Out of the box : moving from categories to dimensions in the phenomenology of depression and anxiety.

Hollander-Gijsman, M.E. den

CHAPTER 2

ROUTINE OUTCOME MONITORING IN THE NETHERLANDS: PRACTICAL EXPERIENCES WITH A WEB-BASED STRATEGY FOR THE ASSESSMENT OF TREATMENT OUTCOME IN CLINICAL PRACTICE

Clinical Psychology and Psychotherapy 2011; 18, 1-12.

Edwin de Beurs
Margien E. den Hollander-Gijsman
Yanda R. van Rood
Nic J.A. Van der Wee
Erik J. Giltay
Martijn S. van Noorden
Rosalind van der Lem
Esther van Fenema
Frans G. Zitman

Abstract
Routine Outcome Monitoring (ROM) is a method devised to systematically collect data on the effectiveness of treatments in everyday clinical practice. ROM involves documenting the outcome of treatments through repeated assessments. Assistants are employed who perform a baseline assessment comprising a standardized diagnostic interview, administration of rating scales, and completion of several self-report measures by the patient. At fixed time intervals assessments are repeated. Dedicated web-based software has been developed to assist in this task. ROM informs therapists and patients on the severity of the complaints at intake, and the waxing and waning of symptoms over the course of treatment. Researchers can use ROM for effectiveness research and managers can use it for benchmarking. The use of ROM for research is illustrated by presenting data on the diagnostic status of patients participating in ROM and data on treatment outcome data of a subgroup of patients (with panic disorder) in our database. The results show that implementation of ROM is feasible and, after some initial reservations, most therapists now consider ROM to be a necessary and important adjunct to the clinical treatment. In addition, ROM furthers research as the data can be used to study the phenomenology of psychiatric disorders and the outcome of treatments delivered in everyday practice.

Key Practioner Message:
- A form of tracking the progress of treatment through Routine Outcome Monitoring (ROM) is described.
- Implementation of ROM appears feasible and can be carried out in large institutions as well as smaller practices.
- Providing feedback about outcome in an appealing format is highly valued by both therapists and patients.
- ROM data enable investigation of the effectiveness of treatments in everyday clinical practice.

2.1 Introduction

Routine Outcome Monitoring (ROM) is the assessment of treatment outcome at regular intervals in order to monitor patients' progress during treatment. It involves the application in everyday clinical practice of assessment technology that was originally developed for randomized clinical trials. Several objectives may be achieved with ROM. It provides information on type and severity of psychopathology before treatment commences, which can be used to optimize allocation of patients to treatment forms. Further, ROM provides feedback to therapist and patients on progress made in treatment. Finally, ROM data can be used for research into the effectiveness of treatments in care as usual.

Already in 1988, Ellwood proposed routine and frequent assessment of patients' health and suggested to build large databases from these data (Ellwood, 1988). Although this idea was well-received in editorials (see, for recent examples, Holloway, 2002; Slade, 2002), in clinical practice routine assessment is seldom realized. In a survey among 396 psychiatrists in England, only 19.4% *"routinely or occasionally"* measured the outcome of the treatment provided (Gilbody, House, & Sheldon, 2002). Since then, some projects have been initiated in which treatment outcome is routinely assessed using different outcome measures.

In the UK two developments are worth mentioning, the Mental Health Minimum Data Set (MHMDS) of the National Health Service (http://www.ic.nhs.uk/services/mental-health/mental-health-minimum-dataset-mhmds) and the Clinical Outcomes in Routine Evaluation (CORE) system. Since 2003, mental health institutes are required by the Department of Health to provide anonymized outcome data on treatments to the MHMDS. The HoNOS is the central part of the MHMDS. The HoNOS was developed as a clinician-rated instrument for routine outcome assessment and appeared a reliable, valid and sensitive outcome measure, especially suitable for the more severe mental disorders (Wing et al., 1998). Until now, the NHS has reported only results on data quality and no outcomes on "spells of care" have been reported yet. The CORE system was designed as a quality evaluation system to evaluate therapy service delivery. Its central measure, the CORE-OM, was developed as a *"user-friendly, pantheoretical and free measure to monitor the outcome of counseling and psychotherapy"* (Barkham, Culverwell, Spindler, & Twigg, 2005). It is best suited for the less severe, more common mental disorders, such as mood and anxiety disorders. Stiles et al. (2006) report on its application in evaluating the outcome of various treatments of patients that were mostly seen in primary care. Interestingly, they found a large treatment effect (average effect size [ES]=1.36 for the pre-post difference), but little difference in outcome was found between theoretically different approaches to treatment.

In Australia a nationwide program of routine outcome measurement was implemented in 2000 (Burgess et al., 2009). Mental health services are

required to provide outcome data for a national database. In this program both clinician-rated (HoNOS) and self-report instruments (e.g., K-10+,Kessler et al., 2003) are used. To analyse the data, the Australian Mental Health Outcomes and Classification Network (AMHOCN) was established in 2003. They not only analyse and report individual and aggregated results (benchmarking) but also take steps to organize the data properly, and give trainings on how to collect and use data.

In the USA, Lambert, Hansen and Finch (2001) coined the term *"patient-focused"* research for routine assessment of the course of symptoms over time. They promote session-by-session assessments and developed a relatively short questionnaire for this purpose: the Outcome Questionnaire (OQ-45, Lambert et al., 1996). The results of the OQ-45 are discussed with the patient, allowing ample time for this in the session. The high frequency of assessments makes short-term changes in psychopathology and functioning visible, but limits the number of items that can be administered and thus the comprehensiveness of the assessment. Apart from reporting on the changes per session, the expected trajectory of scores at future assessments is estimated using the pretreatment score. If a patient's score falls outside a specified range around the projected score the therapist receives a warning of potential premature dropout and/or negative outcome should therapy continue unchanged. A similar approach is advocated by Miller and colleagues (Miller, Duncan, Sorrell, & Brown, 2005). They use an even shorter scale, the Outcome Rating Scale (ORS), comprising only four visual analog scale items to be completed every session. These four items mirror the four subscales of the OQ-45. In addition, at the end of the session the patient also rates the therapeutic alliance and the usefulness of the session (agreement on goal, methods, and overall approach of therapy) on a Session Rating Scale comprising also four visual analog scales. This score is used to assess whether discrepancies exist between what a patient wants from therapy and is receiving (Miller et al., 2005).

This paper presents a method for monitoring treatment outcome in clinical practice which has been implemented in the Netherlands. In contrast to projects described above, we employ a less frequent but more comprehensive assessment battery including both generic and disorder-specific measures, and evaluate treatment outcome from the viewpoint of the patient and an independent rater. The method of ROM is described, as are the experiences with ROM and the use of ROM data by managers and researchers. To illustrate the potential of ROM for research purposes we present the baseline characteristics (diagnoses and comorbidity) of the first cohort of patients, and present outcome data of a subgroup of patients with panic disorder.

2.2 Method

General description

In spring 2002, the mental health clinics of 'Rivierduinen' (an institute serving a region of more than 1 million people) and the Department of Psychiatry of the Leiden University Medical Center (LUMC) started collaboration for routine assessment of the diagnosis at intake, and the severity of complaints at intake. Reassessments take place every 3-4 months during treatment. ROM is restricted to patients referred for treatment of mood, anxiety, and somatoform (MAS) disorders. These patients form a relatively homogenous group with substantial mutual comorbidity (Kessler et al., 1996) and mainly receive outpatient care. So far, patients referred for treatment of other disorders, such as addiction or substance abuse or Axis II disorders, do not participate in ROM. Finally, to be included patients must be literate and have sufficient command of the Dutch language to complete the self-report instruments. For the present report, a group of patients with complete data was selected (N=3,798) and their data were analysed to illustrate the research potential of ROM- data.

Ethical considerations and privacy issues

At intake, patients are informed that ROM is part of the general policy of Rivierduinen to monitor treatment outcome, that outcomes are made available only to their therapist, and that the data will be used for research purposes, but only in anonymized form. If patients object to such use, their data are removed. A comprehensive protocol safeguards anonymity of the patients and ensures proper handling of the data. This protocol is available for patients on request. The Medical Ethical committee of the LUMC approved the regulations and agreed with this policy.

Instruments

At intake the Axis-I diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) is established using the Mini-International Neuropsychiatric Interview-plus (MINI-plus, Sheehan et al., 1998). The Dimensional Assessment of Personality Pathology (DAPP-SF) is administered to assess maladaptive personality traits (Livesley & Jackson, 2006; Van Kampen D., De Beurs E., & Andrea, 2008). Subsequently, a number of instruments are administered at intake, which are also completed at each re-assessment to allow for the evaluation of treatment outcome. Together, these instruments cover change in three areas of functioning: symptom reduction, increased wellbeing, and improvement in general life functioning (Sperry, Brill, Howard, & Grissom, 1996). They are commonly used in treatment outcome research and have good psychometric properties as evidenced by national and international publications (an overview of instruments used is available at http://www.lumc.nl/psychiatry/ROM-instruments). Outcome is assessed

by patients' self-report and by an independent assessor, and includes both generic and disorder-specific measures. Generic instruments are completed by all patients, e.g., the Brief Symptom Inventory (BSI, Derogatis, 1975) and allow the comparison of treatment outcome among all patients irrespective of their disorder. Disorder-specific measures are administered only to those patients meeting criteria for the disorder at hand, e.g., the Beck Depression Inventory Revised (BDI-II, Beck & Steer, 1987) in case of a mood disorder. The latter instruments are more sensitive to change as they assess the intensity of the symptoms which the treatment targets; they provide a more accurate picture of the clinically important improvements or progress of the individual patient (Lee, Jones, Goodman, & Heyman, 2005). The assessment outcomes are made available to the therapist, discussed with the patient, and used to support decision-making for the future course of the treatment.

Specialized staff for ROM

The LUMC and Rivierduinen employ and train psychiatric nurses and psychologists (Master's level) to carry out ROM. They are less costly than therapists, and ratings from a small specialized staff tend to be more reliable than ratings from therapists. ROM assistants administer the MINI-Plus interview, rating scales such as the Clinical Global Impression (CGI, Guy, 1976) and the Global Assessment of Functioning (GAF, Endicott, Spitzer, Fleiss, & Cohen, 1976), and write a brief report (1-2 pages) on the main findings for the therapist.

To date, 20 ROM assistants have been trained in the administration of the MINI-Plus interview and the rating scales. Initially, weekly training sessions were organized. From 2006 on, assistants who had started in 2002 and had at that time ≥ 4 years experience with the ROM instruments, assisted in training new staff. To further sharpen their diagnostic skills, ROM assistants currently still meet every month (for half a day) for instruction (by invited speakers) on the phenomenology of various disorders. In addition, they practice with rating scales to improve interrater reliability. Videotaped interviews with patients are rated and the ratings are afterwards compared and discussed to reach consensus.

Treatments

The diagnostic information from the first ROM assessment is used in conjunction with the standard clinical intake interview to select the optimal treatment for the patient. Psychiatrists and clinical psychologists provide treatment in accordance with the national multidisciplinary guidelines of the National Steering Committee describing evidenced-based treatments for mood and anxiety disorders. Treatment usually consists of medication, mainly selective serotonin reuptake inhibitors (SSRIs), cognitive behaviour therapy (CBT), or

a combination of both. Simultaneous with the start of ROM, a new stepped-care approach to treatment delivery was introduced in which the first treatment of choice is the least invasive/least intensive treatment for which efficacy has been established (e.g., a protocolled CBT or short course of pharmacotherapy). Only when this treatment does not result in sufficient symptom reduction, a more invasive or intensive treatment is offered (e.g., a combination of CBT and pharmacological treatment or, eventually, electroconvulsive therapy).

Clinically significant change

To designate a change from pretest to retest as clinically meaningful we follow the proposal of Jacobson and colleagues (Jacobson, Roberts, Berns, & McGlinchey, 1999) to combine two statistical criteria for clinical significant change. First, the change from baseline to posttest should fall outside the range of the measurement error of the instrument, i.e., a statistically reliable change should be attained (Reliable Change Index). Secondly, the posttest score should be beyond a cut-off point signifying the transgression from dysfunctional to functional, i.e., a clinically significant change. Combining these two criteria provides five possible outcomes: recovery (both criteria are met), mere improvement (only statistically reliable change), no change, deterioration (reliable change in the 'wrong' direction), and relapse (reliable change and a posttest score which falls within the dysfunctional range). In ROM the results of all instruments and subscales are provided in terms of these five possible outcomes.

Feedback to the therapist and patient

ROM provides the therapist with detailed information on the state and progress of their patients. The therapist shares and discusses these results with the patient. The report on the baseline assessment consists of a summary of the results of the diagnostic interview and a selection of the most relevant results of the instruments (Figure 2.1). The re-assessment report describes the progress made since the previous assessment, or presents a review of the course of complaints over successive assessments (Figure 2.2). To accommodate therapists and patients, care is taken to present results in a visually attractive way and to provide feedback without delay. Reports follow within one day of the (re)assessment.

Therapists use the reports to inform their patients about the results. Patients are not granted direct access to their data; it was considered important to assist and inform patients on how to interpret the results in an appropriate way. The results (such as depicted in Figs. 2.1 and 2.2) can be printed and given to the patient to take home. Apart from being used to inform patients, the ROM results are also used in staff meetings were the course of treatment of patients is discussed periodically.

Brief Symptom Inventory (BSI), Assessment 1, (12-12-2008)

Personal data:

Name:	J. Doe
Birth year:	1969
Reg. number:	0321431
Therapist:	E. de Beurs

Results:

			compared to:	
subscale	score	miss	patients	gen. population
SOM	61 #	0%	high	very high
OC	50 #	0%	average	high
I S	46 #	0%	below av.	above av.
DEP	38	0%	low	average
ANX	60 #	0%	above av.	very high
HOS	49 #	0%	average	high
PHO	65 #	0%	high	very high
PAR	48 #	0%	below av.	above av.
PSY	57 #	0%	above av.	very high
TOT	52#	0%	above av.	very high

the score is higher than the cut-off between functional/dysfunctional

In a graph:

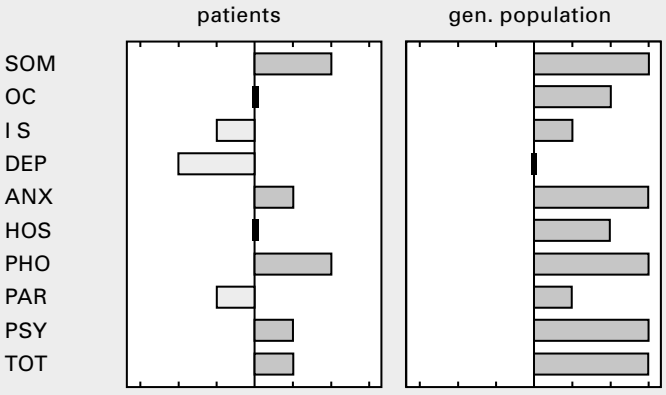

Figure 2.1 An example of the output from a single assessment with the Brief Symptom Inventory (T-scores).

Brief Symptom Inventory (BSI) - course

Personal data:

Name:	J. Doe
Birth year:	1969
Reg. number:	0321431
Therapist:	E. de Beurs

Results:

	Assessment:						
subscale	1 (12-12)		2 (4-9)		3 (8-20)		4(11-22)
SOM	61	>>	42	-	42	-	46 (below av.)*
OC	50	-	39	-	38	-	39 (low)
I S	46	-	44	-	37	-	37 (very low)
DEP	38	-	35	-	35	-	35 (very low)
ANX	60	>	44	-	41	-	39 (low)
HOS	49	-	42	-	42	-	42 (low)
PHO	65	>	48	-	50	-	42 (low)
PAR	48	-	43	-	41	-	39 (very low)
PSY	57	>>	37	-	40	-	40 (low)
TOT	52	>>	39	-	37	-	37 (low)

* normative level of the last score
- no significant change
> improved (sign. change in comparison to the previous assessment)
>> recovered (significant progress and transgression of the cut-off)

In a graph:

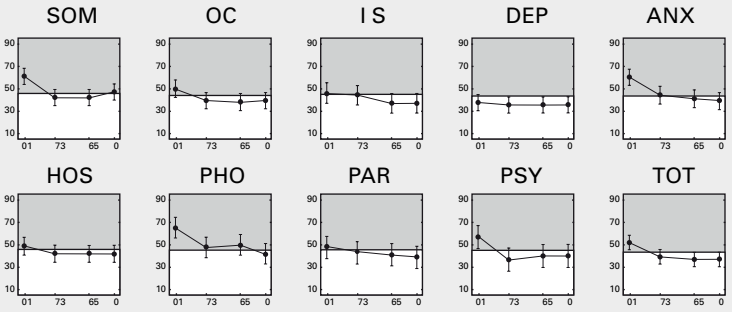

Figure 2.2 The course of complaints according to the Brief Symptom Inventory (T-scores) over time.

Software support

Dedicated web-based computer software has been developed for the administration of the MINI-Plus diagnostic interview, completion of rating scales, administration of self-report measures, and ascertainment of treatment outcome. The software presents each question of the MINI-Plus on the screen of the interviewer together with the response options. The computer software is able to deal with the sometimes complicated scoring rules in this interview and is 'intelligent': if sufficient symptoms are answered as absent to preclude a diagnosis, or sufficient symptoms are rated present to establish a positive diagnosis, no additional questions are asked, after which the module is closed and the next module is started.

The software is also used for completion of self-report questionnaires. For this purpose it has been designed as an open system: any questionnaire can be defined and administered with the software. The assessment can take place at the clinic where touchscreens can be used to accommodate computer-illiterate patients or, if they wish, patients can complete questionnaires at home via the internet.

The software computes (sub)scale scores and compares them with normative values for male/female *patients* and male/female respondents from the *general population* (Figure 2.1) and depicts the course of symptoms over time (Figure 2.2). Furthermore, the software helps in the management of data collection, e.g. allowing to list all patients who need to be invited for an upcoming 'outcome assessment' session. Finally, the software allows for the export of aggregated and anonymized data for analysis with statistical software, such as the Statistical Package for the Social Sciences (SPSS).

2.3 Results

Experiences with ROM

Interrater reliability

Multiple assessments of the same case were available from the training sessions of the ROM staff, but these cannot be used to formally establish interrater reliability, as this would lead to underestimation of the reliability of experienced ROM assistants. Interrater reliability has, however, been formerly assessed with a small subset of patients (n=44) revealing sufficient interrater reliability for the Comprehensive Psychiatric Rating Scale (Goekoop, De Beurs, & Zitman, 2007); average Cohen's κ=0.60), the GAF (average Cohen's κ=0.73) and the CGI (average Cohen's κ=0.55). These indices denote acceptable interrator reliability.

Time investment

The time needed for the first assessment is about 2 hours; 35 minutes for the MINI-Plus, 40 minutes for the rating scales and 45 minutes for the self-

report measures. A ROM re-assessment session takes (on average) 1 hour to complete. Research assistants, however, reported that for some patients there was insufficient time to include all the disorder-specific instruments which should be administered according to the MINI-Plus diagnoses.

Acceptability of the ROM procedure to patients

Patients showed good compliance with the ROM procedure. The percentage of patients with a mood, anxiety, and/or somatoform disorder that participated in ROM increased to 80% by 2009. Reasons for not participating were: the patient's command of the Dutch language was deemed insufficient to complete the questionnaires, or the assessment procedure was considered too invasive for the patient. No patients refused to partake in the ROM procedure, but approximately 5% failed repeatedly to show up at their first assessment. Comparison of the demographic data of patients who did and did not participate in ROM revealed no significant difference for gender, age, or educational level (all $p > 0.20$); however, more patients with a non-Dutch ethnic origin did not participate in ROM.

Patients were satisfied with ROM; they did not feel excessively burdened and the comprehensive assessment made them feel that their problem or disorder was taken seriously by the staff.

Attrition

Even though patients were willing to participate in ROM at intake, in our study sample, on each successive assessment the cohort was reduced in size by about 50%. Half of the patients without a re-assessment had discontinued their treatment and their last assessment can be considered a proper endpoint. However, the other half was still undergoing treatment, should have been assessed, and is considered as real loss to follow-up. Thus, no formal endpoint assessment was available for about 25% of the patients of the baseline sample due to repeated no-show. The assessment session had been rescheduled twice for these latter patients before we gave up on their outcome data. No-show for re-assessment ranges from 10 to 30% of the appointments, making it a costly problem.

Therapists' impressions of ROM

In an early phase of the implementation of ROM we conducted a survey among therapists and managers, investigating their views on the accuracy and usefulness of the data in their day to day clinical work. Therapists reported that they utilized the outcome data to motivate patients by showing them the progress made thus far, and the symptoms that still need attention in treatment. Initially, some resistance from therapists toward standardized assessment had to be overcome. Some felt that ROM was intrusive, violating the privacy of the

therapy dyad. Others felt they were better able to judge the clinical progress than could be done with standardized instruments. In practice, however, it appeared that ROM data supported or supplemented their clinical impression on how the patient fared in treatment; the data sometimes even corrected a false impression. As a result, therapeutic staff became more sensitive to treatment outcome data and eventually the majority enthusiastically accepted ROM.

During staff meetings the ROM results are presented and when they demonstrate lack of progress different courses of action are discussed. Likewise, when the ROM results indicate recovery, i.e., reliable and clinical relevant decreases in scores measuring the intensity of the main complaints, this signals that therapy might be ended, preventing the unnecessarily lingering on of treatment. Thus, ROM more than likely improved the efficiency of the treatments provided in the clinic.

Other use of ROM data

Apart from therapists and patients, researchers and managers may also use ROM data. Managers have just started to use ROM data for internal benchmarking purposes. As yet, only results on the proportion of successfully monitored treatments have been compared among the seven outpatient clinics. Outcomes on differential effectiveness of various treatment programs, locations, departments, or even therapists, have so far not been supplied. For these outcomes more complete data are needed, i.e. less loss of re-assessments.

ROM data have been used for psychometric research (Wardenaar et al., 2010; De Beurs E., Rinne, van, Verheul, & Andrea, 2009; De Beurs, Den Hollander-Gijsman, Helmich, & Zitman, 2007; Den Hollander-Gijsman, De Beurs, Van der Wee, Van Rood, & Zitman, 2010; Van Kampen D. et al., 2008), treatment outcome research, and for basic research (Van Noorden et al., 2010; Veen et al., 2009).

Examples of findings with ROM data

The results described in this paper are based on ROM data collected from January 2004 to December 2006. This dataset consists of 3,798 patients. The average age of the group was 39.6 (SD=13.3) years and 63% were women.

Diagnostic status at intake

According to the MINI-Plus, 1,618 patients (42.6%) met criteria for one MAS disorder, and 1,556 patients (41.0%) had more than one concurrent disorder (967 patients (25.5%) with two comorbid disorders, 403 patients (10.6%) with three, and 186 patients (4.9%) with four or more). Figure 2.3 presents an overview of the various (combinations of) diagnoses found in this sample when grouped in higher-order categories of MAS disorders: 1,788 patients (47.0%) met criteria for one or more *mood* disorders, 1,653 patients (43.5%) for one or more *anxiety*

disorders, 653 patients (17.2%) for one or more *somatoform* disorders, and 851 patients (22.4%) had other disorders (e.g., adjustment disorder, mixed anxiety-depression), or did not meet criteria for a DSM-IV Axis-I diagnosis (16.4%).

Outcome for panic disorder patients

To further illustrate the potential of ROM, we investigated the treatment outcome of patients with panic disorder using the Panic Disorder Severity Scale (PDSS) as an observer rater instrument for the assessment of the intensity of panic disorder symptoms (Shear et al., 1997). A total of 415 patients had a MINI-Plus diagnosis of panic disorder (with current panic attacks) and filled in the PDSS. Their average age was 35.9 (SD=10.7) years; 64% were women and 62% suffered from panic disorder with agoraphobia. On the PDSS the average score at pretest was 12.34 (SD=5.03). A second assessment after (on average) 25.8 (SD=18.7) weeks was available for 238 patients. In this subsample, the PDSS total score dropped from 12.56 (SD=5.01) to 7.04 (SD=5.84), a difference of almost 1 SD. At posttest 68% scored below the cutoff score of 7 on the PDSS indicating clinically significant change (at the pretest 18% of the patients scored below 7). Finally, 58 patients completed four assessments spanning (on average) 62 weeks of treatment. Multivariate analysis of variance for repeated measures was used to test whether the drop in score over time followed a linear pattern: $F_{\text{linear contrast}}(1, 57)=52.11$, $p < .000$, partial $\eta 2=.48$, which denotes a large effect.

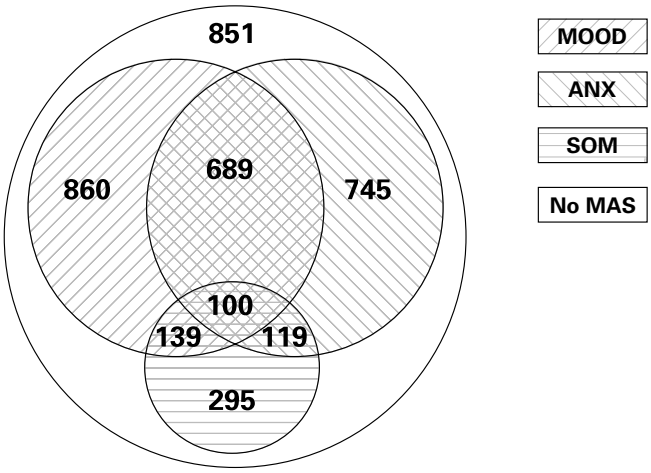

Figure 2.3 Number of patients with mood (MOOD), anxiety (ANX), somatoform (SOM) disorders (or MAS disorders), and those not meeting criteria for mood, anxiety, or somatoform disorders (No MAS) and their pattern of comorbity.

2.4 Discussion

The value of ROM

Our experiences with ROM suggest that ROM offers several benefits; however, these need to be investigated in more detail (preferably in a controlled study) as the present study provides no empirical proof of the positive value of ROM. In the literature different approaches to ROM have been developed and described. Some use a single instrument (e.g., the HoNOS, CORE-OM, or the OQ-45), others use multiple measures from different perspectives (e.g. clients, therapist or clinical raters). Some administer these instruments at pretest and posttest only, others assess periodically, and still others (Lambert, Harmon, Slade, Whipple, & Hawkins, 2005) organize assessments on a session-by-session basis.

The benefits from monitoring and providing feedback to patients have been studied in a few studies using a controlled design. Lambert and colleagues (2005) compared three conditions: informing both the therapist and the patient of the results, informing only the therapist, and informing neither the therapist nor the patient. They report positive effects on treatment outcome of providing feedback. The greatest reduction of symptoms was seen in the condition where both parties were provided with feedback. In this condition the rate of patients demonstrating clinically significant improvement was doubled. Slade et al. (2006) evaluated in their controlled study the effects of 3-monthly feedback to patients treated in a community mental health centre. These patients regularly completed (postal) questionnaires on their mental health. On the primary outcome measures the intervention group did not fare better than the 'treatment as usual' control group. However, informed patients spent significantly less days in in-patient care which made the intervention cost effective. A recent meta-analysis on the effects of providing feedback on therapist and patients concluded that the benefits are present but rather limited in effect size (Knaup, Koesters, Schoefer, Becker, & Pushner, 2009). The research of Lambert and colleagues (Lambert et al., 2005) and Miller and colleagues (Miller et al., 2005) suggests the best results are attained with patients that otherwise would have stopped the treatment prematurely. Thus, additional controlled studies are needed, in view of the substantial efforts and costs involved in obtaining outcome data in a routine manner. In addition, further research is required to determine the minimal assessment battery necessary to serve all stake holders; i.e. therapists and patients, researchers and managers.

Attrition

In our study sample, on each successive assessment the cohort was reduced in size by about 50%. High attrition rates with ROM are frequently reported. For example, in the study of Stiles et al. (2006) posttest data were available for only 33% of the patients that had provided pretest data. The high number of

patients that is lost for follow up precludes conclusions on the effectiveness of the treatments evaluated (Clark, Fairburn, & Wessely, 2008). To increase the number of patients with complete ROM data we have improved communication between ROM assistants and therapist. Now, therapists are required to inform the ROM assistant if treatment is about to conclude. The assessment session can then be scheduled prior to the final treatment session. Until more complete data are available, an intention-to-treat analysis of aggregated data might yield a more valid reflection of results obtained in everyday clinical practice than is provided by a completer analysis (Wood, White, & Thompson, 2004).

Ratings versus self-report data

ROM data are (in part) based on ratings by the ROM staff. Reliability of these ratings ranged from 0.60 to 0.74 (Cohen's κ). This denotes moderate to substantial agreement between the raters. However, these interrater reliability estimates may be somewhat inflated as they are based on the re-rating of a videotaped interview. Subjecting patients twice to two separate interviews was deemed too demanding for patients, but would have yielded more valid (and likely lower) interrater reliability estimates.

Traditionally, outcome research with mood disorders relies predominantly on rating scales whereas with anxiety disorders it is more common to use self-report scales. Using the same instruments in ROM allows for direct comparison between the treatment results attained in clinical practice (efficacy) and in randomized controlled studies. However, there are considerable additional efforts and costs involved in utilizing ratings made by independent observers. The incremental value of using raters and rating scales, compared to assessing outcome by patients' self-report with questionnaires needs to be further investigated.

Computerised assessment

We decided to use computerized administration of questionnaires for ROM as this implies that there are no missing data, instruments are scored straight away, and normed results are immediately available. A disadvantage is that some experience with computers is required, which older people in particular might not have. To solve this problem the assessment sessions are scheduled at the treatment centres were touchscreens are available and ROM staff can help the patient when this is necessary.

The software allows for completion of the self-report questionnaires at home. Although this is patient-friendly, the drawback is that we cannot be 100% certain that the patient completed the measures without the help of family members or others. The option of completing self-report questionnaires at home is, however, still open and we are currently exploring options for this more patient-friendly version of ROM.

Scientific research implications

To illustrate the research potential of ROM data, we investigated the diagnostic data of all the patients participating in ROM between January 2004 and December 2006, and the treatment outcome of panic disorder patients. The diagnostic data of the MINI-Plus reveal that in clinical practice comorbidity abounds: a large proportion of patients meet criteria for two or more diagnoses. This may in part be due to the use of a structured diagnostic interview in which the criteria of a large number of DSM disorders are methodically checked and diagnoses are not easily overlooked. In a clinical interview, the intaker might be more focused on establishing the disorder to treat, disregarding comorbid psychopathology, and may see symptoms which could qualify for a comorbid diagnosis as belonging to the primary diagnosis (See also Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). The importance of diagnostic accuracy was underlined by Jensen-Doss and Weisz (2008) who showed in a meta-analysis that less drop-outs occurred and a better outcome was attained in cases where clinicians and researchers (using a structured diagnostic instrument) agreed about the diagnosis.

The patients monitored in this study form a representative sample of the patients typically seen in clinical practice. The ROM data from this sample can be used to investigate whether these patients differ substantially from patients that participate in clinical trials. Treatment outcome of panic disorder patients was assessed with a widely-used disorder-specific outcome measure, the PDSS. After (on average) 6 months of treatment scores on this rating scale had dropped by almost 1 SD. Barlow, Gorman, Shear, and Woods (2000) in their landmark randomised controlled trial, report a response rate of 60%, defined as a score below the threshold of 7 for clinically relevant complaints. In our sample, 68% of the patients scored below 7 at post-test, comparing favourably with the results of Barlow and colleagues.

The findings on comorbidity and panic disorder treatment illustrate that data collected through ROM can be used for research. Researchers normally do not have easy access to the treatment results attained in mental health institutes. With ROM the interests of both therapists and researchers are served. The ROM structure allows for the collection of additional data, such as information on biological, social, psychological, or cognitive functioning of patients. With these data fundamental research questions can be addressed regarding differences between diagnostic subgroups, associations between the phenomenology of disorders and biological or psychological parameters, and the prognostic value of these variables for treatment outcome. The latter can be advantageous for clinical practice, potentially allowing for a better match between patient needs and treatment. For instance, the choice of medication in the treatment of anxiety or depression is largely a process of 'trial and error' to find an acceptable balance between side-effects and optimal

therapeutic effect. In the future, it might be possible to select medication based on the patient's genetic information. Currently, a biological database is being built with genetic information of the patients who participate in ROM. This will enable future research into the genetic background of the phenomenology of common mental disorders and may yield preliminary data on the interaction of genes, pharmacological agents and treatment response.

2.5 Conclusions

In summary, ROM is a method for the routine assessment of treatment outcomes in clinical practice, which simultaneously serves the interests of patients, therapists, mental healthcare managers, and researchers. Implementation of ROM has been shown to be feasible and created an efficient 'assessment culture' in a mental health institute with little academic tradition.