On real-world patients and real-world outcomes : the Leiden Routine Outcome Monitoring Study in patients with mood, anxiety and somatoform disorders

Noorden, M.S. van

Chapter 7

Summary, discussion and future perspectives

Summary

This thesis focused on *real-world* patients and *real-world* outcomes by using Routine Outcome Monitoring (ROM) in patients with Mood, Anxiety, and Somatoform (MAS) disorders. The primary aims of the series of studies were to investigate correlates of disease characteristics in a large cohort of treatment-seeking patients with MAS disorders, and to discuss these findings in relation to existing data derived from RCTs and general population studies. In addition, using prospective data, we investigated whether baseline characteristics measured with ROM could predict outcomes in real-world or naturalistic treatment settings. The secondary aim was to assess the feasibility of conducting large-scale clinical epidemiological research with ROM data that had been collected in everyday clinical practice, thereby representing the phenotype of 'real' patients. In the remainder of this chapter we will summarise the major findings of the presented studies, discuss these findings and ROM in a broader perspective, and conclude with thoughts and recommendations about future studies.

Summary of major findings

The first study of this thesis (Chapter 2) concentrated on gender differences in patients with Major Depressive Disorder (MDD). Although gender differences in course and symptomatology of depression have frequently been studied, a closer look at the literature reveals that almost all studies have been conducted in population samples or clinical trial samples. In a recent study of our group we have shown that only 20-25% of our MDD patients would qualify for inclusion in a regular clinical trial (van der Lem et al., 2010), so we argued that patients from clinical trials do not adequately represent treatment-seeking everyday or real-world patients. In other words, external validity is limited. The aim of this study therefore was to investigate gender differences in clinical characteristics of a naturalistic sample of MDD patients from our ROM database. We used ROM baseline data of 1,131 MDD patients, and analysed symptom severity, symptom profiles, comorbidity and general health status. The most remarkable findings were that women reported a more severe 'subjective' phenotype on self-rated scales, with more symptoms and more severe symptoms of depression and a lower general health status than men. On more objective observer-rated scales, however, no significant differences in symptom severity were found between women and men. These findings suggest that the burden of MDD is higher in women than in men, but only subjectively. The pattern of subjective symptomatology

was also different in women compared to men, with women reporting more loss of energy, hypersomnia, feelings of worthlessness, loss of appetite, fatigue, loss of interest in sex, feelings of guilt and crying. The majority of these findings were largely consistent with previous studies in non-naturalistic patient samples. In Chapter 3, we extended the cross-sectional analyses on ROM baseline assessments in patients with MDD. In this study we compared clinical characteristics of patients with a pre-adult onset versus adult-onset MDD. Evidence from earlier studies pointed towards a more severe phenotype in pre-adult onset MDD. However, these findings again were mostly derived from studies in either population-based studies or clinical trials. In multivariable analyses in 1,105 MDD patients, we found that patients with pre-adult onset MDD more often had a history of suicide attempts and current suicidal thoughts compared with adult-onset MDD patients. This important study finding replicated earlier findings in clinical samples and supports the hypothesis that pre-adult MDD is characterised by a distinct phenotype compared with adult-onset MDD. Other previous findings, e.g. having more severe symptoms (Zisook et al., 2004; Zisook et al., 2007) or anxiety (Parker et al., 2003) in pre-adult onset MDD, could not be replicated. This is important to ascertain, since our samples more truly represent 'real' patients than previous studies did. Although possibly confounded by recall bias and measurement error, the age of onset of MDD can be obtained relatively easy and may help the clinician in risk-assessment of suicidality.

After these first two studies in MDD patients we expanded our cross-sectional analyses to all patients with MAS disorders. In Chapter 4 we studied prevalence and correlates of deliberate self-harm and suicidal ideation (DSHI) in a large group of patients referred for treatment of either a mood, anxiety or somatoform disorder, or a combination of these diagnoses. We analysed 2,844 patients with a ROM baseline assessment and used the self-harm subscale of the Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF) to assess DSHI. Lifetime DSHI was reported by 55% of the total sample. Being married, having a higher educational status and having high anxiety symptoms were independently associated with lower risks of DSHI. Having a higher number of MAS diagnoses, having more depressive symptoms and higher scores on the emotional dysregulation subscale of the DAPP-SF were associated with higher risks of DSHI. The negative association with anxiety measured with the Brief Anxiety Scale (BAS) in adjusted analyses was unexpected. The most plausible explanation of this finding would be that patients with high anxiety levels but similar levels of depressive symptoms were too afraid to hurt themselves. These results in a unique broad sample of outpatients with common mental disorders demonstrate that DSHI is a common phenomenon in MAS outpatients, with distinct clinical correlates. The findings may help clinicians to identify patients at

increased risk for DSHI and completed suicide.

In the next part of this thesis we shifted the focus to clinical predictors of real-world outcomes, using prospective ROM measurements in addition to the baseline assessments. In Chapter 5 we used ROM baseline and follow-up assessments of up to 24 months to investigate whether individual baseline depressive symptoms would predict outcome in 1,489 MDD patients. In earlier studies, many clinical correlates of MDD-treatment outcome have been identified. Earlier age of onset, more severe complaints and cluster B and C personality traits were among the predictors of less favourable outcome (see for example Frank et al., 2010; Penninx et al., 2011; Souery et al., 2007; Vuorilehto et al., 2009). Most studies had a limited follow-up time of three months to one year. Moreover, the predictive value of specific depressive symptoms at baseline had not been systematically assessed before. In daily clinical practice, being able to predict which patients are at risk to become non-responders or develop a chronic course of the depression is highly important. We analysed the predictive value of all 21 individual depressive symptoms measured with the widely used Beck depression Inventory-Revised (BDI-II; Beck et al., 1988) on naturalistic treatment outcome, defined as remission (score <10) or response (≥50% reduction) on the Montgomery-Åsberg Depression Rating Scale (MADRS; Montgomery, 1979). Using multivariable Cox regression analyses, we found that after adjusting for known clinical variables correlating with treatment outcome, baseline presence of the two major symptoms 'pessimism' and 'loss of energy' strongly and independently predicted a higher chance of a chronic course. In daily clinical practice, the presence or absence of these symptoms are relatively easy to assess, and the presence of pessimism and/or loss of energy should alert the clinician to the higher risk of a chronic MDD course.

In the final study (Chapter 6), we extended the prediction of real-world outcomes to the total group of patients with MAS disorders. In the DSM-IV-TR, MAS disorders show a considerable overlap in diagnostic criteria and in daily practice mutual comorbidity of these disorders is substantial. Although in most clinical trials only patients with certain primary axis-I diagnoses are included, the concept of primary diagnosis is rather vague (van der Lem et al., 2010). The frequent occurrence of MAS-comorbidity complicates the establishment of the primary or main diagnosis. Hence, in daily clinical practice, a patient with for example a MDD and a social phobia is treated according to the MDD guideline, the social phobia guideline, or a combination of both. In this particular study we aimed to assess predictors of 'cross-diagnostic' outcome in clinical practice. We defined outcome as a certain reduction on a combination of two disorder independent measurement scales: the observer-rated Clinical Global Improvement Scale (CGI; Kadouri et al., 2007) and the self-report Brief Symptom Inventory (BSI; Derogatis and Melisaratos, 1983).

We used baseline ROM data as well as prospective data with up to 24 months of follow-up. We used a sample of 892 MAS patients, and validated our results in an independent sample of 1,392 MAS patients. In multivariable Cox regression analyses, we found that after adjusting for age and gender, the following characteristics were independent predictors of unfavourable outcome: advancing age, having MAS comorbidity or a somatoform disorder, the personality traits intimacy problems, affect lability and self-harm, and a low self-reported general health status. These findings were confirmed in a second sample of MAS patients, and could help the clinician to identify patients who are at risk for a chronic course of the MAS disorder. Whether these patients, based on the above profile, could benefit from extra interventions should be the focus of future studies.

General discussion

Routine Outcome Monitoring is the systematic assessment of treatment outcomes in everyday clinical practice. In psychiatric specialty care, ROM assessments include standardised and validated measurement instruments. The ROM-infrastructure in Leiden uses a broad test battery that comprises a range of both self-report and observer-rated scales. The test battery includes a diagnostic instrument, instruments that measure generic and disorder-specific symptom severity, and scales that measure psychosocial functioning. In general, ROM is primarily used as a tool for patient and clinician to monitor treatment progress. Although more specific research is necessary, studies indicate a positive impact of ROM and feedback on mental health status (Carlier et al., 2010). In anonymised form, the systematically obtained data can also be used for clinical epidemiological research purposes (de Beurs et al., 2011), as exemplified by this thesis. Whenever a broad range of instruments is applied, detailed phenotypic information of patients can be derived from ROM. In the Leiden region in the Netherlands, ROM assessments are part of the routine diagnostic and therapeutic procedure in patients with MAS disorders in the outpatient clinics at the Leiden University Medical Center (LUMC) and Rivierduinen (RD; see Box 1.1, Chapter 1). In Leiden, assessments ideally take place directly after referral (ROM baseline assessment) and at any step or change in the treatment protocol. In practice, this is on average every 3-4 months repeatedly during treatment. In Leiden, only patients who are unable to undergo ROM assessments due to language problems, disease severity or refusal are excluded from ROM. With approximately 80% of all patients taking part in the Leiden ROM, results are highly generalisable to everyday clinical practice.

The Leiden ROM infrastructure is an example of an extensive ROM infrastructure. This extensive ROM has several advantages, e.g. detailed phenotyping, staging of psychiatric disease, identification of residual symptoms ('rollback phenomena', see Fava et al. 2011). Furthermore, it provides opportunities for clinical epidemiological research and benchmarking. However, important differences exist among implemented ROM systems worldwide. For example, some institutions apply routine monitoring with the use of only one self-report instrument, that is administered at every visit (e.g. Zimmerman et al., 2005).

The studies presented in this thesis focused on clinical characteristics and predictors of outcome in everyday patients with mood, anxiety and somatoform disorders. Although this seems a trivial fact, studies in naturalistic samples of MAS patients in psychiatric specialty care were scarce. The large-scale ROM infrastructure in the Leiden region provided a unique opportunity to investigate important clinical aspects in a relatively unselected treatment-seeking outpatient population, yielding findings that are easy to generalise to routine clinical outpatient practice. This contrasts with findings from clinical trials that included highly selected patient samples resulting in a more limited generalisability. Findings from population studies do neither reflect a treatment-seeking population. In fact, 75-80% of our ROM MDD-patients would not have met the in- and exclusion criteria for most clinical trials (van der Lem et al., 2010).

The findings of the explorative studies in this thesis have provided unique insights in the phenomenology of 'real' MAS patients in daily clinical practice. Predictors of naturalistic outcome have been established. In addition, the strengths and weaknesses of a large-scale and extended ROM initiative have become tangible. In the next paragraphs, we will discuss our findings in a broader perspective.

Phenotype of real-world MAS patients

In the process of designing the studies presented in this thesis, we decided to start our explorative analyses with a well-studied topic in psychiatric literature. As mentioned before, gender differences in depression have been studied extensively. The naturalistic character of our study sample was a new point of view. The results of this first study partly confirmed and extended previous research. The most important implication of the first study was that for a balanced ROM assessment both observer-rated and self-report scales are necessary, which supported earlier work of Möller (2009). If the ROM assessments would have comprised only self-rating instruments, important nuances would have been missed. After all, we found no differences in number nor severity of depressive symptoms

between men and women on observer-rated scales, while substantial differences existed on self-rated scales. This stresses the need for maintaining the observer-rated scales in ROM, even in an era of inevitable cost reductions in healthcare. Another important conclusion of the first study in this thesis was that the quality of the anonymised patient data provided by our ROM database was successfully applied for the intended analyses.

The main finding of the second study was that MDD patients with disease onset before the age of 18 years were more suicidal compared to those with onset at adult age. This was in part a replication of earlier studies in different samples. The fact that in multivariable analyses all differences in clinical profile, except suicidality, became non-significant could at least in part explain the high variability of the findings in earlier studies, which not always had included a comprehensive set of potential confounders. The higher rates of current and past suicidality in pre-adult onset MDD may have important clinical implications, as completed suicide needs to be prevented in patients with psychiatric disorders, and it is one of the leading causes of death worldwide. The estimated global burden of suicide is a million deaths per year (WHO, 2002), or 14.5 deaths per 100.000 people worldwide. These are probably underestimations. Large differences exist between countries, and men are much more at risk for completed suicide with a male-female ratio of 2-4 to 1 in developed countries (WHO, 2002). Previous self-harm or suicide attempts are major risk factors for completed suicide (Cavanagh et al., 2003; Coryell and Young, 2005). Based on psychological autopsies in completed suicide patients, it is estimated that 90% (Cavanagh et al., 2003) or even more (Ernst et al., 2004) had a DSM-IV Axis I psychiatric disorder at the time of suicide. More than half of the patients who die of suicide meet criteria for current MDD (Cavanagh et al., 2003; Hawton and van Heeringen, 2009), and about 4% of MDD patients die by suicide. Despite dramatic increases in awareness, prevention and treatment efforts, no significant decrease in suicide rates has occurred in the US from 1990 to 2003 (Kessler et al., 2005). These figures stress the need for early and better identification of suicidality, especially in MDD patients, and the need for specific interventions. Based on the results of Chapter 2, clinicians should be especially aware of suicidality in male MDD patients with pre-adult onset MDD and previous suicide attempts.

The finding of more suicidality in pre-adult onset MDD patients increased our interest in the phenomenon of suicidality and the broader concept of deliberate self-harm. Since suicidal patients and patients who engage in self-destructive behaviour are excluded from most clinical trials in psychiatry, we investigated these phenomena in depth in our total group of MAS patients. This third study also in part replicated earlier studies in more strictly selected populations, but also yielded new insights. The most important outcome was the apparent protective effect of state anxiety on deliberate self-harm and

suicidal ideation. Of course, this finding needs replication in future studies. Furthermore, the broad sample of 2,844 treatment-seeking MAS patients with substantial comorbidity allowed us to identify correlates of self-harm irrespective of DSM-IV diagnosis, which increased applicability in clinical practice.

Predictors of real-world outcomes

One of the most important topics for both patients and clinicians regards the chances of getting better. This is of course not specific for psychiatric patients. Large-scale epidemiological studies have identified risk factors for adverse outcomes in a wide range of somatic diseases or disease variants. These studies have resulted in risk-assessment schedules that have been widely adapted in worldwide guidelines. A famous example is the Framingham heart study, a cohort study of over 5,000 inhabitants of Framingham, Massachusetts. This study, initiated in 1948, has calculated detailed chances of cardiovascular events if a patient corresponds to a certain profile, based on gender, age, systolic blood pressure, and serum lipids (Anderson et al., 1991; Dawber et al., 1951). Based on this study, with numerous publications and an offspring study (Kannel et al., 1979), patients with certain risk profiles are considered as subgroups for which risk scores can be calculated and different treatments can be advised. International guidelines have adapted these risk calculators. In Europe, the Systematic Coronary Risk evaluation (SCORE) project has used data of over 200.000 European patients to calculate estimations of ten-year risk of fatal cardiovascular disease (Conroy et al., 2003). As discussed in the introduction of this thesis, for psychiatric disorders, those detailed studies have never been performed. Based on the available epidemiological literature, only rough estimations about the course of the disease can be provided for the total group of patients with a certain disorder (see for example Judd, 1997). Until now, guidelines and treatment algorithms only roughly differentiate between subgroups of patients. For example: The Dutch Multidisciplinary Guideline for the treatment of MDD (www.ggzrichtlijnen.nl) discriminates between first episode or recurrent episode, and between mild or moderate/severe depression. Furthermore, patients with moderately to severe depression have the options of a modern antidepressant or psychotherapy, whereas the treatment algorithm for patients with Major Depression with psychotic features suggests treatment with a tricyclic antidepressant, with possible addition of an antipsychotic agent. Depression with atypical features may better respond to monoamine-oxidase inhibitors than other classes of antidepressants. However, also for psychiatrists, being able to predict the chance of remission in more

sophisticated subgroups of patients is of utmost importance, yet still very imprecise. Based on the findings of the studies in this thesis, in the next part of this discussion we present a theoretical future algorithm of risk assessment in MDD (figure 7.1).

The fourth study of this thesis can be interpreted as an attempt to predict outcomes of MDD in a real-world treatment setting based on baseline symptoms that can easily be obtained. We have demonstrated that ROM assessments can be helpful in risk assessment and that the baseline presence of both lack of energy and pessimism symptoms measured with the widely used BDI-II, were highly predictive of non-remission of MDD after a follow-up of up to2 years. In the fifth and final study of this thesis we aimed to establish predictors of poor outcome in the whole group of patients with MAS disorders. The broad phenotypic assessment with ROM allowed us to closely investigate demographical and clinical risk factors of poor outcome in a naturalistic treatment setting. Furthermore, we used a cross-diagnostic design in which we included all MAS patients thereby abandoning the rather vague concept of primary diagnosis often used in studies. This made the results of this study easy and widely applicable to clinical practice. Although much more research is needed, our longitudinal ROM studies can be seen as one further step towards a more detailed risk assessment, based on extensive phenotypic data. In the future, hopefully, adding genotypes and endophenotypes (biological substrates) to phenotypic data may further enhance the precision of risk-assessment. In addition, important environmental information that might be important in epigenetic effects (e.g. childhood trauma, daily stress) may also be ascertained with ROM. Efforts in finding biological substrates of symptom dimensions of depression and anxiety have already been published (Veen et al., 2011; Wardenaar et al., 2011).

ROM and Comparative Effectiveness Research

In the literature, research based on ROM data is often regarded as 'patient-centered research'. If treatment details are taken into account, ROM-data driven research could be used for Comparative Effectiveness Research (CER). CER is designed to improve the clinical decision-making process by providing research evidence on the effectiveness and risk-benefit profile on different therapeutic options for specific patient subpopulations (Mane et al., 2011; Sox and Greenfield, 2009). In the US, the Patient-Centered Outcomes Research Institute (PCORI) has been established to facilitate CER. The mission of this institute is to help people make informed healthcare decisions, and to improve healthcare delivery and outcomes (Washington and Lipstein, 2011). The US government has offered

a funding of more than 3 billion US Dollar for the next decade, allowing the support of numerous studies of CER, in different areas of healthcare. In Europe, to our knowledge, no comparable large-scale initiatives are being developed. In theory, ROM databases of different institutions could be merged in large collaboration efforts and used for CER. For such overarching initiatives to be successful, many consensus steps have still to be taken. On a national level, for The Leiden Institute for the Advancement and Integration of ROM (LIAIROM), achieving consensus about implementation of ROM and e.g. choice of measurement instruments in order to integrate ROM databases is a priority area. For example: based on our studies, specific interventions for MDD patients meeting the profile of advanced age, comorbidity, cluster B traits and poor reported general health could be investigated. Or if treatment details become available in ROM, it would be possible to investigate which antidepressant is associated with the highest chance of remission in patients with baseline pessimism and lack of energy.

A theoretical future algorithm of risk assessment and treatment recommendations in MDD patients which resembles the cardiovascular disease risk calculation is presented in figure 7.1.

In cardiovascular risk management differentiated treatment recommendations (e.g. lifestyle adaptations with or without treatment with statins) are based on several phenotypic and endophenotypic aspects (e.g. smoking and systolic blood pressure). One could imagine that in future psychiatric practice, comparable treatment recommendations would be possible (e.g. lifestyle adaptations, first-step interventions, versus psychotherapy or pharmacotherapy or extra interventions), based on ROM assessments comprising detailed phenotypic information and information about environmental factors, combined with genotypes and endophenotypes (e.g. cortisol levels).

ROM critically appraised

One of the aims of this thesis was to assess the feasibility of conducting large-scale clinical epidemiological research with data provided by ROM. The studies presented here demonstrate that the ROM infrastructure in RD/LUMC provide vast possibilities in conducting both cross-sectional and prospective analyses, with the opportunity of studying specific phenotypes of MAS patients and risk factors of outcomes and disease course. Several aspects deserve further notification. We will now focus on the strengths and weaknesses of ROM. We will differentiate between the Leiden ROM system and ROM in general.

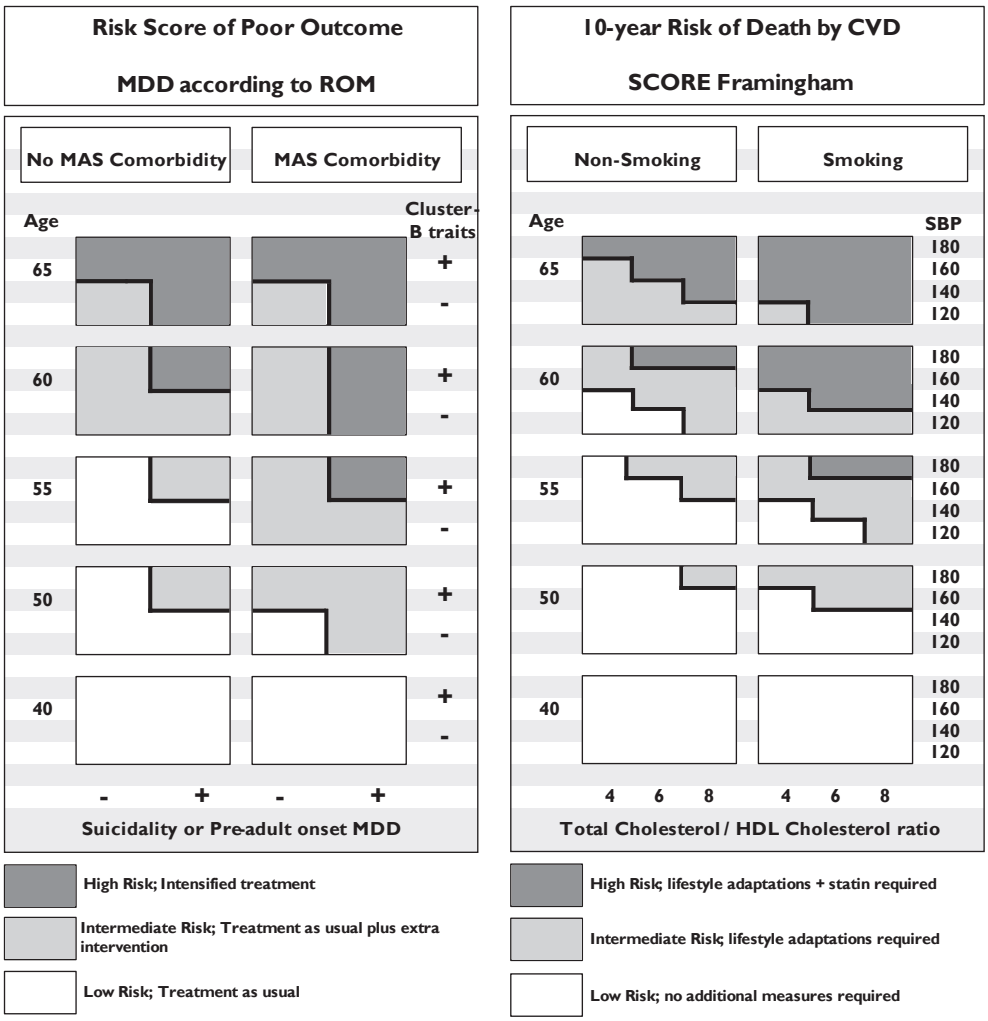

Figure 7.1. Theoretical future risk-assessment calculation table for MDD, in analogy of the (adapted) SCORE/ Framingham risk-assessment calculation table for 10-year risk of death by cardiovascular disease; Abbreviations: ROM: Routine Outcome Monitoring; MDD: Major Depressive Disorder; SCORE: Systematic Coronary Risk Evaluation; MAS: Mood, Anxiety and Somatoform; SBP: Systolic Blood Pressure; HDL: High-Density Lipoprotein; CVD: Cardiovascular Disease. MAS comorbidity denotes the co-occurrence of more than one MAS disorder.

The Leiden ROM infrastructure has been developed and implemented with the combined goals of treatment monitoring for patient and clinician, and to serve as a large database for research and for benchmarking. The use of computerised assessments and dedicated software prevents missing data within specific instruments, and provides high data quality. Another strength is the fact that assessments are performed by specially trained research nurses who are not involved in treatment, ensuring neutral assessments without the risk of measurement bias that could occur when clinicians would rate their own patients. Moreover, the extensive baseline assessments, including a comprehensive and well validated diagnostic measurement instrument (MINI-Plus), an instrument that assesses dysfunctional personality traits (DAPP-SF) and repeatedly administered scales that cover the symptom level and psychosocial functioning, provide detailed phenotypic information at every phase of treatment. Both phenotypic research on a group level and outcomes research can be performed. The choice of measurement instruments at the time of implementation (from 2002) was based on psychometric quality and applicability of the instrument, and on the availability in the public domain. For longitudinal research, obviously, the test battery should remain extensive and ideally unchanged. Unfortunately, in the past years copyright issues forced us to abandon certain scales (e.g. DAPP-SF, BSI). The high rate of ROM baseline assessments of approximately 80% is another strength. Because ROM is integrated in everyday clinical practice, only patients who are not fluent in Dutch or those patients who are too ill to undergo assessments are excluded from ROM.

By utilising ROM data for the studies in this thesis we encountered several limitations of the Leiden ROM infrastructure. The most important limitation of the present Leiden ROM system is the current lack of integration with the (anonymous) electronic patient record forms (PRFs), preventing the researcher from retrieving information about exact treatment modality, detailed psychiatric history, family history of psychiatric diseases, and somatic status. In order to obtain detailed information about e.g. treatment, the researcher has to hand-search the patient record files. This hand-searching could compromise anonymity. Integration of ROM with PRFs would have given us the opportunity to adjust for treatment modality or guideline adherence in the two longitudinal studies (Chapters 5 and 6). In these studies we have not been able to adjust for possible confounding by the treatment type, patient characteristics, and disease characteristics. On the patient level, integration of ROM and PRF could improve decision making by identifying subpopulations of patients with similar characteristics and to determine which treatments have been successful with minimal side-effects (Iglehart, 2009; Mane et al., 2011). This type of research is commonly regarded as Comparative Effectiveness Research (CER). Another limitation of our present ROM system regards the time intervals of ROM assessments. From a clinical point of view,

ideally follow-up assessments should be scheduled before every next treatment step. From a research perspective, ROM measurements at fixed time intervals, independent from disease severity are to be preferred for most longitudinal data analytic techniques. Cox regression analysis, the method we used in Chapter 5 and 6 for predicting outcome, is an example of a technique that is able to use variable numbers of measurements and variable time-intervals. In the current clinical practice in Leiden, measurements are initiated by the research nurse, the clinician or patient, but not always before a change of treatment or a next treatment step according to the guidelines. This unpredictable interval of ROM assessments without automatically scheduled re-assessments may contribute to bias due to attrition and complicates data analysis and interpretation. At present, efforts are made to incorporate a more systematic measurement interval in the Leiden ROM.

A potential limitation or pitfall of ROM in general is important to consider. In this modern era of excessive growth of healthcare and inevitable health-costs, the power of health insurance companies is growing and the professional autonomy of medical specialists is increasingly under pressure. Understandably, policy makers like health insurance companies and governments demand more and more insight in costs of treatments and treatment processes to be able to control these costs. Since ROM is a potentially valuable source of information regarding these processes, many health providers have implemented ROM initiatives over the past years, with benchmarking as one of the major goals. The studies in this thesis may illustrate that benchmarking based on ROM assessments is possible but could also be a hazardous operation. First of all, some institutions have adapted a ROM system in which only a succinct set of outcome scales or only one scale is used. Even with our extended ROM measurement scales, it would be hard to derive reliable benchmarking data because of important inter-patient differences that require complex statistics to take into account. For example, if in a certain clinic more MDD patients with comorbid personality disorders are being treated, outcomes may be worse compared to another clinic that uses the same guidelines but where patients with less complicated complaints seek treatment. Our major concern would be that in the case of a limited ROM assessment battery, policy makers will draw conclusions based on insufficient or inadequately analysed data. For example, in tertiary care clinics or specialised secondary care clinics, typically patients with treatment-resistant complaints, somatic comorbidity, co-existent personality pathology, or a combination of these are being treated. Those patients are likely to have worse treatment-outcomes, irrespective of the quality of treatment, as compared with less complicated patients in general psychiatric specialty care.

In this thesis we have shown that for a balanced ROM assessment it is necessary to apply both self-report and observer-rated scales. Ironically, due to the aforementioned efforts to control healthcare costs, the current tendency is to limit or even omit the observer-rated scales and standardised diagnostic interviews from ROM. To base clinical and benchmark decisions solely on self-report scales would perhaps save salary costs of research nurses. However, eventually these self-report measurements could only reflect part of the process of diagnostics and treatment outcome because more objective information derived from observer-rated scales would be lacking. In this thesis, we have shown that ROM, including observer-rated measurements, provides valuable and more objective information. Finally, the large variety of ROM assessment scales that are being used by different health organisations limits comparability and is another caveat with respect to benchmarking. Calculating standardized scores or z-scores could help with respect to this problem.

A last limitation of ROM in general to discuss here relates to the population under study. The studies described here focused on patients with MAS disorders, also regarded as common mental disorders. Our local decision to implement ROM firstly in MAS disorder patients was based on the research profile of our department. However, substantial overlap of symptoms and mutual comorbidity of MAS disorders, the wide availability of validated rating scales in the MAS domain, and the fact that MAS patients were expected to be more compliant than e.g. psychotic patients or patients with personality disorders as main diagnosis, also played a role in this decision. Indeed, in several ROM initiatives in other patient populations, e.g. psychotic patients or Severe Mental Illness (SMI)-patients, these considerations have been underlined (Mulder et al., 2010).

Future Perspectives

The studies presented in this thesis can be regarded as a first set of explorative studies to demonstrate the possibilities of conducting clinical epidemiological research with ROM data. Several adjustments, e.g. adding treatment details in ROM, may further improve the potential of future studies in our setting (see table 7.1).

Besides for clinical epidemiological research, the Leiden ROM data can serve as basis for research in other domains. Examples of these are biological and psychometric research. In Leiden, many researchers already use the ROM-database for studies in these domains. Results have been published and will be in the near future. To further illustrate the potential of ROM, we will give some examples of current projects that use the Leiden ROM infrastructure.

First, the lack of well-known biological markers in psychiatric disorders complicates the borders of disease. When can someone with certain complaints be classified as a patient? The diagnostic classification system DSM-IV only partially answers that question by operationalising disorders by consensus definitions. Due to the absence of clear markers and borders, the line between 'healthy' and 'sick' will be hard to define. Validated measurement scales are helpful in defining and establishing that line. However, for many validated measurement scales used in ROM no reference values in the general population have been calculated. NormQuest is a study initiated in Leiden that aims to assess those reference values for the commonly used measurement scales in MAS patients (Schulte et al., in press).

Second, the ROM infrastructure with naturalistically obtained data also allows for add-on research. The Mood, Anxiety, Somatoform disorders and Hypothalamus pituitary adrenal (HPA) axis Biobank (MASHBANK) is an example of this type of research. The MASHBANK has been founded to investigate the link between genetic variants, functioning of the HPA axis and the phenotype in patients with MAS disorders, and comparing those with patients form the general population. Patients routinely enrolled in ROM have been asked informed consent to donate DNA for the MASHBANK, after MEC approval of the protocol. So far, almost 2000 samples of MAS patients and control subjects have been collected.

Third, as discussed earlier, ROM measurement scales should ideally be free of copyright in order to allow broad implementation. For longitudinal research, a stable test battery is paramount. Unfortunately, the past years publishers have claimed copyrights for measurement scales used in our ROM test battery forcing us to abandon those scales. The ROM Research Center (COROM) in Leiden facilitates ROM-related research of which the development of new measurement scales for the free domain is one of the topics. The development and validation of freely available scales comparable to copyright-protected scales are needed. An example of the latter is the construction of a new measure for the assessment of psychological distress. We developed and validated the Symptom Questionnaire-48 (SQ-48), a measurement scale that assesses 48 symptoms in nine symptom domains (Carlier et al., in preparation). Furthermore, translation and subsequent validation of measurement scales in different languages is necessary to allow non-Dutch patients to be enrolled in ROM.

As mentioned before, the studies in this thesis are examples of the potential of ROM. Many more studies and collaborations are necessary to develop useful risk calculators, to integrate biological markers in ROM, and to develop specific interventions for subgroups of patients. Although practical obstacles may have to be faced, lessons

from the cardiovascular field (Framingham, SCORE) but also oncology demonstrate that large-scale collaboration may dramatically improve outcomes step by step in large groups of patients. For example, acute leukemia is the most common form of childhood cancer, comprising approximately 30% of all malignancies in children. Survival rates for Acute Lymphatic Leukemia have increased dramatically since the 1980s, with current five-year overall survival rates of over 85% (Gatta et al., 2005; Pui et al., 2004; Pui and Evans, 2006). These improved survival rates are due to large-scale collaborations and treatment of large groups of patients according to standardised research protocols, based on staging of disease and constant monitoring of outcomes. These protocols have evolved over and over according to outcomes of trials and findings of more biological studies (Lee et al., 2000; Pui and Evans, 2006). In psychiatry, establishment of the international schizophrenia consortium (ISC) has resulted in large-scale genetic studies in schizophrenia and bipolar disorder (e.g. Purcell et al., 2009). In Europe, the GENDEP consortium aims to use genetic profiles to predict outcome of antidepressant treatment (e.g. Uher, et al., 2010). These initiatives demonstrate that large-scale collaborations are possible. Indeed, initiatives for staging and profiling in psychiatry are emerging. In a recent review of Fava et al. (2011), the authors stress the importance of staging in psychiatry. "Staging differs from conventional diagnostic practice in that it not only defines the extent of progression of a disorder in a particular point in time but also reveals a person's current location on the continuum of the course of illness". Detailed information about symptoms, comorbidity, psychosocial functioning, quality of life and response to treatment derived by ROM assessments could serve as basis for staging.

Concluding remarks

The studies in this thesis demonstrate that a proper ROM infrastructure is a valuable source for clinical epidemiological research in naturalistic or 'real' patients. We have studied in detail aspects of the phenotype of MAS patients, and established risk factors of poor outcome in both MDD and a treatment-seeking MAS population. Major findings were the phenotype of more suicidality in pre-adult onset MDD patients, and correlates of deliberate self-harm in a large group of MAS patients. Pessimism and lack of energy were found to be predictors of poor outcome in MDD patients, and advancing age, dysfunctional cluster-B personality traits, MAS comorbidity and poor reported general health status were risk factors for poor outcome in MAS disorders. These risk factors need future replication in other naturalistic cohorts, and could result in detailed risk-assessment

calculators comparable to those used in cardiovascular medicine. Integration of treatment details in ROM would allow for comparative effectiveness studies in specific subgroups of patients. Integration of biological markers in ROM and (inter)national collaboration could bring ROM to a higher level. On a national level, regional mental health providers should give priority to reaching consensus about uniform implementation of ROM. We believe that consensus is important to be able to respond to demands of policy makers. A long-term vision is necessary to achieve this goal.

References

Anderson, K.M., Odell, P.M., Wilson, P.W., Kannel, W.B., 1991. Cardiovascular disease risk profiles. American. Heart Journal 121, 293-298.

Beck, A.T., Steer, R.A., Carbin, M.G., 1988. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical Psychology Review 8, 77-100.

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.

Carlier, I.V., Meuldijk, D., van Vliet, I.M., van Fenema, E., 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).

Carlier, I.V., Schulte-van Maaren, Y.W., Wardenaar, K.J., Giltay, E.J., van Noorden, M.S., Vergeer, P., Zitman, F.G. Development and validation of the 48-item Symptom Questionnaire in patients with depressive, anxiety and somatoform disorders. (in preparation).

Cavanagh, J.T., Carson, A.J., Sharpe, M., Lawrie, S.M., 2003. Psychological autopsy studies of suicide: a systematic review. Psychological Medicine 33, 395-405.

Conroy, R.M., Pyorala, K., Fitzgerald, A.P., Sans, S., Menotti, A., De Backer, G., De Baquer, D., Ducimetiere, P., Jousilahti, P., Keil, U., Njolstad, I., Oganov, R.G., Thomsen, T., Tunstall-Pedoe, H., Tverdal, A., Wedel, H., Whincup, P., Wilhelmsen, L., Graham, I.M., 2003. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European Heart Journal 24, 987-1003.

Coryell, W., Young, E.A., 2005. Clinical predictors of suicide in primary major depressive disorder. Journal of Clinical Psychiatry 66, 412-417.

Dawber, T.R., Meadors, G.F., Moore, F.E., Jr., 1951. Epidemiological approaches to heart disease: the Framingham Study. American Journal of Public Health and the Nation's Health 41, 279-281.

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

Ernst, C., Lalovic, A., Lesage, A., Seguin, M., Tousignant, M., Turecki, G., 2004. Suicide and no axis I psychopathology. BMC Psychiatry 4, 7.

Fava, G.A., Rafanelli, C., Tomba, E., 2011 The clinical process in psychiatry: A clinimetric approach. Journal of Clinical Psychiatry (in press).

Frank, E., Cassano, G.B., Rucci, P., Thompson, W.K., Kraemer, H.C., Fagiolini, A., Maggi, L., Kupfer, D.J., Shear, M.K., Houck, P.R., Calugi, S., Grochocinski, V.J., Scocco, P., Buttenfield, J., Forgione, R.N., 2011. Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. Psychological Medicine 41, 151-162.

Gatta, G., Capocaccia, R., Stiller, C., Kaatsch, P., Berrino, F., Terenziani, M., 2005. Childhood cancer survival trends in Europe: a EUROCARE Working Group study. Journal of Clinical Oncology 23, 3742-3751.

Hawton, K., van Heeringen, K., 2009. Suicide. Lancet 373, 1372-1381.

Iglehart, J.K., 2009. Prioritizing comparative-effectiveness research--IOM recommendations. New England Journal of Medicine 361, 325-328.

Judd, L.L., 1997. The clinical course of unipolar major depressive disorders. Archives of General Psychiatry 54, 989-991.

Kadouri, A., Corruble, E., Falissard, B., 2007. The improved Clinical Global Impression Scale (iCGI): development and validation in depression. BMC Psychiatry 7, 7.

Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J., Castelli, W.P., 1979. An investigation of coronary heart disease in families. The Framingham offspring study. American Journal of Epidemiology 110, 281-290.

Kessler, R.C., Berglund, P., Borges, G., Nock, M., Wang, P.S., 2005. Trends in suicide ideation, plans, gestures, and attempts in the United States, 1990-1992 to 2001-2003. Journal of the American Medical Association 293, 2487-2495.

Lee, S.J., Earle, C.C., Weeks, J.C., 2000. Outcomes research in oncology: history, conceptual framework, and trends in the literature. Journal of the National Cancer Institute 92, 195-204.

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.

Mane, K.K., Bizon, C., Schmitt, C., Owen, P., Burchett, B., Pietrobon, R., Gersing, K., 2011. VisualDecisionLinc: A visual analytics approach for comparative effectiveness-based clinical decision support in psychiatry. Journal of Biomedical Informatics (in press).

Möller, H.J., 2009. Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. World Journal of Biological Psychiatry 10, 6-26.

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.

Parker, G., Roy, K., Hadzi-Pavlovic, D., Mitchell, P., Wilhelm, K., 2003. Distinguishing early and late onset non-melancholic unipolar depression. Journal of Affective Disorders 74, 131-138.

Penninx, B.W., Nolen, W.A., Lamers, F., Zitman, F.G., Smit, J.H., Spinhoven, P., Cuijpers, P., de Jong, P.J., van Marwijk, H.W., van der Meer, K., Verhaak, P., Laurant, M.G., de Graaf, R., Hoogendijk, W.J., van der Wee, N.J., Ormel, J., van Dyck, R., Beekman, A.T., 2011. Two-year course of depressive and anxiety disorders: Results from the Netherlands Study of Depression and Anxiety (NESDA). Journal of Affective Disorders (in press).

Pui, C.H., Evans, W.E., 2006. Treatment of acute lymphoblastic leukemia. New England Journal of Medicine 354, 166-178.

Pui, C.H., Sandlund, J.T., Pei, D., Campana, D., Rivera, G.K., Ribeiro, R.C., Rubnitz, J.E., Razzouk, B.I., Howard, S.C., Hudson, M.M., Cheng, C., Kun, L.E., Raimondi, S.C., Behm, F.G., Downing, J.R., Relling, M.V., Evans, W.E., 2004. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. Blood 104, 2690-2696.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748-752.

Schulte-van Maaren, Y.W., Carlier, I.V., Giltay, E.J., van Noorden, M.S., de Waal, M.W., van der Wee, N.J., Zitman, F.G., 2011 Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study (in press).

Souery, D., Oswald, P., Massat, I., Bailer, U., Bollen, J., Demyttenaere, K., Kasper, S., Lecrubier, Y., Montgomery, S., Serretti, A., Zohar, J., Mendlewicz, J., 2007. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. Journal Clinical Psychiatry 68, 1062-1070.

Sox, H.C., Greenfield, S., 2009. Comparative effectiveness research: a report from the Institute of Medicine. Annals of Internal Medicine 151, 203-205.

Uher, R., Perroud, N., Ng, M.Y., Hauser, J., Henigsberg, N., Maier, W., Mors, O., Placentino, A., Rietschel, M., Souery, D., Zagar, T., Czerski, P.M., Jerman, B., Larsen, E.R., Schulze, T.G., Zobel, A., Cohen-Woods, S., Pirlo, K., Butler, A.W., Muglia, P., Barnes, M.R., Lathrop, M., Farmer, A., Breen, G., Aitchison, K.J., Craig, I., Lewis, C.M., McGuffin, P., 2010. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. The American Journal of Psychiatry 167, 555-564.

Veen, G., van Vliet, I.M., De Rijk, R.H., Giltay, E.J., van Pelt, J., Zitman, F.G., 2011. Basal cortisol levels in relation to dimensions and DSM-IV categories of depression and anxiety. Psychiatry Research 185, 121-128.

Vuorilehto, M.S., Melartin, T.K., Isometsa, E.T., 2009. Course and outcome of depressive disorders in primary care: a prospective 18-month study. Psychological Medicine 39, 1697-1707.

Wardenaar, K.J., Vreeburg, S.A., van Veen, T., Giltay, E.J., Veen, G., Penninx, B.W., Zitman, F.G., 2011. Dimensions of depression and anxiety and the hypothalamo-pituitary-adrenal axis. Biological Psychiatry 69, 366-373.

Washington, A.E., Lipstein, S.H., 2011. The Patient-Centered Outcomes Research Institute - Promoting Better Information, Decisions, and Health. New England Journal of Medicine (in press).

World Health Organisation. World report on violence and health. 2002, WHO edn., Geneva.

Zimmerman, M., Posternak, M.A., Chelminski, I., Friedman, M., 2005. Standardized clinical outcome rating scale for depression for use in clinical practice. Depression and Anxiety 22, 36-40.

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.

Zisook, S., Rush, A.J., Albala, A., Alpert, J., Balasubramani, G.K., Fava, M., Husain, M., Sackeim, H., Trivedi, M., Wisniewski, S., 2004. Factors that differentiate early vs. later onset of major depression disorder. Psychiatry Research 129, 127-140.