Chapter 4

The reproducibility of the Canadian Occupational Performance Measure

I.C.J.M. Eyssen
A. Beelen
C. Dedding
M. Cardol
J. Dekker

Abstract

Objective
To assess the reproducibility (reliability and inter-rater agreement) of the client-centred Canadian Occupational Performance Measure (COPM).

Design
The COPM was administered twice, with a mean interval of seven days (SD 1.6, range 4-14), by two different occupational therapists. Data analysis was based on intraclass correlation coefficients, the Bland and Altman method and Cohen's weighted kappas.

Setting
Occupational therapy departments of two university medical centres.

Subjects
Consecutive clients, with various diagnoses, newly referred to the outpatient clinic of two occupational therapy departments, were included. They were all over 18 years of age and perceived limitations in more than one activity of daily life. Complete data on 95 clients were obtained: 31 men and 64 women.

Results
Sixty-six per cent of the activities prioritized at the first assessment were also prioritized at the second assessment. The intraclass correlation coefficients were 0.67 (95% confidence interval (CI) 0.54-0.78) for the mean performance score and 0.69 (95% CI 0.56-0.79) for the mean satisfaction score. The limits of agreement were -2.5 to 2.4 for the mean performance score and -2.7 to 2.7 for the mean satisfaction score. For the separate prioritized problems, the weighted kappas ranged from 0.37 to 0.49.

Conclusions
Inter-rater agreement of the prioritized problems was moderate. The reproducibility of the mean performance and satisfaction scores was moderate, but it was poor for the scores of the separate problems. Therefore, the mean scores should be used for individual assessment.

Introduction

Practising from a client-centred perspective acknowledges that the individual is at the centre of the assessment, intervention and evaluation of a treatment process[1]; a process in which the client actively participates and in which the client's priorities, needs and social context are taken into account. Client-centred practice emphasizes a collaborative partnership between the client and the professional, and may lead to improved client satisfaction, increased adherence to a health service programme and improved functional outcome.[2,3] An outcome measure that is capable of reflecting a client-centred perspective is the Canadian Occupational Performance Measure (COPM).[4]

The COPM is designed as an individualized outcome measure to detect changes in the selfperception of the client's performance and satisfaction over time. It is a semi-structured interview, in which the therapist supports the client in identifying activities that the client wants, needs or is expected to perform, and in which the client experiences problems with performance or satisfaction. The COPM yields the client's most important problems in daily activities and scores for performance and satisfaction of these most important problems in activities.

The COPM uses a structured approach, in that there are specific instructions and methods for administering and scoring.[5] However, the test is not norm-referenced because the theoretical basis of the COPM, the Canadian Model of Occupational Performance (CMOP), conceptualized occupational performance as an individual experience, as the ability to choose, organize and satisfactorily perform meaningful occupations.[5-7]

Although the implementation of the COPM in clinical practice and in research[8-14] is expanding, the clinimetric properties of this measure are not fully known. An important item in evaluating the clinimetric quality of an instrument is its reproducibility. The reproducibility is the extent to which an instrument is free of measurement error.[15] It can be assessed by rating the test-retest reliability that refers to the extent to which the same results are obtained from repeated assessments with the same questionnaire when no change in physical functioning is expected.[16,17] For evaluative instruments, reliability should be established with a measure of agreement.[15] Test-retest reliability of the performance and satisfaction of the COPM appeared to be moderate to high (intraclass correlation coefficient (ICC) 0.63-0.89).[5,18] The intra-rater agreement of the content of the problems appeared to be moderate in stroke patients; 56% of the problems prioritized in the

Chapter 4

first assessment were also prioritized in the second assessment.[19] No information is available on the inter-rater agreement.

The present study addresses the following research questions:

1) Do clients prioritize the same problems when assessed twice by two different therapists (inter-rater agreement of the prioritized problems)?

2) (a) To what extent do the mean performance scores and mean satisfaction scores correspond when clients repeatedly evaluate the same problems (reliability and agreement of the mean performance score and the mean satisfaction score) and (b) to what extent do the performance scores and the satisfaction scores of the separate problems correspond when clients repeatedly evaluate the same problems (agreement of the performance score and the satisfaction score of the separate prioritized problems)?

Methods

Study population

The study population consisted of a consecutive series of newly referred outpatients in the occupational therapy departments of two academic hospitals in Amsterdam, the VU University Medical Centre (VUmc) and the Academic Medical Centre (AMC). During the study period 170 outpatients who fulfilled the inclusion criteria were invited to participate. Inclusion criteria were: age above 18 years, perceived limitations in more than one activity of daily life, and no difficulty in understanding the Dutch language. Clients with various diagnoses were eligible. One hundred and five clients agreed to participate. The medical ethics committees of both medical centres approved the protocol of this study. All participants gave their written informed consent to participate in a study on the reproducibility and validity of the COPM. The results of the validity study were published by Dedding and colleagues.[20]

Assessment

Four assessors, two in each hospital, collected the data. The assessors were occupational therapists who were trained and had at least two years of experience in administering the COPM. The official Dutch translation of the COPM[21] was used. The administration of the COPM consists of five steps. In the first step, the client identifies problems in occupational performance that are important and relevant. In the second step the client determines priorities by rating the importance of each problem on a 10-point scale (from 'not important at all' to 'extremely important').

In the third step the client identifies the five most important problems he or she perceives during daily activities. In the last two steps the client is asked to rate on a 10-point scale the performance (from 'not able to do it at all' to 'able to do it extremely well') and satisfaction (from 'not satisfied at all' to 'extremely satisfied') with regard to each of the five prioritized problems. Mean scores can be calculated for performance and satisfaction.[5]

Procedure

To evaluate the agreement of the content of the problems identified during the COPM interview, each client was interviewed twice with an interval of seven days (SD 1.6, range 4-14). On the first occasion, the first assessor administered the COPM. On the second occasion, the COPM was administered by a different assessor, who was blinded for the results of the first assessment. In addition, the second assessor asked the client to score again the prioritized problems identified during the first assessment, while both the client and the assessor were blinded for the scoring during the first assessment. On each occasion the COPM interview resulted in a maximum of five prioritized problems with performance and satisfaction scores for each prioritized problem. No occupational therapy was given between the two assessments.

Data analysis

Inter-rater agreement of the prioritized occupational performance problems
To establish the consistency of the prioritized problems, the problems identified during the first assessment were compared with the problems identified one week later during the second assessment. Two assessors, who were blinded for each other's ratings, compared the problems. For each client, agreement was expressed as the number of problems prioritized at the second assessment that had also been prioritized at the first assessment, divided by the total number of problems prioritized at the first assessment.

A secondary analysis was performed to calculate the percentage of the number of all problems identified at the second assessment (which may or may not have been given priority) that had been prioritized at the first assessment.

Reliability and agreement of the mean and separate performance and satisfaction scores
The reliability and agreement were examined on the basis of the performance and satisfaction scores of the first assessment and the scores of the assessment after

one week. Mean scores (sum scores divided by the number of identified problems) were calculated for performance and satisfaction.

To test for systematic differences in the performance and satisfaction scores for each of the five prioritized problems, the Wilcoxon signed rank test was used. Reliability of the mean performance and satisfaction scores was assessed by calculating the intraclass correlation coefficients (ICC, one-way random effect model). In addition, the Bland and Altman method[22] was used to assess agreement between the repeated mean scores. Calculations included the mean difference (d), the 95% confidence intervals (CI) of these differences, the standard deviation (SD) of the differences, the reliability coefficient and the 95% limits of agreement. Diagrams were also plotted to illustrate the distribution of results. Cohen's weighted kappa[23,24] was calculated for the separate problem scores.

Results

Study population

Of the 105 clients who met the inclusion criteria, complete data on 95 clients were obtained. The reasons for incomplete data were illness at the time of the second appointment ($n = 3$), scheduling problems ($n = 2$), failure to attend the second appointment ($n = 2$) and the therapist stopped the assessment because it was not possible to formulate or rate the perceived problems ($n = 3$) (e.g., because it was difficult for the client to tell his or her story and to express what the perceived problems were or because the client indicated that he or she did not want the problems inventoried or rated this way).

The characteristics of the study population are presented in Table 4.1. The median duration of the disease/disorder was two years (range 0-61). Seven clients indicated that there had been a change in their therapy or medication between the two assessments.

Inter-rater agreement of the prioritized occupational performance problems

The median number of prioritized problems at the individual assessment was five (range 2-5) at the first assessment and five (range 2-5) at the second assessment. A total of 432 problems were prioritized at the first assessment and 437 at the second assessment. Of the problems that were prioritized at the first assessment, the median (p25, p75) percentage that were also prioritized at the second assessment was 66% (50,80). Of the problems that were prioritized at the first assessment, a

median of 80% (60,100) was also identified in the whole item pool of problems identified at the second assessment.

When the clients who had indicated a change in therapy or medication ($n = 7$) were excluded from the analyses, the agreement improved from 66% to 71%.

Table 4.1 Characteristics of the study population (n = 95) (mean (SD) and range in years)

	n	Mean (SD)	Range
Age		47 (15)	19-80
Gender			
Male / Female	31/64		
Living situation			
Living alone / together with others	31/64		
Cultural background other than Dutch	3		
Therapy/medication change between assessments	7		
Disease duration (years)			
Disorders of wrist, hand and arm (e.g. repetitive strain injury, reflex sympathetic dystrophy, fractures, tendon injuries)	29	4 (8)	0-40
Central neurological disorders (e.g. multiple sclerosis, cerebral injuries)	23	5 (7)	0-21
Neuromuscular diseases (e.g. post-polio syndrome, limb girdle dystrophy)	16	20 (21)	0-61
Other (e.g. diabetes mellitus, arthritis, arthrosis, neoplasm)	27	8 (11)	0-47

Reliability and agreement of the mean and separate performance and satisfaction scores

The ICC for the mean scores for performance and satisfaction were respectively 0.67 (95% CI 0.54-0.78) and 0.69 (95% CI 0.56-0.79). The limits of agreement for the mean values of performance and satisfaction were in the range of respectively -2.5 to 2.4 ($\bar{d}$ -0.05, SD 1.2) and -2.7 to 2.7 ($\bar{d}$ -0.01, SD 1.4). Cohen's weighted kappa for each of the five problems ranged from 0.37 to 0.49 for performance scores and 0.38-0.49 for satisfaction scores. An illustration of the distribution of the mean performance scores and the mean satisfaction scores during the first assessment and the reassessment can be found in Figures 4.1 and 4.2.
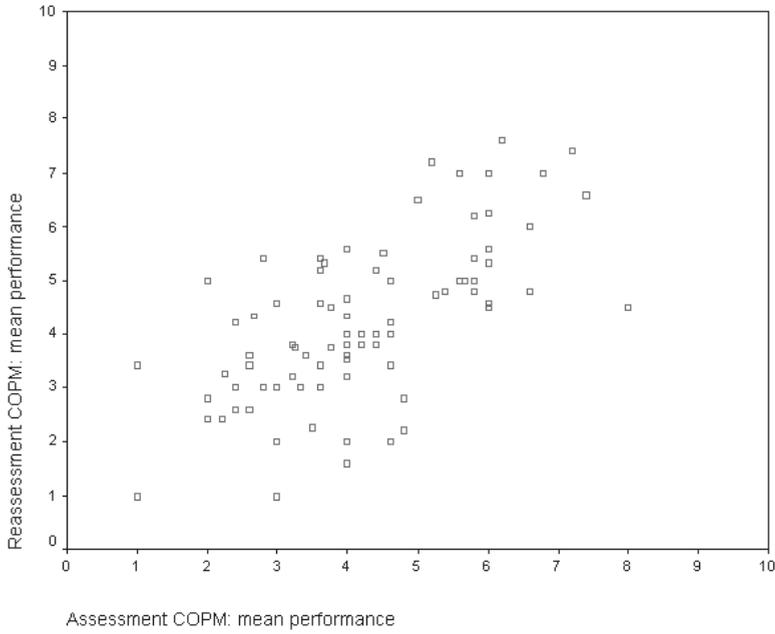
Chapter 4

Figure 4.1 Distribution of the results of the first COPM assessment and the reassessment: mean performance scores
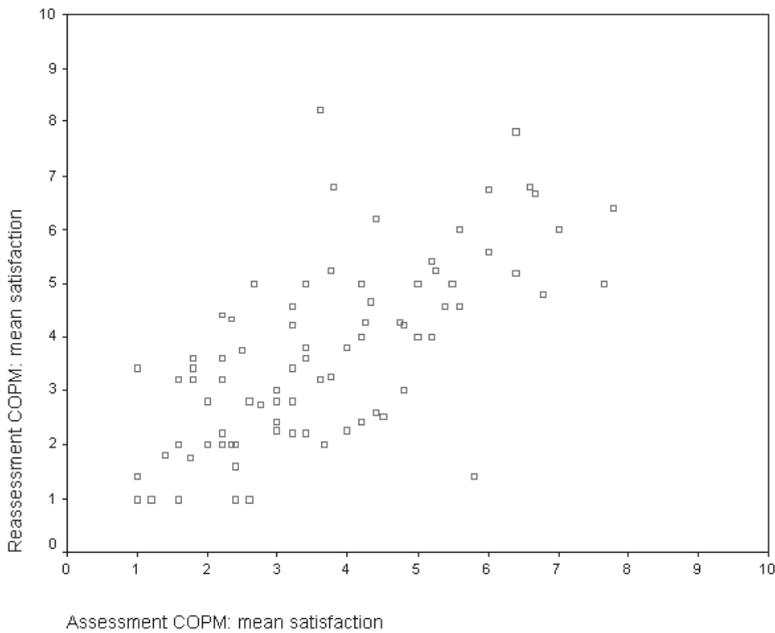

Figure 4.2 Distribution of the results of the first COPM assessment and the reassessment: mean satisfaction scores

Discussion

Agreement in prioritized problems

When clients were interviewed twice by two different therapists, approximately three of the five problems (66%) prioritized at the first assessment were also prioritized at the second assessment. The difference in priority was only to a limited extent due to a change in therapy or medication ($n = 7$). If not only the prioritized activities, but all activities identified at the second assessment are taken into account (including problems in activities which do exist but were not prioritized), the inter-rater agreement is higher (80%). This suggests that at the two assessments the clients mainly reported the same problems, but they rated the importance of these problems in a different way. The difference in prioritizing may be related to the client's perception of the problems. The perceptions and lives of people are not static, but can change; they can vary with time and experience and are modified by coping, adaptation, recent experiences or hope.[25] This may have changed the priority of the problems perceived by the clients in this study.

Another explanation for the moderate inter-rater agreement can be found in the semi-structured design of the COPM. This design is essential in guiding the client-centred assessment process. Although the assessors were trained in administering the COPM, and although there are specific instructions and methods for the assessment[6], the semi-structured design of the COPM may introduce variability in the assessment.[26] This can create inherent differences in the outcome of the identified problems.

One may question whether the inter-rater agreement of the COPM is sufficient. In clinical practice, the therapist can discuss any changes in priorities with the client. This can produce useful information for directing client-centred treatment. However, when the COPM is used as a measure for comparing the nature of clients' prioritized problems in activities assessed by different assessors, one should be aware of the degree of error in this measurement.

Reliability and agreement of performance and satisfaction scores

The reliability of the mean performance scores and mean satisfaction scores for the problems was moderate. The ICC values were in the lower range of the values found in other studies.[5,18,19] Based on the criterion, described by Lee et al.[27] that the lower limit of the 95% CI for the ICCs should be at least 0.75, it can be concluded that in the present study none of the ICCs met this criterion. The ICC indicates how much clients differ from each other, but it gives no indication of the magnitude

Chapter 4

of disagreement between repeated measurements within an individual. Since the COPM focuses on the identification of perceived problems within clients, the analyses were complemented by calculation of the Bland and Altman 95% limits of agreement. The Bland and Altman[22] method reveals the actual difference between measurements within individuals. The 95% limits of agreement indicated a variability of -2.5 to 2.4 points in the mean performance score and -2.7 to 2.7 points in the mean satisfaction score. This implies that for individuals, only large differences (>2.5 to 2.7 points) in mean performance and satisfaction scores (e.g., in response to treatment) can be detected.

In groups of clients the possibility of detecting differences in mean performance or satisfaction scores will be better. For example, based on the standard deviation of the differences in mean performance scores found in the present study (SD 1.2), a mean change in performance score of 0.8 points can be detected with paired observations in a group of 25 clients (alpha of 0.05 and a power of 90%).

Because in clinical practice each of the separate problems and their scores are often analysed, the agreement of the performance and satisfaction scores for each of the prioritized problems was also investigated. Although no systematic differences in scores were found between the two assessments, the agreement, as indicated with Cohen's weighted kappa, was just fair (0.37) to moderate (0.49).[24] Therefore the performance and satisfaction scores for each of the prioritized problems lack the necessary precision for individual client assessment. This is unfortunate, since the client-centred COPM was developed to identify, prioritize and evaluate the client's perceived problems in activities. Nonetheless, the reproducibility of the mean scores of the COPM shows better results than the reproducibility of the scores of each of the prioritized problems. Thus the mean scores should be used instead of the separate scores for each of the prioritized problems.

To a certain degree, the rather low reliability and agreement found in this study may be due to the methodology that was applied. Subsequent to the second interview, the clients again scored performance and satisfaction with regard to the prioritized problems identified during the first interview. These scores can be different, due to the context in which they were obtained. At the first assessment the scores were obtained in a context in which both the client and the assessor knew exactly what was meant by the prioritized problems. But at the second assessment the scores were obtained without the exact context of the first assessment. Due to the first and second interview a process can be induced that makes the client more conscious of how he or she really perceives the prioritized activities.

Variability in scores can also be due to different interpretations of the prioritized problems. This can especially be the case if the description of the problems is less specific. For example, for 'doing the housework' or 'dressing' there can be larger differences in scores than for 'vacuuming' or 'tying laces', since doing the housework and dressing both consist of several activities, whereas vacuuming and tying laces are single activities.

In order to improve the reproducibility of the content and the scores of the prioritized problems, a more distinct process of prioritizing, describing and selecting problems is advocated. However, the semi-structured client-centred design of the COPM should be retained, since this provides a process through which a client's unique values and preferences can be identified.[18,20]

It can be concluded from the results of this study that the inter-rater agreement with regard to the content of the problems and the reproducibility of the mean performance scores and the mean satisfaction scores was moderate. The reproducibility of the performance and satisfaction scores of the separate problems is weak. For individual assessment, the mean scores should therefore be used instead of the scores for each of the prioritized problems. Although the semi-structured design of the COPM is valuable, and should definitely not be lost, it seems advisable to standardize some parts of the COPM; a more distinct way of prioritizing the problems and a more specific manner of formulating the problems in terms of activities may improve the reproducibility of the COPM. However, standardization may not be achieved at the expense of client-centredness. Therefore, further research is needed to determine whether a more distinct method of assessment in the COPM, while preserving its client-centredness, can indeed improve the reproducibility.

Clinical messages

- The semi-structured design of the Canadian Occupational Performance Measure (COPM) is valuable, but may lead to moderate inter-rater agreement.
- For individual assessment, the mean scores should be used instead of the scores for the separate prioritized problems.
- A more distinct way of formulating and prioritizing problems may improve the reproducibility.

Chapter 4

References

1. Sumsion T. A revised occupational therapy definition of client-centred practice. Br J Occup Ther 2000;63:304-309.
2. Law M, Baptiste S, Mills J. Client-centred practice: what does it mean and does it make any difference? Can J Occup Ther 1995;62:250-57.
3. Wressle E, Eeg-Olofsson AM, Marcusson J, Henriksson C. Improved client participation in the rehabilitation process using a client-centred goal formulation structure. J Rehabil Med 2002;34:5-11.
4. Donnelly C, Carswell A. Individualized outcome measures: a review of the literature. Can J Occup Ther 2002;69:84-94.
5. Law M, Baptiste S, Carswell A, McColl MA, Polatajko H, Pollock N. Canadian Occupational Performance Measure, third edition. Ottawa: CAOT Publications ACE, 1998.
6. Sumsion T ed. Client-centred practice in occupational therapy: a guide to implementation. Edinburgh: Churchill Livingstone, 1999.
7. Canadian Association of Occupational Therapists. Enabling occupation: an occupational therapy perspective. Ottawa: CAOT Publications, 1997.
8. Bodiam C. The use of the Canadian Occupational Performance Measure for the assessment of outcome on a neurorehabilitation unit. Br J Occup Ther 1999;62:123-26.
9. Carpenter L, Baker GA, Tyldesley B. The use of the Canadian occupational performance measure as an outcome of a pain management program. Can J Occup Ther 2001;68:16-22.
10. Gilbertson L, Langhorne P. Home-based occupational therapy: stroke patients' satisfaction with occupational performance and service provision. Br J Occup Ther 2000;63:464-68.
11. Richardson J, Law M, Wishart L, Guyatt G. The use of a simulated environment (easy street) to retrain independent living skills in elderly persons: a randomized controlled trial. J Gerontol A Biol Sci Med Sci 2000;55:578-84.
12. Ripat J, Etcheverry E, Cooper J, Tate RB. A comparison of the Canadian Occupational Performance Measure and the Health Assessment Questionnaire. Can J Occup Ther 2001;68:247-53.
13. Samuelsson KA, Tropp H, Gerdle B. Shoulder pain and its consequences in paraplegic spinal cord-injured wheelchair users. Spinal Cord 2004;42:41-46.
14. Walsh DA, Kelly SJ, Johnson PS, Rajkumar S, Bennetts K. Performance problems of patients with chronic low-back pain and the measurement of patient-centered outcome. Spine 2004;29:87-93.
15. Bot SDM, Terwee CB, Van der Windt DAWM, Bouter LM, Dekker J, De Vet HCW. Psychometric evaluation of self-report questionnaires; the development of a checklist. Proceedings of the Second Workshop on Research Methodology. Amsterdam: VU University, 2003:161-68.
16. Lohr KN, Aaronson NK, Alonso J et al. Evaluation quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996;18:979-92.
17. De Vet HC, Bouter L, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 2001;17:479-87.
18. Pan AW, Chung L, Hsin-Hwei G. Reliability and validity of the Canadian Occupational Performance Measure for clients with psychiatric disorders in Taiwan. Occup Ther Int 2003;10:269-77.
19. Cup EHC, Scholte op Reimer WJM, Thijssen MCE, Van Kuyk-Minis MAH. Reliability and validity of the Canadian Occupational Performance Measure in stroke patients. Clin Rehabil 2003;17:402-09.
20. Dedding C, Cardol M, Eyssen ICJM, Dekker J, Beelen A. Validity of the Canadian Occupational Performance Measure, a client-centred outcome measurement. Clin Rehabil 2004;18:660-67.

21. Van Duijn HM, Niezen AA, Cardol M, Corder-Bartels IE, Verkerk GJQ. Canadees meetinstrument voor handelingen en vaardigheden. Amsterdam: AMC/VU, 1999.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;8:307-10.
23. Cohen, JA. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20: 37-46.
24. Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.
25. Cardol M, Beelen A, Van den Bos GA, De Jong BA, De Groot IJ, De Haan RJ. Responsiveness of the Impact on Participation and Autonomy questionnaire. Arch Phys Med Rehabil 2002;83:1524-29.
26. Chan CCH, Lee TMC. Validity of the Canadian Occupational Performance Measure. Occup Ther Int 1997;4:229-47.
27. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. Compute Biol Med 1989;19:61-70.

Chapter 4