing well in treatment and that is only a small portion of the patients that are in therapy. Measuring all patients frequently can be time consuming and expensive. Finding a tradeoff between time invested in measuring and preventing negative outcomes must be found. Using the prediction models from Chapter 4 might be instrumental in that. Patients that are at risk for negative functioning at the end of treatment or have negative expected treatment recovery curves could be measured more intensively, whereas patients that are expected to progress well might get a lighter measurement schedule. It should be noted that we have found models with many predictor variables, which may not always be practical to use in clinical practice, since it requires that values on all predictor variables are known. This is often not the case; especially the early treatment predictors (expectancies and early working alliance) are often not available for all patients, which was one of the reasons we encountered so many missing variables in our study.

An alternative option, if one does not have access to prediction models for patients at risk, is to measure all patients frequently (every session) for the first 3-5 sessions of therapy to get an idea of the direction the progress is taking. Research shows that early change is highly predictive of outcome at the end of therapy (e.g. Haas, Hill, Lambert, & Morrell, 2002; Lambert, 2005; Lutz, Stulz, & Kock, 2009), so if patients do not improve in the beginning of therapy, that could be an indication to monitor their progress more closely. In the Netherlands, many outcome monitoring systems in outpatient settings measure outcomes not more than once every three months. Depending on how much therapy is provided in that period, that may not be frequent enough to detect negative outcomes early on.

Feedback to the therapist does not necessarily have to be provided every time outcome is measured in patients. In the study in Chapter 5 we measured outcome every session for the first five sessions of therapy, but provided feedback to the therapists at sessions 1, 3 and 5. Especially when outcome is measured on a session by session basis, providing feedback to therapists every session could be too much, especially in longer

treatments, where providing feedback to therapists may have a reduced effectiveness (see Chapter 6). One model could be to provide feedback only when patients are not progressing well and do not report actively to the therapist if the patient is on track. However, this may demotivate therapists, as they will only get negative feedback in that case. Another option is to provide feedback regularly (e.g. every five sessions) if patients are on track and more often if the patient is not progressing well. Riemer and Bickman (2011) propose a hierarchical system in which one brief feedback signal is provided to therapist (e.g. red or green), and more information is provided only if the patient is not progressing well.

Studies show that feedback is more effective if is more specific (Kluger & DeNisi, 1996). This is not surprising, after all just getting the message that the patient is not progressing well might not be very helpful. It may trigger further inquiry and promote a discussion with the patient on the lack of progress, but it would be better if more specific information were available. Lambert has implemented an instrument that assesses the patients' motivation, social support and alliance with the therapist, as well as events that happened in the patients' life that might help explain why the patient is not doing well. In addition to the assessment, suggestions are given on what interventions might be helpful to improve problems in these areas. A recent meta-analysis showed that using the clinical support tools – as this system is called– has superior outcomes compared to feedback on outcome alone (Shimokawa, Lambert, & Smart, 2010).

In this thesis we have focused on measuring outcomes, but other factors might be useful to monitor as well. Therapists often prefer adding process measures to the outcome monitoring (Riemer & Bickman, 2011). Process measures that have been used include the working alliance, expectancies and motivation. Results on how monitoring these concepts influence outcomes has been scarce and inconsistent. For instance, one study that monitored the working alliance besides outcome showed a positive overall effect on outcome (Reese, Norsworthy, & Rowlands, 2009), but no effect of monitoring the alliance was found in other studies (Crits-Christoph, et al., 2010).

What applies to feedback at the individual level, probably also applies to feedback at the organizational level. Other than individual patient feedback, team level feedback is usually on the outcome of completed cases and has the objective to learn from outcomes on an aggregated level and inform treatment policy. It is usually provided periodically (e.g. once a year). Teams are most likely more willing to accept feedback from a source they perceive as valid and if the information is presented in a clear and easy way. It can be complicated to compare different teams, especially if they differ in patient populations and in the level of implementation of outcome monitoring. It is expected that in organizations feedback on outcome will be most effective for 'not on track' teams - teams that perform below the standard - since they will be motivated

to make changes to improve outcomes. An important question is what should be considered the standard? In the Netherlands, a national benchmark for mental health care aims to provide such a standard, although the discussion on whether we are trying to compare apples and oranges will probably continue, as any comparison has its limitations. By applying statistical case mix correction or providing case mix based comparison groups, the quality of such a comparison can be improved. An interesting alternative approach has been developed by Lambert, who uses prediction models to compute the expected outcomes for a mental health care organization. In that way, the institution can be benchmarked against their own expected outcomes (Hansen, Lambert & Forman, 2002; Hansen & Lambert, 2003).

Therapist and organizational factors

The study in Chapter 5 shows that traits of the therapist may moderate the effect of feedback. Gold and Stricker (2011) indicate that therapists have their 'own needs to avoid the perception of failure' and getting feedback about cases that do not progress well in therapy might cause resistance in the therapists. This may especially be the case for therapists with low self-efficacy (see chapter 5). A recent exploratory study showed that therapists with a high external feedback propensity and who perceived the outcome monitoring feedback as more valid, had a more positive attitude towards outcome monitoring feedback (De Jong, 2012). In Chapter 5 it was demonstrated that having a positive attitude towards the feedback before implementation of outcome monitoring, was predictive of actual use of the feedback.

Getting negative feedback about one's patients can be unpleasant. This unpleasant feeling is referred to in the professional literature as cognitive dissonance and is theorized to be the driving mechanism in feedback. By experiencing cognitive dissonance people get motivated to change their behavior (for instance change their treatment strategy). However, changing behavior is not the only option. As the original experiments by Festinger demonstrated, changing one's cognitions is another possibility to reduce cognitive dissonance (Festinger, 1957). Through the process of causal attribution, therapists may attribute the reasons for lack of progress in the patient outside themselves (e.g. 'it is part of the complaints of this patient'), which may reduce the effectiveness of feedback. On the other hand, therapists that hold themselves accountable and get negative feedback on a regular basis may be at risk for burnout (Riemer & Bickman, 2011). Although traits in the therapists are not within the influence of someone who wants to implement outcome monitoring, it is important to understand the dynamics within the therapists and discuss these with them.

Beside personality traits and external causal attribution, other barriers may be present, that may cause therapists to disregard the feedback. A survey amongst

therapists showed that 'other tasks that asked for attention' (40%), not having enough time (21%) and having trouble interpreting the feedback (21%) were frequent barriers for therapists to use the feedback (De Jong, 2012). Organizational factors that may be barriers to look at the feedback are high administrative pressure or full caseloads. Another factor is accountability. Riemer and Bickman (2011) state that if therapists are not hold accountable for using the feedback, implementation may have little impact. Creating external pressure will increase the likelihood that the therapist will have attention for the feedback.

Future research

Methodological innovations

Although many prediction models that aim to predict outcome in psychosocial interventions are available, it remains complicated to predict negative outcomes. More specifically, the patients that show no (reliable) change seem most complicated to predict. Patients that deteriorate can be predicted reasonably well, but model performance for the no change group is much worse (Finch, Lambert, & Schaalje, 2001; Lutz, et al., 2006). Better prediction models are needed and increasingly, statistical techniques from other disciplines are applied in our field in order to increase prediction precision. One technique that is gaining popularity in psychotherapy research is the use of latent variable models, such as growth mixture modeling. In these models, homogeneous change patterns are identified and based on the characteristics of the change patterns group membership is predicted (e.g. Stulz, Lutz, Leach, Lucock, & Barkham, 2007). This approach uses a reverse method compared to traditional regression models and may provide new insights in which patients are likely to improve and deteriorate.

Many of the techniques used in predicting outcomes have been regression based techniques in which the relation between predictor variables and outcome is assumed to be linear. The CART analysis in Chapter 4 showed that linearity may not always be correct. Kendler (2008) states that we need to start looking at more complex interactions in explanatory models for psychiatric illness and apply models that consider predictors at different levels – inside and macro, within and outside the individual – to better understand what risk factors are relevant in psychiatry. Over the last decade, new applications of statistical techniques have been introduced in clinical psychology that allow for more complex interactions, including state-space models that allow for dynamic individual change patterns (e.g. Fisher, Newman, & Molenaar, 2011) and case-based time-series analysis, that uses bootstrapping as a benchmark for individual change (e.g. Borckardt, et al., 2008).

Most of the methods mentioned above use outcome measures at the scale level, but repeatedly completing questionnaires can place a considerable burden on patients. A promising approach is the application of computerized adaptive testing. Adaptive testing involves the administration of a questionnaire in such a way that an optimal amount of information if obtained in a minimal amount of time. Based on the responses of the patients, only relevant questions that are necessary to determine the severity of the patients complaints are administered, which typically leads to a reduction of items of 50% (Weiss & Kingsbury, 1984). An important initiative in this area is a joint initiative working on building a Patient-Reported Outcomes Measurement Information System (PROMIS; http://www.nihpromis.org). The aim of this network is to develop a large bank of items that measure patient-reported outcomes and allow efficient assessment in clinical research of a wide range of diseases, using state of the art scientific techniques. The outcome measures in PROMIS are constructed according to item-response theory, which has several advantages over classical test-theory (Embretson & Reise, 2000).

Feedback studies

Most of the studies on the effectiveness of feedback have been performed in the United States, using outpatients in university based clinics and counseling centers. Often, patients were students and staff at the university were the study was performed (e.g. Harmon, et al., 2007; Lambert, et al., 2001; Reese, et al., 2009; Slade, et al., 2008; Whipple, et al., 2003). Therapists often were faculty at the university and in that sense used to exposure to research in their clinical work. Our studies were performed in the real world and have demonstrated that implementing feedback may be more complex in those settings. More research should be done in settings outside universities and with other patient groups than outpatients. Some work has been done in inpatient settings (Berking, Orth, & Lutz, 2006; Newnham, Hooke, & Page, 2010) and with substance abuse counseling (Crits-Christoph, et al.).

Another line of research that is necessary is optimizing the feedback. In our studies, we did not use expected treatment recovery curves. So far, Lambert's group is the only one that we know of that has incorporated prediction models in the feedback. More research should be done evaluating the use of prediction models as a benchmark for patients' progress. Especially the combination of the CART and multilevel model (see Chapter 4), which is a new approach, would be interesting to use as a base for feedback. In addition, the clinical support tools should be validated – there is no research supporting the premise that feedback on the alliance, social support and motivation has a specific effect in these domains. An expansion of the clinical support tools could include information based on multidisciplinary treatment guidelines.

Our study (Chapter 5) was among the first studies to research the effect of recipient characteristics in feedback on mental health care. The results demonstrated that therapist characteristics are relevant and more research in this area is needed. Therapist characteristics that might be interesting to study include attribution style, locus of control, personality traits of the therapists and emotional stability. In addition, it would be important to get more insight in the dynamics of how feedback works for whom. Are there therapists that perform worse if they are provided with feedback? Is feedback helpful for all patients?

Another line of research that would be interesting is to combine feedback studies with basic research. For instance, one problem with feedback is that it does not seem to have a learning effect in therapists. This issue could be studied in an experimental design, and is currently being studied by the research group of Andrew Page (personal communication). Another experiment would be to try and manipulate the therapists characteristics by training specific feedback related characteristics. Finding an optimal way to present the feedback to therapists could also be tested in small experimental studies.

Finally, for the further development of feedback, it is crucial that the premises of feedback theory are tested in a clinical context, since most theories originate from social and organizational psychology. Feedback effects are considered context specific and currently the contextualized feedback intervention theory (CFIT; Riemer & Bickman, 2011) is the only available theory that focuses on clinical practice. CFIT is complex and for the largest part untested, therefore alternative theoretical models could be explored as a basis to generate new hypotheses about how feedback works in clinical practice.

Concluding thoughts

Feedback is a complicated process and there are many things we still do not know about it, especially on how it works in clinical practice. This thesis demonstrated that providing outcome monitoring feedback to therapists and patients certainly does not improve outcomes under all circumstances. The metaphor of learning archery blindfolded by Sapyta et al (2005) was used earlier to illustrate why therapists may need outcome monitoring feedback about their patients' progress. They stated that "without direct feedback on how their clients are progressing, clinicians are essentially wearing a blindfold while shooting at a target" (p. 152). This – somewhat bold – statement may be true, but the question remains whether outcome monitoring feedback provides the kind of feedback that is necessary for mastering "hitting the target" and whether it is the only source of feedback. In addition, targets may change during the course of therapy.

People have sometimes asked me if we should consider stopping with outcome monitoring, considering that effects on outcome are smaller than anticipated and the effort to implement it can be large. I do not think that we should. Although effects are small now, with new developments in research and better implementation into clinical practice effects may very well increase. Even with relatively small effects on outcome, and feedback mainly being effective for a portion of patients that are not progressing well, providing feedback on patients' progress can still be useful. As long as patients leave therapy deteriorated or unchanged and there is a chance for change, we owe it to our patients to give it our best shot.