

Methodological issues like the loss of information and the potentially reduced reliability of longitudinal patient reports in MS research can partially be solved by using proxy respondents. In this thesis we outlined agreement and disagreement between patients and proxy respondents and gave a comprehensive insight in the underlying factors of disagreement. Finally a model was developed to estimate patients' long term outcomes on the physical impact of MS, using proxy scores. Results described in this thesis can be considered as a step towards the use of proxy reported outcomes in MS research and clinical trials.

## **DISCUSSION OF OUR FINDINGS**

### **VALIDATION OF PATIENT REPORTED OUTCOME SCALES FOR PROXY USE**

When using proxy responses, one should be able to rely on the measurement scales to yield the same results for patients and proxies. In this thesis we assessed patient and proxy responses on five PROs: The MSIS-29 physical and psychological scale, the MSWS-12, the GNDS and the MSNQ.

The MSIS-29 was evaluated for proxy use in earlier studies with good results.<sup>1</sup> The MSWS is an objective scale with a construct (walking) that is relatively easy to observe, in which we did not expect problems when asking questions to a proxy respondent in the third phrase (“was he/she limited in his/her ability to walk?”). The GNDS is an interview in which a member of the research team asks questions to a proxy and in which understanding of the questions can be directly checked. With respect to the MSNQ, we found opposing results in the existing literature and therefore we aimed to test the validity and interpretability of the Dutch version of the MSNQ patient and informant form.<sup>2-5</sup>

The results in chapter 2 demonstrated that we could confirm the unidimensionality of the MSNQ. It assesses one underlying construct, which is believed to be neuropsychological competence for both the patient and the informant version. The internal consistency was high. Concerning the construct validity, the results confirmed earlier findings that the patient

version correlates higher with anxiety and depression of the patient than with neuropsychological testing (BRB-N). The informant form on the other hand showed the desired result, higher correlations were found with impaired subtest scores of the patient on the BRB-N compared to anxiety and depression of the patient. Inter rater agreement on scale and item level were moderate, indicating that the patient and informant form were not able to discriminate cognitive impaired patients similarly. Regarding the interpretability, the sensitivity of the patient version was too low to discriminate for cognitive impairment. We could define a cut-off point for the informant version and further analyses showed that when combining both versions, there was no added value of the patient version upon the use of the informant version. Although these results on the MSNQ were in favor of the use of proxy ratings, they were not very promising for the analyses on agreement between patient and proxy in the next chapter.

The population of MS patients in this thesis had a good cognitive status. Only about as few as 10% of our patients could be classified as cognitively impaired, whereas cognitive deficits are described to be present in 40-70% of the MS patients and can often be detected early in the disease.<sup>6-10</sup> This finding indicates that our patient population may not reflect the general population of MS patients or that perhaps others were more liberal in defining cognitive impairment. For most of the work in this thesis it was an advantage to work with cognitively preserved patients, since the assessments that were obtained from them could be considered as reliable. However, with respect to assessment of cognitive dysfunction and the possible influence of cognition on (dis)agreement between patients and proxies, an under-representation of cognitive impaired patients is clearly a disadvantage, making any estimation of the value of scales for cognition less reliable.

## **AGREEMENT BETWEEN PATIENT AND PROXY**

In chapter 3 we described agreement between patients and proxies in a cross sectional study. From the perspective of an earlier study with a relatively small group of patients and proxies with the MSIS-29, we aimed to extend this to other PROs that are frequently used in MS research with larger groups of MS patients and proxies. We used a range of MS scales from

what we hypothesized was ‘more easy to observe’ (walking ability, impact of physical functioning), ‘intermediate difficult to observe’ (overall disability) and ‘difficult to observe’ (psychological impact and functioning).

In chapter 3.1 we could confirm our hypothesis that there was agreement between patients and proxies on PROs measuring physical impact of MS and walking ability of the MS patient. Our data supports previous findings that proxies can rate the disease status of patients in an accurate way,<sup>1</sup> when the focus is on specific and observable disease aspects or (physical) limitations.<sup>11</sup> However, for psychological scales (MSNQ and MSIS Psychological) and overall disability (GNDS) the differences between patients and proxy respondents were large and dependent of scale score.

Although these results were promising for the use of physical MS scales rated by proxy respondents, we should stress that this was an overall impression of mean differences between patients and proxies for five scales on group level, which could not automatically be generalized to individual patient-proxy couples. To achieve that goal, we aimed to gain more insight in the underlying factors of patient-proxy agreement.

## **HOW TO EXPLAIN DIFFERENCES BETWEEN PATIENT AND PROXY?**

We suggested that factors influencing the differences between patients and proxies on MS PROs should be further investigated. The objective in chapter 3.2 was to examine whether MS patients’ cognitive status, physical status and levels of depression and anxiety and their proxies’ health, levels of caregiver strain and depression explained the differences we found between patient and proxy ratings. The results demonstrated that on all PROs except MSNQ, proxies reported more difficulties or impact of MS than the patients themselves. Caregiver strain explained a significant part of the differences, proxies who had a higher level of caregiver strain were more negative than the patients on the disease related complaints and disease impact. It is an important observation that patients’ and proxies ratings on MS PROs diverge as caregiver strain increases. This might indicate that proxies experience stress related to their partners need for care as worsening of the patients’ health or as increasing

impact on their daily lives. Besides caregiver strain, we also observed that patients' anxiety and proxies' depression were explanatory for a substantial amount of the rating differences.

The results according to the differences on the MSNQ were opposite to the other PROs under investigation. In chapter 3.2 different explanations are mentioned, it could be related to starting cognitive problems which cannot be observed by the neuropsychologist or the proxy, but in the meantime really bother patients. Thereby part of the differences between patients and proxies on the MSNQ can be explained by anxiety of patients, which could be a reasonable reaction on incipient cognitive problems or fear for future problems. However, according to the results reported in chapter 2, we interpret MSNQ patient-proxy results with caution.

#### **DO PATIENT AND PROXY AGREE ABOUT CHANGE OVER TIME?**

Although we know from previous studies that we can use PROs to monitor changes in MS over time, only a little is known about how proxies behave in rating the patients' MS status over time.<sup>12-16</sup> Keeping in mind the positive results with patient-proxy measurements on physical scales, as seen in the previous chapters, chapter 4.1 focused on the physical impact of MS and walking ability of the MS patient rated by patients and proxies over a 2-year period. We compared longitudinal changes as assessed by patients and proxies on these scales and investigated which factors could be of influence when finding disagreement about change over time.

The majority of the couples agreed about the change over time, indicating that patient and proxy both scored higher, lower or equal after 2 years in comparison to their baseline scores. However, 44% of the patient-proxy couples disagreed with respect to the change of experienced physical impact of MS and 38% of the couples disagreed about changes in walking ability of the patient over a 2-year period. Despite these contradictions, complete disagreement ('higher' versus 'lower' scores) was rare for both scales (2% for MSIS, 5% for MSWS). In the group in which the patients were more positive on the change than their proxies, a higher age of patients and proxies was found on both scales and we observed a

longer duration of the patient-proxy relationship (MSWS) and a longer disease duration (MSIS physical) for this group. And although these results did only reach statistical significance at 0.10 level, we found increased levels of anxiety, depression and caregiver burden in the group in which proxies were more negative than patients on the change of physical impact of MS in 2 years.

These results are promising for the use of proxy reports in (longitudinal) MS research. However, it must be emphasized that these results are found on scales that assess 'easy to observe' aspects of the disease. Reaching agreement over time for scales that assess 'more difficult to observe' aspects of MS is probably more challenging.

### **A FIRST STEP TOWARDS THE USE OF PROXY MEASUREMENTS IN MS STUDIES**

In all previous studies and chapters we concluded that proxy reports should be interpreted with caution and that proxy responses cannot serve as a direct substitute for patients' responses in (longitudinal) MS studies. However, in daily practice researchers have to deal with the loss of information in longitudinal studies, as we described earlier. Hence, the question keeps coming up: To what extent are proxies able to rate patients' follow up scores when patients are not able to participate in follow up measurements? If we would be able to answer this question, we could also be able to write a clear conclusion of this thesis: Is it worth continuing research on proxy measurements in MS?

In chapter 4.2 we focused on the MSIS physical scale, aiming to evaluate whether (missing) patient scores in longitudinal MS studies on this scale can be estimated using proxy outcomes. We were aware of the fact that a proxy follow up score on itself would not be sufficient, therefore we were interested in the variables that, in addition to the longitudinal proxy scores, contributed to the most accurate estimation of longitudinal patient scores on the physical impact of MS. The results demonstrated that the most accurate prediction of the patient follow up score on the MSIS physical scale was based on a constant value (4.43) and a combination of the proxy follow up score ( $\beta=0.47$ ) and patient baseline score ( $\beta=0.47$ ). Correlations between the real and predicted patient follow up scores were high ( $\geq 0.86$ ), even

when follow up lengths varied. The model was validated in an external cohort of MS patients and proxies with good result.

Although these findings can be considered as a first step towards the use of proxy reported outcomes in future studies, caution must be applied as these findings might not be perfectly convertible to other MS populations and further validation is needed. Moreover it is unknown whether this will be feasible for other PROs.

## **PRESENT POSITION**

The results of this thesis illustrate the feasibility of using assessments obtained from proxies in MS research. It shows the limitations which were in part predictable (easy to observe 'objective' aspects are more easy to assess from a different perspective than hard to observe 'subjective' aspects) and in part unpredictable (caregiver burden causes discrepancies in the assessments between patients and proxies). However, both limitations are meaningful and face valid. In cross sectional settings it is hard to deduce patients scores from proxy assessments. Even when taking into account aspects that may account for differences, agreement on a group level is moderate and on an individual level poor. However, for longitudinal measurements our first results are promising. Factors that may induce differences between patients and proxies might filter out when using changes over time. This was confirmed in an independent sample of MS patients and proxies. Nevertheless, although progress is made in the research reported in this thesis, it is too early to use this approach in intervention studies or clinical practice.

## **FUTURE PERSPECTIVES**

### **CONDITIONS OF PATIENTS AND PROXIES**

We are aware of the limitation in our project that, at baseline, proxies were able to participate in the hospital, during the patient visit, or from home. The conditions at home

cannot be controlled, so there is a possibility that the proxy consulted the patient when completing the questionnaires or that the proxy would have had questions for a member of the research team while completing the PROs. Although we didn't find any significant differences between the baseline groups completing PROs in the hospital or at home, we would advice to keep the conditions equal for the whole study population in future studies.

### **SCREENING OF PROXIES**

In this thesis we assessed not only patients', but also proxies' anxiety and depression. Thereby we used two scales to indicate the strain and burden of partners in taking care of the MS patient, which seemed to be important variables in explaining the differences between patient and proxy reports. Future research might need to go beyond that and include more information on proxy health. Despite the use of the Short Form Health Survey (SF-36), we didn't get an objective impression of proxies' health and cognitive state and we didn't collect information on medication use of proxies. It might be important to take these aspects into account in future studies.

### **CARE FOR PROXIES?**

We found high levels of caregiver strain/burden and depression in MS patients' partners, which had impact on their assessment of the disease of their partner. High level of burden and depression could raise the question whether it shouldn't be more common to also provide care to proxies? If so, the question may be: Who should be responsible for that care? Should that be their general practitioner or the neurologist who takes care of the MS patient? Apart from being an interesting question for future researchers (what is the need for care in partners of MS patients?) if provided this may affect the caregiver burden and the depression. In turn this may affect the (dis)agreement between patients and proxies and thus potentially impact the results of this study.

## **FUTURE USE OF PROXY REPORTS**

The model applied in chapter 4.2 for use of proxy reports regarding the physical impact of MS should be tested in other MS patient populations. In addition, in future research should be evaluated whether the same approach can be successful for other PROs in MS. Based on our own results we are convinced that proxy responses, especially for physical aspects of the disease, do have value. The question remains how to handle the situation where we can use proxy reports on physical (impact) scales but not on psychological or cognitive scales in the future? Will proxy measures than still retain value? Longitudinal studies in MS do not only concern the physical status of MS patients, many researchers are currently focused on the psychological and cognitive aspects of the disease. If they are not able to use proxy responses on these scales, their data collection would still be incomplete. What is the consequence of using proxy ratings for only part of the spectrum? An additional question is whether we can use patient and proxy ratings interchangeably? To answer these questions we advise to continue the collection of both proxy and patient data in larger cohorts with diverse profiles and different follow up periods.

## **CONCLUSION**

This thesis was focused on the evaluation of patient and proxy responses on patient reported outcome scales that are often used in MS research, to optimize the quality of long term data collection. We can conclude that proxies, rating the patients' disease state or disease impact, can be a valuable source of information, especially on PROs measuring physical aspects of MS. For the physical impact of MS we developed a model to estimate missing patient scores in longitudinal studies with good accuracy. However, we also observed that there are some factors causing discrepancies between patients and proxies, especially on (neuro)psychological scales. We found patients' anxiety, some neuropsychological test components and proxies' depression and caregiver burden to be explanatory for differences between patients and proxies. However, we are hopeful that in the future, with good models to estimate long term patient scores, proxy reported outcomes can be a valuable source of information.



## REFERENCES

- 1 van der Linden FA, Kragt JJ, Klein M et al. Psychometric evaluation of the multiple sclerosis impact scale (MSIS-29) for proxy use. *J. Neurol. Neurosurg. Psychiatry* 2005; 76: 1677-81.
- 2 Benedict RH, Munschauer F, Linn R et al. Screening for multiple sclerosis cognitive impairment using a self-administered 15-item questionnaire. *Mult. Scler.* 2003; 9: 95-101.
- 3 Benedict RH, Cox D, Thompson LL et al. Reliable screening for neuropsychological impairment in multiple sclerosis. *Mult. Scler.* 2004; 10: 675-8.
- 4 O'Brien A, Gaudino-Goering E, Shawaryn M et al. Relationship of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) to functional, emotional, and neuropsychological outcomes. *Arch. Clin. Neuropsychol.* 2007; 22: 933-48.
- 5 Vanotti S, Benedict RH, Acion L, Caceres F. Validation of the Multiple Sclerosis Neuropsychological Screening Questionnaire in Argentina. *Mult. Scler.* 2009; 15: 244-50.
- 6 Amato MP, Portaccio E, Goretti B et al. Cognitive impairment in early stages of multiple sclerosis. *Neurol. Sci.* 2010; 31: S211-S214.
- 7 Bobholz JA, Rao SM. Cognitive dysfunction in multiple sclerosis: a review of recent developments. *Curr. Opin. Neurol.* 2003; 16: 283-8.
- 8 Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology* 1991; 41: 685-91.
- 9 Rao SM, Leo GJ, Ellington L et al. Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning. *Neurology* 1991; 41: 692-6.
- 10 Rogers JM, Panegyres PK. Cognitive impairment in multiple sclerosis: evidence-based analysis and recommendations. *J. Clin. Neurosci.* 2007; 14: 919-27.
- 11 Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J. Clin. Epidemiol.* 1992; 45: 743-60.
- 12 McGuigan C, Hutchinson M. Confirming the validity and responsiveness of the Multiple Sclerosis Walking Scale-12 (MSWS-12). *Neurology* 2004; 62: 2103-5.
- 13 Giordano A, Pucci E, Naldi P et al. Responsiveness of patient reported outcome measures in multiple sclerosis relapses: the REMS study. *J. Neurol. Neurosurg. Psychiatry* 2009; 80: 1023-8.
- 14 McGuigan C, Hutchinson M. The multiple sclerosis impact scale (MSIS-29) is a reliable and sensitive measure. *J. Neurol. Neurosurg. Psychiatry* 2004; 75: 266-9.
- 15 Motl RW, McAuley E, Mullen S. Longitudinal measurement invariance of the Multiple Sclerosis Walking Scale-12. *J. Neurol. Sci.* 2011; 305: 75-9.
- 16 van der Linden FA, Kragt JJ, van Bon M et al. Longitudinal proxy measurements in multiple sclerosis: patient-proxy agreement on the impact of MS on daily life over a period of two years. *BMC. Neurol.* 2008; 8: 2.