Introduction

Psychological assessment in mental-health practice involves the diagnosis and prediction of human behavior. For example, a psychologist may want to assess whether a patient suffers from generalized anxiety disorder. Or, a psychiatrist may want to predict whether a patient will show the greatest reduction in symptoms by receiving psychotherapy, pharmacotherapy, or a combination of both. Diagnosis and prediction of human behavior requires a practitioner to perform two basic tasks: collection and interpretation of data (Dawes, Faust & Meehl, 1989). In this thesis, I will address a broad range of issues in clinical assessment, and introduce new approaches to improving the efficiency and accuracy of the collection and interpretation of data in clinical practice.

According to the view on assessment put forth by Cronbach en Gleser (1965), in many practical settings, the ultimate purpose of testing is decision making. Therefore, in the collection and interpretation of data in clinical practice, the decision for which attributes are assessed should be taken into account. This view gives rise to many opportunities for improving efficiency and accuracy of assessment. For example, item administration within a test can be halted whenever administration of further items can no longer change the resulting decision. Or, a sequential approach to assessment can be taken, by selecting the attribute to be evaluated next, based on what is already known about a patient. This can greatly reduce assessment length, because attributes that are redundant for making a decision, given the value of attributes already assessed or known, need not be assessed. In turn, reducing assessment length will reduce the burden of assessment for both patients and practitioners, because less data has to be collected and interpreted for making a prediction.
Introduction

Optimizing accuracy and efficiency of clinical assessment can be seen as a feature selection task. In feature selection, it is assumed that many features (attributes) in a dataset are either irrelevant or redundant for prediction of the outcome variable, and an attempt is made to only select the subset of features that are relevant for prediction (Dash & Liu, 1997). One class of feature selection methods that is preeminently suited for optimizing accuracy and efficiency of clinical assessment are recursive partitioning methods (RPMs). RPMs separate observations in a dataset into subsets for which the distribution of the outcome variable are most different. This is done by recursively splitting the space spanned by all predictor variables into a set of rectangular areas (Strobl, Malley & Tutz, 2009). Because partitions are non-overlapping, they can be visually represented with a tree structure, providing easy-to-use prediction tools for clinical practitioners. RPMs can also be used for subgroup detection: because the partitions are defined by values of predictor variables, every partition defines a specific subgroup of respondents. Detection of subgroups with, for example, a high risk for developing a mental disorder, or a greater expected benefit of psychotherapy than of pharmacotherapy, is of great importance for decision making in clinical practice.

In the first chapters of this thesis, I will explore the use of several RPMs for prediction of clinical outcomes. Application of RPMs may provide results that are more efficient and accurate for clinical assessment and decision making, compared to many standard data-analytic approaches applied in clinical research.

The fidelity of clinical assessment, however, does not only depend on the data-analytic method used, but also on the reliability and validity of outcome variables. Biased measurement of mental health constructs will likely hamper the accuracy of decision-making in clinical practice. One of the measurement biases that may occur in clinical research and practice is response shift bias (Oort, 2005). Response shift bias can occur when change is assessed by means of total scores on self-report inventories. These total scores may reflect true change (change in a patient's level on the target construct), but they can also be confounded by response shifts (changes in patient's values or standards for measurement). Before treatment effects in clinical trails can be assessed using item
Introduction

responses on self-report inventories, it should be tested whether changes in these responses are confounded by response shift bias. Therefore, in the last chapter of this thesis I will focus on the detection of response shift bias in the assessment of treatment effects of mental health interventions.

Outline

The first two chapters of this thesis focus on diagnosis in mental health practice, presenting methods to improve efficiency and accuracy when one or several test scores are used for making classification decisions (e.g., in screening or diagnosis of mental disorders).

In Chapter 1, I will discuss the application of curtailment, an algorithm for reducing the number of items administered within a test, to an existing dataset with item responses to several well-known mental health self-report inventories. With curtailment, administration of items within a test is halted when the probability of obtaining one of the possible outcomes (e.g., "at risk" or "not at risk") exceeds a pre-specified threshold. One of the advantages of curtailment is that the aforementioned probability can be determined without specifying a predictive model (e.g., logistic regression), using empirical proportions only. I will evaluate the performance of curtailment by comparing its diagnostic accuracy and assessment-length reduction with that of an adaptive test based on item response theory.

Often in clinical assessment, instead of a single test, a battery of tests may be administered for making classification decisions in clinical assessment. In Chapter 2, I present an algorithm for reducing assessment length of test batteries used for classification, that combines curtailment with the classification and regression trees algorithm (CART; Breiman, Friedman, Olshen & Stone, 1984). The new algorithm creates a classification tree, which determines the subset and the order of tests to be administered, and provides the cut-off values used for classification. In turn, curtailment is applied in every node of the classification tree. I will apply the combined algorithm to an existing dataset of item responses on
Introduction

a battery of mental-health questionnaires, to assess its diagnostic accuracy and efficiency.

The third and fourth chapters of this thesis focus on prediction in mental health practice, introducing data-analytic methods that provide accurate and user-friendly prediction tools, when the variables relevant for making a decision are not yet known.

In Chapter 3, I will discuss the potential use of rule-based methods, more specifically the RuleFit algorithm (Friedman & Popescu, 2008), for assessment in clinical practice. RuleFit is an algorithm that creates a so-called ensemble of prediction rules. These prediction rules are derived from the nodes of a large number of CART trees, which in turn are grown on subsets of the data. Using sparse regression, RuleFit selects a small subset of prediction rules that improve predictive accuracy of the final ensemble. Because prediction rules are much simpler than the trees from which they are derived, and only a small number of the rules are included in the final ensemble, RuleFit ensembles may provide prediction tools for clinical practice that are easier to apply than the results of traditional actuarial methods, and with similar accuracy.

In Chapter 4, I will focus on the use of RPMs for decisions on treatment allocation. This chapter presents an extension of the model-based recursive partitioning framework (Zeileis, Hothorn & Hornik, 2008) for clustered datasets, aimed at the detection of treatment-subgroup interactions. Treatment-subgroup interactions are subgroups in a dataset, which show differential effects for two or more treatments. Detection of such subgroups is important for personalized medicine, where selection of the optimal treatment is based on a patient’s characteristics. A linear model tree is a model-based recursive partitioning method that is preeminently suited for the detection of treatment-subgroup interactions. However, when datasets have a clustered structure, for example, when observations are clustered within research centers, or within participants in longitudinal datasets, this clustered structure should be taken into account by estimation of random effects. Tree-based methods that allow for estimation of random effects, as well as detection of treatment-subgroup interactions have not yet been developed. Therefore, in Chapter 4, I will introduce the Imertree algorithm, that
combines linear model trees and estimation of random effects, and evaluate it’s performance in detecting treatment-subgroup interactions in simulated datasets with and without cluster-specific effects.

The accurate detection of treatment-subgroup interactions as discussed in Chapter 4, also requires unbiased measurement of treatment outcomes. Therefore, in Chapter 5, I will discuss response-shift bias: bias that may occur in the measurement of change, when treatment outcomes are quantified in terms of total scores on self-report inventories, as is often the case in trials in clinical psychology. I will test whether such bias is present in total scores on a self-report inventory on depressive symptomatology, in a well-known clinical trial comparing the effects of psycho- and pharmacotherapy for depression.

In the general discussion at the end of this thesis, I will summarize the results of the studies presented in this thesis, discuss overarching topics, and point out some directions for future research.