SUMMARY

DISABILITY IN CHRONIC LOW BACK PAIN: PSYCHOMETRIC PROPERTIES OF ADL- AND WORK-RELATED INSTRUMENTS

Chronic low back pain (CLBP) is one of the most common health problems in Western society, and due to the considerable impact on daily functioning, sickness absence and work disability, it is also a heavy financial burden on society. In Rehabilitation Medicine, several treatment programs have been developed to improve participation in society by reducing activity limitations. The main focus is on the limitations and restrictions in the performance of domestic activities in daily living (ADL). Since low back pain is the primary cause of work absence, with huge costs, the impact on functioning at work needs to be considered in rehabilitation, also. Due to differences in area of functioning (at home or at work) both ADL and work-related instruments should be used to assess the effectiveness of treatment.

To obtain evidence about the usefulness of instruments in Rehabilitation Medicine, the psychometric properties of the instruments should be acceptable. Important psychometric properties are reliability and validity. Reliability refers to the extent to which a measurement is consistent and free from errors. The most common measures of reliability are test-retest reliability and inter-observer reliability. Consistency involves the extent to which repeated measures in individuals remain stable over time, under the same circumstances in the absence of treatment. Errors influence the extent of stability over time in the absence of treatment. The validity of an instrument is the ability to measure what it is intended to measure. Besides content-, construct- and criterion-related validity, responsiveness is an important form in measuring treatment effectiveness. Responsiveness refers to the ability to detect clinically important changes in the construct being measured. No clarity exists in literature about responsiveness. Different external criteria are used to determine whether a patient has achieved a clinically important change in functional status, due to the absence of a gold standard to measure clinically important change. It is known that the choice of external criteria influences responsiveness measures.

Several questionnaires have been developed to assess the limitations of CLBP patients to perform ADL activities. The Roland Morris Disability Questionnaire is one of the most frequently used questionnaires in rehabilitation practice. The Dutch language version of the RMDQ (RMDQ-Dv) proved to be valid and responsive; however the reliability was not investigated. Furthermore, the consequences of using different external criteria on the responsiveness of the RMDQ-Dv were not investigated also.
Assessment of work limitations is a relatively new phenomenon in rehabilitation; therefore, work-related instruments have hardly been developed and introduced in daily practice. In other disciplines, several instruments were developed and are used to assess work limitations. In the Netherlands, the Work & Handicap Questionnaire (WHQ) was developed by the Netherlands Organization for Applied Scientific Research (TNO) within the scope of a vocational handicap research program. The Functional Information System (FIS) and the Functional Ability List (FAL) were developed in Dutch Social Insurance Medicine as standardized forms to determine capability to work. Until 2000 the FIS was used, nowadays the FAL is used. The social insurance physicians use the results of history taking and physical examination to fill out the standardised form about limitations in the performance of work-related activities. The ‘Isernhagen Work Systems Functional Capacity Evaluation’ (IWS FCE) is a battery of standardized tests designed to assess a person's functional performance related to work-related activities in a laboratory setting. In these work-related instruments, different assessment perspectives are distinguished for the assessment of work limitations: patient’s self-report, physician’s report based upon history taking and physical examination and performance results based upon functional testing. Several studies showed that the assessed limitations inferred from different perspectives will correlate weakly; the extent of differences between the assessed limitations of the three perspectives, however, has not been investigated.

The aim of this thesis was to investigate the reliability of the mentioned instruments in CLBP patients in Rehabilitation Medicine, to determine the responsiveness of the RMDQ-Dv and to determine the differences in limitations inferred from self-report, clinical examination and functional testing. The main research questions in this thesis were:

**ADL-related instrument**
- What is the reliability and stability of the Dutch language version of the Roland Morris Disability Questionnaire?
- What is the responsiveness of the Dutch language version of the Roland Morris Disability Questionnaire using different external criteria?

**Work-related instruments**
- What is the reliability of the Work & Handicap Questionnaire, the Functional Information System, the Functional Ability List, and the Isernhagen Work Systems Functional Capacity Evaluation?
Summary

- What are the differences in limitations inferred from self-report, clinical examination and functional testing?

In *chapter 1* instruments and research questions were described. In *chapter 2* the test-retest reliability and stability of the RMDQ-Dv for the assessment of ADL limitations in CLBP patients were described. Thirty CLBP filled out the RMDQ-Dv twice with a two-week interval and before starting the rehabilitation program. Intra Class Correlation (ICC) was used as a measure for reliability and the limits of agreement were calculated for quantifying the stability of the RMDQ-Dv. An ICC of 0.75 or more was considered as an acceptable reliability. No criteria for limits of agreement were available. However, smaller limits of agreement indicate more stability because it indicates that the natural variation is small. The RMDQ-Dv showed good reliability, with an ICC of 0.91. A large amount of natural variation of 6 points was found to the total scoring range of 0 to 24 which should be taken into account when interpreting the results of interventions.

In *chapter 3* the responsiveness of the RMDQ-Dv was studied, in particular, the consequences of using different external criteria. The choice of the external criterion determines the accuracy in discriminating improved from non-improved patients. Due to absence of a gold standard to measure clinically important change, four external criteria were used in responsiveness studies: 1. Global perceived effect of change in complaints. 2. Global perceived effect of change in ability to take care of oneself. 3. Change in rating of pain intensity. 4. Smallest Real Difference (SRD). Considerable differences were found in responsiveness statistics, when using different external criteria. However, good responsiveness is found for the RMDQ-Dv for all four external criteria and therefore, it can be concluded that the RMDQ-Dv is able to distinguish between improved and non-improved patient, independent of the used external criterions.

*Chapter 4* presented the test-retest reliability of two self-reports, the WHQ and the FAL. The FAL was in this study used as questionnaire. Thirty patients filled out both questionnaires twice with a two-week interval and before starting the rehabilitation program. A Kappa value of more than 0.60, absolute agreement of more than 80% and ICC of more than 0.75 were considered as acceptable. The test-retest reliability proved to be unacceptable for most items of the WHQ and FAL. An interesting finding was that the reliability of the ADL-items of the WHQ (ICC = 0.82) was considerably higher than the reliability of the work-items (ICC = 0.42 and 0.33). It is not clear why the reliability of the ADL part of the WHQ shows a higher reliability than the work part of the same questionnaire. In future research it is of interest to investigate why patients may be less able to score in a reliable way on work-related questionnaires than on ADL-related questionnaires and in what way it would be possible to improve the reliability of work-related self-reports in such a way that it is also feasible in clinical practice.
Chapter 5 presented the intra- and inter-observer reliability of the FIS and FAL as scoring forms of social insurance physicians. Four registered Dutch social insurance physicians (3 men, 1 woman) enrolled the study, thirty CLBP patients were assessed. To assess the intra-rater reliability a social insurance physician filled out the FIS and FAL twice after history taking and physical examination of CLBP patients with a two-week interval. To assess the inter-rater reliability, two social insurance physicians performed the diagnostic procedure independently and filled out the FIS and FAL. The first physician carried out the history taking and performed the physical examination, and the next day the other physician did the same. Each subject was examined 4 times. A Kappa value of more than 0.60 and a percent absolute agreement of 80% or more were considered to be acceptable. The results showed unacceptable intra- and inter-observer reliability for almost all items of the FIS. The FAL showed better results, however, for a great part of the items, the statistical values were below the criteria for acceptance also.

Chapter 6 presented the test-retest reliability of all 28 of the tests of the IWS FCE. Thirty CLBP patients who were admitted for rehabilitation treatment participated in the study. Two FCE sessions were held with a 2-week interval in between. Acceptable test-retest reliability was demonstrated for 15 of the 28 tests (79%) based on kappa values (≥ 0.60) and percentage of absolute agreement (≥ 80%). Eleven of the 28 tests (61%) showed acceptable test-retest reliability based on ICC values (≥ 0.75).

Test-retest reliability of the non-material handling tests was difficult, due to ceiling- and criterion effects of the tests. In these tests, most patients were not limited in the performance and could meet the test criteria. Despite the acceptable reliability of these tests, a substantial natural variation should be taken into account when interpreting the results of interventions. In chapter 7 it is described that this substantial variation on the IWS FCE tests may in part be attributed to the CLBP patients.

Chapter 8 presented the comparison of work limitations inferred from different perspectives. The FAL was used as a single scoring form to compare the results of the three perspectives. Ninety-two CLBP patients who participated in this study. The work limitations of each patient were filled out on the FAL by the patient, one of the four social insurance physicians (based upon history taking and physical examination) and the IWS FCE evaluator (based upon the IWS FCE results). A Kappa value of more than 0.60 and a percent absolute agreement of 80% and Spearman correlation of 0.75 or more were considered to be acceptable. Comparison of self-report, clinical examination (history taking and physical examination) and IWS FCE for the assessment of work limitations in CLBP patients showed considerable differences in results. Self-reported limitations were considerably larger than those derived from clinical examination and IWS FCE. Additionally, the limitations derived from the clinical examination were larger than those derived from the IWS FCE.
In the general discussion in chapter 9, the main findings and the clinical implications of the research were reviewed while giving advice for the use of the instruments in rehabilitation practice in CLBP patients. In addition, possible directions for further research about the different topics were given.

The strengths and weaknesses of the different instruments were discussed. The RMDQ is a useful instrument in daily practice to assess the limitations in ADL because it is reliable and responsive. However, the natural variation should be taken into account when using it clinically.

The WHQ, FIS and FAL are not useful instruments to assess work limitations because of the unacceptable reliability. It is recommended to improve the properties of the currently used instruments or to develop new instruments and to investigate their psychometric properties before using it in practice. Those disciplines which use these instruments in daily practice should investigate the reliability of these instruments in their daily practice also.

The reliable tests of the IWS FCE can be used to assess the level of physical limitations in the performance of work-related activities in a laboratory setting. The usefulness of the IWS FCE to evaluate effectiveness of treatment should be studied in a responsiveness study, because of its ceiling and criterion effects and the large amounts of natural variation. In those patients, with reported limitations, but good performances on the IWS FCE, the IWS FCE can be useful to make patients aware of their possible own level of functioning.

Because of the differences in work limitations inferred from different perspectives, using a combination of different perspective assessments is recommended in order to obtain a comprehensive picture of the patient’s work limitations in Rehabilitation Medicine.

The translation of laboratory results to daily practice is argued. It is hypothesized that, if a weak correlation exists between laboratory and daily practice, the assessment of limitations in CLBP patients should occur in a broader view to daily practice instead of a laboratory setting. Than a more realistic assessment of patient’s limitations in daily practice and of the influence of personal and environmental factors on performance will be obtained. In treatment of CLBP patients, professionals (physiatrists, company doctors, social insurance doctors, psychologists) should cooperate in the establishment of the relevant limitations, the determination of need for intervention, design and planning of treatment and documenting of treatment effectiveness. One case-manager should be assigned and be responsible for the treatment process.