Chapter 1

General introduction and thesis outline

M.S. van Noorden
N.J.A. van der Wee
F.G. Zitman
E.J. Giltay

A modified version of this Chapter has been accepted for publication as: "Routine Outcome Monitoring in psychiatric clinical practice: backgrounds and overview"

Introduction

Routine Outcome Monitoring (ROM) is the systematic measurement of treatment outcomes in routine clinical practice. ROM can be used as a tool for both the patient and the clinician in monitoring treatment progress. With ROM, depending on the choice of measurement instruments, detailed information about psychiatric diagnosis, several domains of symptoms and complaints, and psychosocial functioning can be ascertained, in every phase of treatment. Furthermore, on a group level, anonymised ROM data can be used for conducting epidemiological research, as well as for purposes of benchmarking. ROM is a potentially important source of information regarding the *effectiveness* of treatment in daily –or *real-world*– practice, in addition to the available information about *efficacy* of specific interventions derived from Randomised Controlled Trials (RCTs; Ellwood, 1988; Holloway, 2002; Relman, 1988; Zimmerman et al., 2006). Despite these potential advantages, ROM has not yet been broadly implemented in psychiatry (Carlier et al., 2010; de Beurs et al., 2011; Gilbody et al., 2002; Slade, 2002a).

In the Leiden region in the Netherlands, an extensive ROM infrastructure has been developed and implemented since 2002 by a collaboration of the Regional Mental Health Provider (RMPH) Rivierduinen (RD) and the Department of Psychiatry of the Leiden University Medical Center (LUMC), respectively secondary and tertiary psychiatric specialty care settings (de Beurs, et al., 2011). This thesis focused on real-world patients with Mood, Anxiety and Somatoform (MAS) disorders and real-world outcomes in daily practice, by using ROM-data collected in RD and the LUMC. The remaining part of this introductory chapter provides background information and definitions, as well as the main aims and a thesis outline.

Psychiatric diagnosis

An important precondition for a doctor to adequately treat an ill patient, is a reliable and valid diagnosis. This core criterion is true for all areas in medicine (Goodwin et al., 1996). The study of symptoms and occurrence of diseases, and hence the classification and definition of diagnoses, are within the scope of epidemiology. Ideally, knowledge of underlying pathophysiological disturbances is used for disease classification. Usually, a clinician gathers a medical history, physical examination and often laboratory tests and/or imaging tests to obtain a diagnosis (Fauci et al., 2008). Whenever a reliable and valid diagnosis has been established, a treatment plan can be proposed, and informed consent of the

patient has to be obtained. After initiation, the effect of treatment has to be monitored. In theory, treatment effect can be measured in several domains: disease activity in terms of pathological processes or biological parameters, subjective symptoms as experienced by the patient, symptoms observed by the clinician, (psychosocial) functioning, and health-related quality of life (Fauci, et al., 2008; Smith et al., 1997).

Despite major research efforts in psychiatry during the past decades, knowledge about the pathophysiological mechanisms underlying most psychiatric disorders is still limited. This is in contrast with many somatic disorders, where large breakthroughs in understanding of pathophysiology have been accomplished. This lack of knowledge about pathophysiology of aetiology of psychiatric disorders has implications for both diagnosis and monitoring of treatment effect. Firstly, the value of laboratory tests and other biomarkers in psychiatric diagnostics in the individual patient is merely marginal (Quinones et al., 2009). The psychiatrist uses medical history taking, i.e. the patient's report of internal phenomena and the systematic mental-state examination to ascertain the symptoms and complaints of the patient. Instead of laboratory or imaging tests, rating scales that measure psychopathology can be applied. Secondly, the monitoring of treatment effect is limited to standardised rating of symptoms and psychosocial functioning, because at present, no biological parameters (i.e. biomarkers) can be used as measures of disease activity. However, as mentioned above, monitoring outcomes on a routine and standardised basis with ROM has not yet become standard practice in psychiatry. The various reasons for this lack of implementation will be discussed later in this chapter.

Until 1980, no well-defined, international accepted diagnostic criteria existed in psychiatry (Mayes et al., 2005). The need for reliable and valid diagnoses urged the American Psychiatric Association (APA) to introduce the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) in 1980 (American Psychiatric Association, 1980). This document was based on validity field trials in the United States (US) and consensus of an APA task force, and contrary to the first two editions (DSM-I and DSM-II) it comprised detailed descriptions of symptom clusters and diagnostic criteria of psychiatric disorders. Since the introduction of DSM-III, psychiatric disorders are being classified based on the presence of symptoms, providing syndromal diagnoses. This is exactly what the influential German psychiatrist Emil Kraepelin proposed almost a century earlier (Mayes & Horwitz, 2005). The introduction of the DSM-III has caused a revolution in psychiatry as it dramatically increased the possibilities of conducting epidemiological research with results that were internationally applicable. The current version of the DSM, the DSM-IV (introduced in 1994 with a text revision in 2000; DSM-IV-TR) is the result of

ongoing epidemiological research and consensus (American Psychiatric Association, 1994; American Psychiatric Association, 2000). The publication of the DSM-V is due in 2013 (www.DSM5.org). As examples of common MAS disorders, table 1.1 shows adapted DSM-IV-TR criteria for Major Depressive Episode and Social Phobia.

Table 1.1. Examples of common MAS disorders : adapted DSM-IV-TR criteria of Major Depressive Episode and Social Phobia

Major Depressive Episode	Social Phobia
A. ≥5 of the following symptoms present ≥2 weeks and represent a change from previous functioning ; at least one of the symptoms is either 1 or 2.	A. An intense fear of ≥1 social or performance situations in which the person is exposed to unfamiliar people or to possible scrutiny by others. The individual fears that he or she will act in a way (or show anxiety symptoms) that will be humiliating or embarrassing
1. depressed mood	
2. markedly diminished interest or pleasure	B. Exposure to the feared situation almost invariably provokes anxiety, which may take the form of a situationally predisposed panic attack
3. significant weight loss or weight gain, or decrease or increase in appetite	
4. insomnia or hypersomnia	
5. psychomotor agitation or retardation	C . The person recognizes that the fear is excessive or unreasonable
6. fatigue or loss of energy	
7. feelings of worthlessness or excessive or in appropriate guilt	D. The situation is avoided or else is endured with intense anxiety or distress
8. diminished ability to think or concentrate , or indecisiveness	
9. recurrent thoughts of death or suicide	
	E. The avoidance, or distress in the feared situation interferes significantly with the person's normal routine , occupational (or academic) functioning, or social activities or relationships, or there is marked distress about having the phobia
B. The symptoms do not meet criteria for a mixed episode	
C. The symptoms cause significant distress or impairment in social, occupational, or other important areas of functioning	
	F. The fear or avoidance is not due to the direct physiological effects of a substance or a general medical condition and is not better accounted for by another mental disorder
D. The symptoms are not due to the direct physiological effects of a substance (e.g. drug of abuse, medication) or a general medical condition (e.g. hypothyreoidism)	
	G. If a general medical condition or another mental disorder is present, the fear in criterion A is unrelated to it
E. The symptoms are not better accounted for by bereavement.	
	Specify if: generalized if the fears include most social situations .

Abbreviations: MAS: Mood Anxiety and Somatoform; DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision.

Measurement in psychiatry

In order to measure or classify psychiatric disorders, psychiatric symptoms preferably have to be assessed in an objective and reproducible, standardised manner (Zitman, 1990). Since most psychiatric symptoms have a large subjective component (e.g. delusive thoughts, hallucinations, disturbed mood, somatic sensations), objective measurement is a challenge. These symptoms are not easily observed or verified by an examiner (Gilbody et al., 2003). The need for objective measurement of psychiatric symptoms has resulted in the development of psychometrics: the science of psychological assessment.

A psychometric test is an instrument designed to produce a quantitative assessment of some psychological attribute(s). According to the psychometric principles, a psychometric test should be valid, reliable and free of bias (Ishak et al., 2002). Validity indicates that the test assesses the true state of the phenomenon being measured, reliability refers to the extent of reproducibility of the test and bias is a systematic error in the design of the test or study, or in data analysis. For use in ROM, a test should also be sensitive to clinically important change over time (Smith et al., 1997). Measurement in psychiatry can take place on the level of syndromal diagnosis, on the level of symptom severity, on the level of psychosocial functioning and on the level of health related quality of life. Ideally, a ROM test battery consists of measurement instruments that cover all these levels (de Beurs et al., 2011).

DSM Diagnostic measurement instruments

Syndromal classifications are potentially less fundamental than classifications that make use of clearly disturbed biological etiological processes, e.g. the detection of tumor cells in cancer or the occurrence of a pathogen in infectious diseases. Nevertheless, the introduction of the DSM-II and its successive worldwide use has greatly facilitated the development of structured diagnostic measurement instruments, necessary for psychiatric epidemiologic research. Until the 1980's, psychiatric epidemiology was hampered by methodological shortcomings, most importantly because of fuzzy definitions of diagnoses and outcomes (Tohen et al., 2000).

The 1980 DSM-III criteria were used for the development of the Diagnostic Interview Schedule (DIS), for use in the first large US community epidemiologic study on mental health: the Epidemiological Catchment Area (ECA) study, sponsored by the US National Institute of Mental Health (NIMH). This structured interview could be administered by lay interviewers because of its closed-ended questions that did not require clinical judgment (Robins et al., 1981; Tohen et al., 2000). Some years later, the Composite International Diagnostic Interview (CIDI) was developed in collaboration with the World Health Organisation (WHO; Robins et al., 1988). A modified version of the CIDI was used in the next large US community epidemiological study: the National Comorbidity Survey (NCS; Kessler et al., 1994). After the development of these structured diagnostic interviews, reliable prevalence estimates of psychiatric disorders were possible. Because of the extensive format of the CIDI, which limited use in clinical practice, Lecrubier and colleagues developed in a European-US collaboration a short validated structured diagnostic instrument: the Mini International Neuropsychiatric Interview-Plus (MINI-Plus). The MINI-Plus was validated versus the CIDI with satisfactory results (Lecrubier, 1997).

Symptomatic and functional measurement instruments

While diagnostic measurement instruments measure DSM diagnosis in a standardised manner, symptomatic and functional measurement instruments measure symptom severity and health status on a functional level. The latter two categories of instruments can be regarded as monitoring instruments, which may be applied at several time-points during treatment to evaluate progress of treatment and disease. Symptom-based scales may be generic or disorder-specific, and self-report or observer-rated.

The Hamilton Depression Rating Scale (HDRS) and Montgomery-Åsberg Depression Rating Scale (MADRS) are examples of well-known disorder-specific observer-rated rating scales that measure symptom severity in Major Depression (Hamilton, 1960; Montgomery, 1979). Of these two scales, the HDRS has been predominantly used in RCTs, but the MADRS seems superior for outpatient use (Uher et al., 2008). The Brief Symptom Inventory (BSI) is an example of a generic self-report rating scale that measures psychopathological symptom severity in several domains, e.g. anxiety, depression, hostility, and somatic complains (de Beurs et al., 2006; Derogatis et al., 1983).

Domains of functional monitoring instruments include general health status, quality of life (QoL) and social and role functioning. These instruments are ordinarily regarded as QoL scales, although a distinction between QoL and (psychosocial) functioning can be made. An example of a widely used instrument that measures general health status is the self-report Short Form-36 (SF-36; Ware, 1992).

Psychiatric epidemiology

The large-scale epidemiological community studies facilitated by the sophistication of psychiatric diagnosis and the subsequent development and validation of comprehensive diagnostic measurement instruments have provided valuable data on prevalence and incidence of psychiatric disorders in the general population. Furthermore, these studies have described in great detail clinical characteristics and correlates of most psychiatric disorders. A disadvantage of these epidemiological studies, however, is the fact that the responding subjects do not necessarily reflect treatment seeking populations, even when they meet criteria for psychiatric disorders. This may limit generalisability or external validity of these findings to the daily clinical practice in psychiatric specialty care. Nevertheless, these studies have played a major role in the development of the field of psychiatric epidemiology and the development of current psychiatry. Psychiatric epidemiology can be defined as "the study of the distribution and determinants of mental

disorders in specified populations and of the risk factors associated with their onset and course" (Tohen et al., 2000). In analogy with epidemiology in somatic medicine, psychiatric epidemiology can be subdivided in *community* psychiatric epidemiology and *clinical* psychiatric epidemiology.

In contrast with *community epidemiology*, which aims to describe disease phenomena and to estimate prevalence rates of diseases in the general population, the main goals of *clinical epidemiology* are to investigate the effects of presumed causal risk factors on the onset and course of illness in clinical patients, to evaluate the validity of diagnostic tests and to study predictors of treatment response that might be targeted in subsequent interventions (Kessler, 2007; Tohen et al., 2000). In the last three decades, descriptive community psychiatric epidemiologic research, with the ECA and NCS as examples, has prospered. On the other hand, clinical psychiatric epidemiology has remained under-developed as compared to clinical epidemiology in other fields of medicine (Kessler, 2007). This is mostly because of the fundamental problem of establishing psychiatric diagnoses (assessment of caseness) as compared with somatic diagnoses, because of the limited validity of most psychiatric diagnoses. Another reason for this difference is the fact that the treatment of psychiatric disorders is diverse –despite the availability of evidence-based guidelines– making it more difficult to conduct clinical epidemiological research of naturalistic variation in treatment response (Kessler, 2007). Finally, because many patients with psychiatric disorders do not seek treatment, representative descriptive data of psychiatric disorders in the general population are not necessarily applicable to everyday patients in clinical practice (Burger et al., 2007; Kessler, 2007).

Descriptive community epidemiological studies like the ECA and NCS have yielded valuable insights in prevalence rates and phenomenology of psychiatric disorders in the general population. Since these US community studies in the 1980's, replications have been conducted (National Comorbidity Survey-replication; NCS-R), as well as community studies in Europe. The first and second Netherlands Mental Health Survey and Incidence Study (NEMESIS 1 and 2) are examples of the latter (Bijl et al., 1998; de Graaf et al., 2011). The prevalence rates of most psychiatric disorders appear to be quite consistent over time and across continents (table 1.2).

Contrary to community epidemiological studies, clinical psychiatric epidemiological studies have the important potential of evaluating interventions in daily clinical practice. Kessler (2007) stated that "In addition to studying the aggregate magnitude of treatment effects, clinical epidemiological studies are needed to study the predictors of individual differences in treatment response". This type of work would ideally involve investigating baseline (i.e., as from the onset of treatment) predictors of course of

illness in broadly representative clinical samples". Examples of large scale, truly naturalistic studies are scarce nowadays. In order to conduct these studies, a ROM infrastructure could be used, in which outcome data of large naturalistic samples are collected.

Randomised Controlled Trials

The development of psychopathology measurement instruments several decades ago has also dramatically increased the possibilities of evaluating treatment efficacy by means of clinical trials. *Randomised Controlled Trials* (RCTs) have widely been accepted as the gold standard in evaluating treatment efficacy in medicine (Atkins et al., 2004; Kaptchuk, 2001). For a new drug to be approved by the regulatory authorities, superior efficacy compared to placebo in RCTs is needed. Indeed, the design of a typical RCT, in which two or more specific interventions with or without a placebo condition are directly compared in a double-blind way in a sample of patients with a specific disease, aims to maximise internal validity of the trial at hand. In other words, whenever an effect is found, it is most likely being explained by the intervention under study because confounding is largely eliminated through the randomisation process. This high level of internal validity can only be realised if both the disease under study is strictly defined, if the intervention is strictly defined, and if the sample is homogeneous in terms of comorbidity and other clinical characteristics. This means that often a large and strict set of inclusion and exclusion criteria are being used in RCTs. If those strict conditions are met, and the sample is large enough to detect clinically meaningful differences, in theory an RCT with maximised *internal validity* will provide the strongest possible evidence for superiority of a certain intervention (Rorty, 1977).

However, in the real world those perfect 'laboratory circumstances' do not exist. The RCT populations usually are highly selected, and suffer from limited *external validity*. Yet, most evidence-based treatments in medicine are largely based on findings from RCTs. Of course, the findings from RCTs are a major leap forward in terms of *evidence based medicine* as compared to the mere descriptive 'clinical expertise' and case studies from the pre-RCT era.

A major disadvantage of RCTs, however, is the possible lack of external validity or generalisability (Licht et al., 1997; Wells, 1999; Zimmerman et al., 2002). In a recent study of our group we found that only 20-25% of our depressive outpatients would meet general inclusion criteria for RCTs (van der Lem et al., 2010). In other words, real-world patients most likely differ from RCT-patients. Apart from evidence about treatment efficacy, RCTs

have also played a role in our knowledge about characteristics of psychiatric disorders (e.g. symptom profiles, comorbidity patterns). Typically, symptomatology of specific disorders has been analysed using only the baseline measurements in large RCT populations, and reports about these clinical characteristics are being published secondary to the main paper describing the primary outcome of the intervention (see for example Marcus et al., 2005 and Zisook et al., 2007).

Routine Outcome Monitoring

Historical perspective of Routine Outcome Monitoring

The limited generalisability of findings from RCTs and population studies to daily clinical practice and the lack of insight in processes and patient's experiences of treatment has inspired Ellwood for his 1988 Shattuck lecture in which he pleaded for "assessing routinely and frequently the health of patients using appropriate reliable and valid measurement instruments and to build large databases with these data" (Ellwood, 1988). He predicted "a new revolution in health care", and stimulated to systematically assess clinical, financial and health outcomes. Although this idea was well received in editorials (Holloway, 2002; Slade, 2002b), recent reviews have shown only a limited number of published studies of routinely assessed outcomes or ROM in psychiatric specialty care (Carlier et al., 2010). Institutions that have adopted ROM usually used a slim test battery (Burgess et al., 2009; Lambert et al., 2001). Several reasons for this lack of routine implementation of ROM in clinical practice are proposed: ROM is costly and time consuming, and requires a relatively complicated technical infrastructure. More important, no consensus exists about the optimal choice of measurement instruments so that outcomes are not easily comparable across clinics and across studies. Probably, parameters like treatment setting, patient population, and the limited availability of measurement instruments free of copyright may contribute to this lack of consensus. Furthermore, the aim of ROM may vary, as several parties have different interests, e.g. policy makers, insurance companies, patients, clinicians and researchers (Carlier et al., 2010; Norquist, 2002).

Aims and methodological issues of Routine Outcome Monitoring

In theory, ROM can provide both clinician and patient with valuable information about symptoms and treatment outcomes in daily clinical practice, and *effectiveness* of treatments in real-world treatment settings. *Evidence-based* treatments are based on efficacy trials, and may not be effective for every patient in the 'real world'. The main aim

of ROM is improvement of the quality of patient care by measuring progress and giving feedback to the patient. Secondary aims of ROM are understanding mechanisms of disease and treatment, establishing cost effectiveness and benchmarking. For understanding the relationship between patients' health status (outcomes), disease status and treatment (process of care) it is necessary to have access to detailed information about the type of treatment (Smith et al., 1997).

If observer-rated measurement scales are being used, it is important that the interviewer is well trained because clinical interpretation of symptoms or complaints is essential for reliable and reproducible results. To increase objectivity, preferably, measurement instruments are applied by an interviewer who is not directly involved in the treatment of the patient. In addition, inter-rater variability between interviewers should be minimised by recurrent training sessions in which calibration takes place.

Ideally, ROM measurement instruments should be clinically relevant, sensitive to change, minimally burdensome to the patient, to the staff, and to the institution in terms of costs of collection and data analysis (Dickey, 2002). This implies that a balanced selection of well-validated measurement instruments free of copyright is to be preferred. It is evident that different instruments may be applied in different patient groups. In the international literature, no consensus exists about the choice of measurement instruments, about the interval of measurement, and about the groups of patients or treatment settings in which ROM may be applied.

Since the data gathering in ROM is naturalistic and observational, no experimental designs can be used if outcome data are routinely assessed. Hence, instead of causal inferences, only correlations can be established on group level (Kessler, 2007). Another methodological issue when analysing ROM data is the problem of confounding and selection bias, since the treatment that a patient receives will often be determined by a number of factors that are related to outcome, such as disease severity (Gilbody et al., 2002).

The use of patient-based measures of health may itself be useful in improving treatment outcomes, because of the possibility to provide feedback of ROM assessments to both patient and clinician. This may enable clinicians to detect problem areas in treatment that would have been missed without the use of data derived from ROM (Carlier et al., 2010; Greenhalgh et al., 1999), and may increase patient's compliance to treatment protocols (Carlier et al., 2010; Hysong, 2009). A limited amount of studies have demonstrated a positive impact of ROM on monitoring treatment, and on the quality of communication between clinician and patient (Carlier et al., 2010; Knaup et al., 2009). In the meta-analysis by Carlier et al., a favourable outcome of feedback by ROM on mental health was found on the short term only.

ROM in Mood, Anxiety and Somatoform disorders

MAS disorders are highly prevalent disorders with a large disease burden (Wittchen et al., 2011). The frequent chronicity of depression contributes substantially to the global burden of disease. By the year 2020, depression is projected to reach the second ranking place of disability-adjusted life years for all ages in both sexes (Murray et al., 1997). In the second NEMESIS study, a representative survey of 6,646 adults in the Netherlands was conducted between 2007 and 2009 (de Graaf et al., 2011). Lifetime as well as 12-month prevalence rates of mood and anxiety disorders were highly comparable with NEMESIS-1, which dated from 1996 (Bijl et al., 1998), and with NCS/NCS-R data (see table 1.2). Prevalence rates of somatoform disorders have not been ascertained in those studies.

MAS disorders show a considerable amount of overlap in diagnostic criteria and probably also share genetic and environmental factors, so these disorders occur frequently as comorbid disorders. MAS disorders are often regarded as 'common mental disorders' as opposed to the less prevalent spectrum of psychotic disorders or 'severe mental disorders'.

Table 1.2. Lifetime prevalence rates of common mood and anxiety disorders across population studies [a]

	NCS [b]	NCS-R [c]	NEMESIS-1 [d]	NEMESIS-2 [e]
Any mood disorder	19.3 (0.7)	20.8 (0.6)	19.0 (0.5)	20.2 (0.6)
Major Depression	17.1 (0.7)	16.6 (0.5)	15.4 (0.4)	18.7 (0.6)
Dysthymic Disorder	6.4 (0.4)	2.5 (0.2)	6.3 (0.3)	1.3 (0.2)
Any anxiety disorder	24.9 (0.8)	28.8 (0.9)	19.3 (0.5)	19.6 (0.7)
Panic Disorder	3.5 (0.5)	4.7 (0.2)	3.8 (0.2)	3.8 (0.3)
Social Phobia	3.3 (0.7)	12.1 (0.4)	7.8 (0.3)	9.3 (0.5)
Generalised Anxiety Disorder	5.1 (0.3)	5.7 (0.3)	2.3 (0.2)	4.5 (0.3)

Abbreviations: NCS(-R): National Comorbidity Survey(-Replication); NEMESIS: Netherlands Mental Health Survey and Incidence Study; DSM: Diagnostic and Statistical Manual of Mental Disorders; (WMH-)CIDI: (World Mental Health-)Composite International Diagnostic Interview.

[a] No prevalence rates were ascertained for somatoform disorders.

[b] DSM-III-R disorders measured with the CIDI .

[c] DSM-IV disorders measured with the WMH -CIDI.

[d] DSM-III-R disorders measured with the CIDI .

[e] DSM-IV disorders measured with the WMH -CIDI.

Since ROM is time-consuming and intensive, especially when a large test battery is applied, implementation of ROM in a treatment setting of MAS disorders might be more feasible than in a treatment setting of psychotic disorders. After all, psychotic patients may not be able to complete an extensive set of instruments (Mulder et al., 2010). These considerations have resulted in the implementation of ROM in an outpatient treatment setting of MAS disorders in the Leiden Region (Box 1.1).

Aims and outline of this thesis

Aims of this thesis

This thesis focused on several aspects of ROM in outpatients with MAS disorders. The main aims were to investigate clinical aspects –symptom profiles, comorbidity, general health status and psychosocial functioning– in a large cohort of treatment-seeking real-world patients with MAS disorders, and to compare these aspects with data from RCTs and general population studies. In addition, using prospective data, we aimed to assess whether baseline characteristics measured with ROM could predict real-world –or naturalistic- treatment outcomes. The secondary aim of this thesis was to establish the feasibility of using routinely obtained outcome data for clinical epidemiological research purposes.

General description of the patient population

The implementation of ROM (Box 1.1) in the department of psychiatry of LUMC and in the outpatient departments of RD from 2002 has taken place in several phases. In 2004, most locations had incorporated ROM in routine clinical practice. As from 2004, reliable ROM data has been systematically collected for research purposes. From January 1, 2004 until December 31, 2006, a total of 3,798 patients who had been referred for treatment of a MAS disorder to one of the outpatient clinics in LUMC and RD had a baseline ROM assessment, and were included in the first ROM Baseline cohort of the Leiden Routine Outcome Monitoring Study.

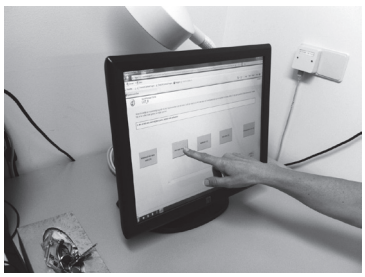
In the first three studies of this thesis, the ROM baseline cohort has been used. The average age of this group was 39.6 (standard deviation [SD] =13.3) years and 63% were women (de Beurs et al., 2011). Diagnostic status was established with the MINI-Plus. It is important to stress that clinical or primary diagnoses have not been taken into account in all ROM analyses. The major advantage of this modus operandi is that diagnostic status is entirely based on a validated measurement instrument.

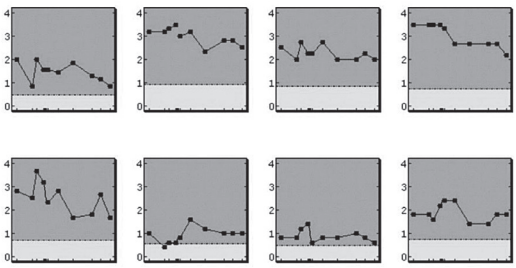
BOX 1.1. ROM in the Leiden University Medical Center & Rivierduinen

In spring 2002, the Regional Mental Health Provider (RMHP) 'Rivierduinen' (an institute serving a region with more than 1 million inhabitants) and the Department of Psychiatry of the Leiden University Medical Center (LUMC) started collaboration for routine assessment of the DSM-IV diagnosis as well as the symptom severity, well-being and health status at time of the first interview of outpatients referred to the RMPH Rivierduinen.

At the start, ROM was 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 they mainly receive outpatient care. To be eligible, patients had to have sufficient mastery of the Dutch language and had to be able to complete self-report instruments. Patients who are considered (by their clinician) to be too ill to complete questionnaires or refuse to be assessed are excluded from ROM assessment.

All patients are assessed by an independent psychiatric research nurse at the start, and during follow up at intervals of three to four months, at the beginning of a new treatment step and at the end of the treatment.





A disadvantage may be that in the case of multiple diagnoses, one diagnosis could be more prominent, and treatment may be based on this clinically more relevant diagnosis. For example, a patient with a Major Depressive Disorder (MDD) and a social phobia, may either be treated according to the social phobia treatment or the major depression guidelines. Until now, no detailed information about treatment is ascertained in the Leiden ROM. Hence, this particular patient may be analysed both as a MDD case and a social phobia case (or as patient with comorbidity), irrespective of the primary diagnosis and of treatment.

BOX 1.1. Continued

During the first session, a standardised diagnostic interview is administered and observer- and self-reported ratings are determined. At baseline 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 interviews are performed by psychiatric research nurses who have been extensively trained and supervised. The Dimensional Assessment of Personality Pathology (DAPP-SF) is administered to assess maladaptive personality traits (Livesley et al., 2006; van Kampen et al., 2008). Until now, in ROM no detailed treatment information is available.

Subsequently, a number of symptom severity rating scales are administered at baseline, 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 et al., 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 (observer-rated), and includes both generic and disorder-specific measures. Clinicians receive a report on the results of the baseline assessments as well as follow-up reporting on treatment outcome in the above mentioned domains. Results of the assessments are provided in detail by the research nurses as well as in a summarised form. The summaries facilitate clinicians to discuss the results with their patients and use them as a tool to evaluate the treatment. Results are also used, in an anonymous form, for scientific purposes.

Since ROM-data are primarily being used by clinicians and patients to monitor treatment progress, no specific informed consent is needed. The use of anonymised data for research purposes has been approved by the Medical Ethical Committee of the LUMC.

According to the MINI-Plus, of the 3,798 patients, 1,788 patients (47.0%) met criteria for one or more mood disorders, 1,653 patients (43.5%) had one or more anxiety disorders, 653 patients (17.2%) had one or more somatoform disorders, and 851 patients (22.4%) had no MAS disorder (figure 1.1). These latter patients had other diagnoses, e.g. adjustment disorders, attention deficit hyperactivity disorder, substance abuse disorders, or did not meet criteria for a Axis I DSM-IV diagnosis (de Beurs et al., 2011).

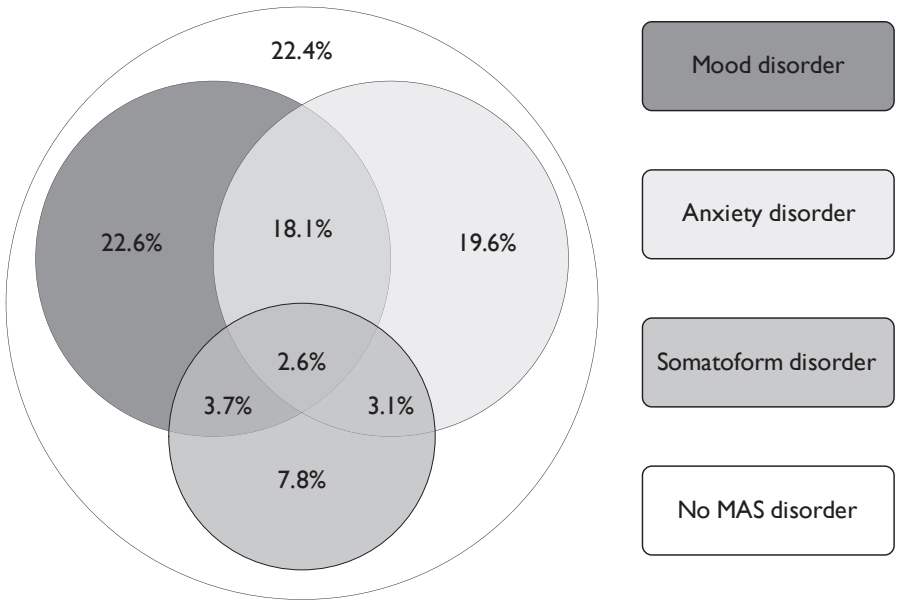


Figure 1.1. Percentages of patients with mood, anxiety and somatoform (MAS) disorders and those not meeting DSM-IV criteria for MAS disorders and comorbidity patterns in the first ROM baseline cohort 2004-2006 (n=3,798). The 'No MAS disorder' group consisted of patients with e.g. adjustment disorders, alcohol and drug abuse disorders, attention deficit and hyperactivity disorder and personality disorders, or no DSM-IV-TR diagnosis at all (e.g. social problems; de Beurs *et al.* 2011).

Thesis outline

In *Chapter 2*, we focused on gender differences in point prevalence rate, symptom profile and comorbidity patterns in 1,131 outpatients with MDD. The aim of this exploratory study was to investigate whether gender differences in a real-world sample of treatment-seeking MDD patients existed and whether these differences would deviate from differences found in clinical trial or population-based samples.

Chapter 3 describes a study on differences in clinical characteristics including demographic correlates, comorbidity, symptomatology and general health status in patients with MDD onset before adulthood versus patients with MDD onset during adulthood. Again, results were compared with findings from non-naturalistic studies.

In *Chapter 4*, we shifted the focus to lifetime deliberate self-harm and suicidal ideation (DSHI) in the total group of MAS patients. It is well-known that self-harm is a common problem in psychiatric specialty care, but no large-scale studies have described

correlates of DSHI in a real-world sample of outpatients with common mental disorders. The aim of this study was to investigate prevalence and correlates of DSHI in a large real-world outpatient mental health population.

In *Chapter 5*, we used longitudinal ROM data to identify predictors of outcome in MDD. Specifically, we aimed to investigate whether individual depressive symptoms measured with the Beck Depression Inventory-Revised (BDI-II) predict response or remission after follow-up of up to 24 months.

Chapter 6 describes a study on predictors of outcome in MAS outpatients using longitudinal ROM data. Since MAS disorders often have overlapping symptoms and mutual comorbidity is substantial, we investigated cross-diagnostic predictors of outcome measured on common generic rating scales: the BSI and the Clinical Global Impression (CGI) scale.

Finally, in *Chapter 7*, we summarised the main results of the studies. Furthermore, we discussed these results and we provided recommendations for further improvement of ROM and future studies.

References

American Psychiatric Association, 1980. Diagnostic and statistical Manual of Mental Disorders. Washington, DC.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised. Washington, DC.

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC.

Atkins, D., Best, D., Briss, P.A., Eccles, M., Falck-Ytter, Y., Flottorp, S., Guyatt, G.H., Harbour, R.T., Haugh, M.C., Henry, D., Hill, S., Jaeschke, R., Leng, G., Liberati, A., Magrini, N., Mason, J., Middleton, P., Mrukowicz, J., O'Connell, D., Oxman, A.D., Phillips, B., Schunemann, H.J., Edejer, T.T., Varonen, H., Vist, G.E., Williams, J.W., Jr., Zaza, S., 2004. Grading quality of evidence and strength of recommendations. British Medical Journal 328, 1490.

de Beurs, E., den Hollander-Gijsman, M.E., van Rood, Y.R., van der Wee, N.J., Giltay, E.J., van Noorden, M.S., van der Lem, R., van Fenema, E.M., Zitman, F.G., 2011. 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 18, 1-12.

de Beurs, E., Zitman, F.G., 2006. The Brief Symptom Inventory (BSI): Reliability and validity of a practical alternative to SCL-90. Maandblad Geestelijke Volksgezondheid 61, 120-141.

Bijl, R.V., Ravelli, A., van Zessen, G., 1998. Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Social Psychiatry and Psychiatric Epidemiology 33, 587-595.

Burger, H., Neeleman, J., 2007. A glossary on psychiatric epidemiology. Journal of Epidemiology and Community Health 61, 185-189.

Burgess, P.M., Pirkis, J.E., Slade, T.N., Johnston, A.K., Meadows, G.N., Gunn, J.M., 2009. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Australian and New Zealand Journal of Psychiatry 43, 615-623.

Carlier, I.V., Meuldijk, D., van Vliet, I.M., van Fenema, E.M., van der Wee, N.J., Zitman, F.G., 2010. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. Journal of Evaluation in Clinical Practice (in press).

Derogatis, L.R., Melisaratos, N., 1983. The Brief Symptom Inventory: an introductory report. Psychological Medicine 13, 595-605.

Dickey, B., 2002. Outcome measurement from research to clinical practice. In: Ishak, W.W., Burt, T., Sederer, L.I. (eds.), Outcome measurement in psychiatry. American Psychiatric Publishing, Washington, DC.

Ellwood, P.M., 1988. Shattuck lecture: Outcomes management. A technology of patient experience. New England Journal of Medicine 318, 1549-1556.

Fauci, A.S., Braunwald, E., Kasper, D.L., Hauser, S.L., Longo, D.L., Jameson, D.L., Loscalzo, J., 2008. The Practice of Medicine. In: Fauci, A.S., Braumwald, E., Kasper, D.L., Hauser, S.L., Longo, D.L., Jameson, D.L., Loscalzo, J. (eds.), McGraw-Hill, New York.

Gilbody, S., House, A.O., and Sheldon, T.A. Outcomes measurement in psychiatry: a review of outcomes management in psychiatric research and practice. Centre for Reviews and Dissemination report 24 . 2003. University of New York, New York.

Gilbody, S.M., House, A.O., Sheldon, T.A., 2002. Outcomes research in mental health. Systematic review. The British Journal of Psychiatry 181, 8-16.

Goodwin, D., Guze, S., 1996. Psychiatric diagnosis. Oxford University Press, New York.

de Graaf, R., Ten Have, M., van Gool, C., van Dijk, S., 2011. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. Social Psychiatry and Psychiatric Epidemiology (in press).

Greenhalgh, J., Meadows, K., 1999. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. Journal of Evaluation in Clinical Practice 5, 401-416.

Hamilton, M., 1960. A rating scale for depression. Journal of Neurology, Neurosurgery & Psychiatry 23, 56-62.

Holloway, F., 2002. Outcome measurement in mental health--welcome to the revolution. The British Journal of Psychiatry 181, 1-2.

Hysong, S.J., 2009. Meta-analysis: audit and feedback features impact effectiveness on care quality. Medical Care 47, 356-363.

Ishak, W.W., Burt, T., Sederer, L.I., 2002. Outcome Measurement in Psychiatry: a critical review. American Psychiatric Association, Washington, DC.

van Kampen, D., de Beurs, E., Andrea, H., 2008. A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): The DAPP-SF. Psychiatry Research 160, 115-128.

Kaptchuk, T.J., 2001. The double-blind, randomised, placebo-controlled trial: gold standard or golden calf? Journal of Clinical Epidemiology 54, 541-549.

Kessler, R.C., 2007. Psychiatric epidemiology: challenges and opportunities. International Review of Psychiatry 19, 509-521.

Kessler, R.C., Mcgonagle, K.A., Zhao, S.Y., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H.U., Kendler, K.S., 1994. Lifetime and 12-Month Prevalence of Dsm-III-R Psychiatric-Disorders in the United-States - Results from the National-Comorbidity-Survey. Archives of General Psychiatry 51, 8-19.

Kessler, R.C., Nelson, C.B., McGonagle, K.A., Liu, J., Swartz, M., Blazer, D.G., 1996. Comorbidity of DSM-III-R major depressive disorder in the general population: Results from the US National Comorbidity Survey. The British Journal of Psychiatry 168, 17-30.

Knaup, C., Koesters, M., Schoefer, D., Becker, T., Puschner, B., 2009. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. The British Journal of Psychiatry 195, 15-22.

Lambert, M.J., Hansen, N.B., Finch, A.E., 2001. Patient-focused research: using patient outcome data to enhance treatment effects. Journal of Consulting and Clinical Psychology 69, 159-172.

Lecrubier, Y., 1997. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 12, 224-231.

van der Lem, R., van der Wee, N.J., van,Veen, T., Zitman, F.G., 2010. The generalisability of antidepressant efficacy trials to routine psychiatric out-patient practice. Psychological Medicine 41, 1353-1363.

Licht, R.W., Gouliaev, G., Vestergaard, P., Frydenberg, M., 1997. Generalisability of results from randomised drug trials. A trial on antimanic treatment. The British Journal of Psychiatry 170, 264-267.

Livesley, W.J., Jackson, D.N., 2006. Manual for the Dimensional Assessement of Personality Problems -basic questionnaire. Sigma, Port Huron, Michigan.

Marcus, S.M., Young, E.A., Kerber, K.B., Kornstein, S., Farabaugh, A.H., Mitchell, J., Wisniewski, S.R., Balasubramani, G.K., Trivedi, M.H., Rush, A.J., 2005. Gender differences in depression: findings from the STAR*D study. Journal of Affective Disorders 87, 141-150.

Mayes, R., Horwitz, A.V., 2005. DSM-III and the revolution in the classification of mental illness. Journal of the History of the Behavioural Sciences 41, 249-267.

Montgomery, S.A., 1979. A new depression scale designed to be sensitive to change. The British journal of psychiatry 134, 382-389.

Mulder, C.L., Van der Gaag, M., Bruggeman, R., Cahn, W., Delespaul, P.A., Dries, P., Faber, G., de Haan, L., van der Heijden, F.M., Kempen, R.W., Mogendorff, E.S., Slooff, C.J., Sytema, S., Wiersma, D., Wunderink, L., van Os, J., 2010. [Routine Outcome Monitoring for patients with severe mental illness: a consensus document]. Tijdschrift voor Psychiatrie 52, 169-179.

Murray, C.J., Lopez, A.D., 1997. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet 349, 1498-1504.

Norquist, G.S., 2002. Role of outcome measurement in psychiatry. In: Ishak, W.W., Burt, T., Sederer, L.I. (eds.), Outcome measurement in psychiatry. American Psychiatric Publishing, Washington, DC.

Quinones, M.P., Kaddurah-Daouk, R., 2009. Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. Neurobiology of Disease 35, 165-176.

Relman, A.S., 1988. Assessment and accountability: the third revolution in medical care. New England Journal of Medicine 319, 1220-1222.

Robins, L.N., Helzer, J.E., Croughan, J., Ratcliff, K.S., 1981. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Archives of General Psychiatry 38, 381-389.

Robins, L.N., Wing, J., Wittchen, H.U., Helzer, J.E., Babor, T.F., Burke, J., Farmer, A., Jablenski, A., Pickens, R., Regier, D.A., 1988. The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Archives of General Psychiatry 45, 1069-1077.

Rorty, R., 1977. Philosophy and the mirror of nature. Princeton University Press, Princeton.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry 59 Suppl 20, 22-33.

Slade, M., 2002a. What outcomes to measure in routine mental health services, and how to assess them: a systematic review. Australian and New Zealand Journal of Psychiatry 36, 743-753.

Slade, M., 2002b. Routine outcome assessment in mental health services. Psychological Medicine 32, 1339-1343.

Smith, G.R., Jr., Manderscheid, R.W., Flynn, L.M., Steinwachs, D.M., 1997. Principles for assessment of patient outcomes in mental health care. Psychiatric Services 48, 1033-1036.

Sperry, L., Brill, P.L., Howard, K.I., Grissom, G.R., 1996. Treatment outcomes in psychotherapy and psychiatric interventions. Brunner/Mazel, Philadelphia, PA.

Tohen, M., Bromet, E., Murphy, J.M., Tsuang, M.T., 2000. Psychiatric epidemiology. Harvard Review of Psychiatry 8, 111-125.

Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., Mors, O., Elkin, A., Williamson, R.J., Schmael, C., Henigsberg, N., Perez, J., Mendlewicz, J., Janzing, J.G., Zobel, A., Skibinska, M., Kozel, D., Stamp, A.S., Bajs, M., Placentino, A., Barreto, M., McGuffin, P., Aitchison, K.J., 2008. Measuring depression: comparison and integration of three scales in the GENDEP study. Psychological Medicine 38, 289-300.

Ware Jr, J.E., 1992. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. Medical Care 30, 473.

Wells, K.B., 1999. Treatment Research at the Crossroads: The Scientific Interface of Clinical Trials and Effectiveness Research. The American Journal of Psychiatry 156, 5.

Wittchen, H.U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., Steinhausen, H.C., 2011. The size and burden of mental disorders and other disorders of the brain in Europe 2010. European Neuropsychopharmacology 21, 655-679.

Zimmerman, M., Mattia, J.I., Posternak, M.A., 2002. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? The American Journal of Psychiatry 159, 469-473.

Zimmerman, M., Ruggero, C.J., Chelminski, I., Young, D., Posternak, M.A., Friedman, M., Boerescu, D., Attiullah, N., 2006. Developing brief scales for use in clinical practice: the reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. Journal of Clinical Psychiatry 67, 1536-1541.

Zisook, S., Lesser, I., Stewart, J.W., Wisniewski, S.R., Balasubramani, G.K., Fava, M., Gilmer, W.S., Dresselhaus, T.R., Thase, M.E., Nierenberg, A.A., Trivedi, M.H., Rush, A.J., 2007. Effect of age at onset on the course of major depressive disorder. The American Journal of Psychiatry 164, 1539-1546.

Zitman, F.G., 1990. Standaardisering van psychiatrische diagnostiek. In: Abraham, R.E., Giel, R., Rooijmans, H.G.M., Zitman, F.G. (eds.), Diagnostiek in de psychiatrie, mogelijheden en grenzen. Boerhaave Commissie, Leiden.