Part I

The HORVAN Study, Part I

Chapter 2

Routine Outcome Measures in Adult Psychiatry in The Netherlands: An Evaluation of Measurement Instruments

V.J.A. Buwalda, M.A. Nugter, S.Y. van de Brug, S. Draisma,

J.H. Smit, J.A. Swinkels, W. van Tilburg

Chapter 2 | Routine Outcome Measures

Abstract

Background: An evaluation of the most commonly used ROM measures in Dutch psychiatry is lacking, both for severe mental illnesses and for common psychiatric disorders.

Aim: To provide an overview of the characteristics and quality of outcome measures.

Method: A literature study yielded six outcome measures. The psychometrical, clinical, and practical aspects of these scales are described.

Results: The measures are suitable and are of adequate quality.

Discussion: It remains to be seen whether any of the outcome measures are suitable for both serious and less-serious mental illnesses. The use of a combination of a self-rating scale and an observer-rating scale that measure symptoms and domains of functioning may be the most promising choice.

Introduction

There are no strict guidelines to assist in the selection of measurement instruments for Routine Outcome Monitoring (ROM). Nugter and Buwalda (2012) describe clinical, practical, and scientific criteria.

In this article we discuss the generic ROM measurement instruments that are most frequently used in The Netherlands to measure the course of symptoms and relational and social functioning in adult psychiatry. The two questions we have tried to answer are to what extent these often-used instruments comply with scientific and practical criteria and are they therefore suitable for measuring **change in a patient's condition.**

Scientific criteria

Useful ROM-instruments must meet the minimal criteria of reliability and validity, as described in, e.g., the COTAN (Evers et al., 2010) and Van Beljouw and Verhaak (2010). Recently, scientific criteria have been developed to determine clinically meaningful changes in the severity of mental disorders. On an aggregate level, the effect size (ES) of the score difference is usually used at various points of time in the treatment process. The ES is calculated by dividing the difference in scores by the standard deviation of the baseline (pre-) measurement. An ES indication of 0.2 is deemed small, 0.5 average, and 0.8 large (Sapyta, 2005).

To determine clinically relevant changes at an individual level, Jacobson and **Truax (1991) developed the concepts "reliable change index" (RC) and "clinically significant change" (CSC).** A change is reliable when the difference between the measurements is significant and corrected for measurement errors of the instrument. Such a change does not have to be clinically significant. For a clinically significant improvement, **the patient's** post-measurement score must fall in the range of normal functioning individuals. In order to make well-founded decisions, standard scores of reference groups of healthy individuals, as well as different - more and less severely ill - patient groups, are therefore required.

Discussion of measurement instruments

Six measurement instruments are eligible for the Dutch ROM. These include two observer-rating scales, the *Clinical Global Impression Scale* (CGI) and the *Health of the Nation Outcome Scales* (HoNOS), and four self-rating scales, the *Outcome Rating Scale* (ORS), the *Symptoms Checklist-90* (SCL-90), the *Brief Symptom Index* (BSI), and the *Outcome Questionnaire* (OQ-45)

Chapter 2 | Routine Outcome Measures

In the following sections, we evaluate the psychometric quality of these instruments and we briefly discuss the clinical and practical features.

For the psychometric quality, we describe both the conventional standards of validity and reliability, as well as the extent to which sensitivity to change was examined. Also, we indicate whether the instrument can be used for patients with severe mental illness (SMI) or for patients who do not have SMI (non-SMI).

CGI

General: The most commonly used rating scale is the CGI. The CGI consists of two questions with a seven-point rating scale. The *Severity of Illness Scale* (CGI-S) measures the severity of the disorder during the previous week. The *Global Improvement Scale* (CGI-I) measures the changes in relation to the beginning of the treatment.

Psychometrics: Psychometric research of the CGI in The Netherlands is not known. Kadouri et al. (2007) found intra-class correlation coefficients (ICCs) of 0.64 to 0.94. Berk et al. (2008) calculated difference scores for symptoms improvement by scoring CGI-I directly and using the CGI-S to calculate the difference between scores at admission and discharge. The correlation between the difference in CGI-S and CGI-I score was 0.71 and significant. In terms of change in patients with bipolar, depressive, and psychotic disorders, a significant improvement between admission and discharge was found. The convergent validity was good, including the correlation with the SCL-90 (Kadouri et al., 2007) and the HoNOS (Berk et al., 2008). There are no standard scores or effect sizes reported.

Practical: The rating scale is short (two items), available free of charge, and a Dutch-language version is available. A short training session is sufficient.

Clinical: The CGI provides only two general scores for functioning. The scale is used for both SMI and non-SMI patients.

HoNOS

General: The *Health of the Nation Outcome Scales* (HoNOS) is an observer-rating scale for SMI patients and is rated by trained clinical staff. Twelve items scored on a five-point scale measure four domains: behavioral problems, impairments, symptoms, and social problems (Bebbington et al., 1999).

Psychometrics: Mulder et al. (2004) found an acceptable Cronbach's alpha of 0.78 for the total score and an ICC of 0.92. The convergent validity - coherence with other instruments - was good. In determining discriminant validity, it appeared that total scores significantly differed among patients with different intensities of treatment and diagnoses.

Broersma et al. (2008) and Aartsen et al. (2010) confirm these findings for geriatric psychiatry. Aartsen et al. also determine change through difference scores, but do not test these. Broersma and Sytema (2010) report a high effect size of 0.84 on the **HoNOS 65+ over the total score and on subscale "symptoms," and a moderate ES on the subscale "s**ocial problems."

In an international study of change, Sharma et al. (1999) found an improvement for different diagnoses on the subscales "social problems" and "symptoms" using paired t-tests. Audin et al. (2001) found improvement in outpatients, demonstrated by a good ES (0.69) and CSC in 24% of outpatients, mainly concerning three items (depression, other problems, and social contacts). They conclude that the HoNOS is less applicable to non-SMI. Buwalda et al. (2011), on the other hand, did find the HoNOS useful to non-SMI outpatients. Is seems that the CSC method is rather strict, resulting in a low percentage of patients with a clinically significant change.

Practical: The HoNOS is short. There are no costs involved, a validated Dutch-language version is available, but training is necessary.

Clinical: The instrument includes a variety of dimensions that are relevant to treatment of both SMI and non-SMI patients.

ORS

General: The *Outcome Rating Scale* (ORS) is a self-rating scale consisting of four items with a visual analog scale, scored from 0-10 (Miller et al., 2003). The items examine individual, relational, social and overall functioning. These domains are derived from the OQ-45 (Lambert et al., 1996). The ORS is meant to be completed after each consultation.

Psychometrics: In The Netherlands, Beljouw et al. (2010) studied the use of the ORS among general psychologists. They found an alpha of 0.82 and moderate correlations with the SCL-90. Miller et al. (2003) and Bringhurst et al. (2006) found higher alpha values. Hafkenscheid et al. (2010) studied change using the CSC method and found that with the ORS changes can be made visible.

Practical: The ORS is short and two versions are available. Scott **Miller's translated** version is not available free of charge. A Dutch-language version is available.

Clinical: The ORS is generally used in psychotherapy but does not refer explicitly to psychiatric disorders.

SCL-90

General: The *Symptom Checklist-90* (SCL-90; Derogatis et al., 1973; Arrindell & Ettema, 1986, 2003) is a multidimensional self-rating scale of 90 items with a five-point scale for the main psychopathological symptom areas. The SCL-90 is widely used in The Netherlands. The Dutch version measures eight symptom dimensions: mistrust and interpersonal sensitivity, somatization, depression, anxiety, agoraphobia, anger-hostility, inadequate thought and action, and sleeping disorders.

Psychometrics: **Cronbach's alphas for the sub**-scales vary from 0.73-0.97 with test-retest coefficients of 0.62-0.91 (Arrindell & Ettema, 2003). The authors also describe norm groups. They find that the effects of short-term as well as long-term treatments can be measured with the SCL-90 and describe studies where change is measured using paired t-tests. Buwalda et al. (2011) found clinically significant improvement in one in four outpatients.

Practical: Completion takes about 20 minutes; each assessment with the SCL-90 **costs €1.66 (about $2.27).**

Clinical: The SCL-90 covers several areas of psychiatric disorders and is mainly used for non-SMI.

BSI

General and psychometrics: The *Brief Symptom Index* (BSI; De Beurs & Zitman, 2006), a self-rating scale, is a derivative of the SCL-90 with 53 items and has comparable psychometric properties (alpha higher than 0.71 and good correlations with other instruments).

Practical: Testing with the BSI takes less time than with the SCL-90, and an **assessment sheet costs €1.75 (about $2.40). A validated Dutch**-language version is available.

Clinical: The BSI covers several areas and is normally used for non-SMI patients.

OQ-45

General: The *Outcome Questionnaire-45* (OQ-45; Lambert et al., 2004), a self-rating scale, consists of 45 questions scored on a five-point scale and measures three domains: symptom distress, interpersonal relations, and social role performance.

Psychometrics: In The Netherlands, De Beurs et al. (2005), De Jong and Nugter (2004), and De Jong et al. (2007) studied the OQ. Cronbach's alphas varied on sub-scales and total scores from 0.68-0.95. Test-retest reliability varies from 0.68-0.83. Also, correlations higher than 0.75 were found for coherence with other instruments. De Jong et al. (2007) and De Beurs et al. (2005) found a good discriminant validity and yielded standard scores for individuals without symptoms. They also found changes after a short treatment on the total score, with a t-test and an effect size of 1.33. Buwalda et al. (2011) found clinically significant improvement in one in three outpatients.

Practical: The OQ can be completed in 10 minutes; the cost per patient is €0.73 (around $0.98), independent of the number of assessments per patient. A validated Dutch-language version is available.

Clinical: Several areas are measured. There has been no study on SMI patients in The Netherlands yet.

Discussion

The six outcome measures discussed in this paper together cover the Axis-I spectrum of disorders of the DSM-IV-TR classification; based on their sufficient reliability and validity in the national and international literature, they are eligible for use for ROM in The Netherlands. The self-rating scales SCL-90 and BSI purport to have a wide application domain, but in practice these measurement instruments are seldom used for SMI patients (such as those with psychotic disorders). In addition, they are rather expensive to use. The OQ-45 is also used for a milder category of patients and is not free of charge. The ORS and CGI are not (yet) sufficiently studied in The Netherlands. The HoNOS is free of charge, but training is required for its use. The observer-rating scales for SMI patients are available free of charge, the self-rating scales for the non-SMI group are not.

A combination of the HoNOS and the OQ (respectively, observer-rating and self-rating) would not only cover the widest spectrum of disorders, but would also

provide insight into complaints as well as social and interpersonal functioning of the patient.

Perhaps it is also the cheapest combination. Validation research on each instrument has been published. Research into the validation of the combination of the HoNOS and OQ-45 is published (Buwalda et al., 2011) and the results seem to support this combination.

References

Aartsen MT, Spitsbaard AK, van Baarsen C, Dhondt ADF, Mascini M, Nefs A, e.a. Een multicenterstudie naar betrouwbaarheid, validiteit en gevoeligheid voor verandering van de HoNOS65+ binnen de ouderenpsychiatrie. Tijdschr Psychiatri 2010; 53: 543-553.

Arindell WA, Ettema JHM. SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator. Lisse: Swets & Zeitlinger 1986; 2003.

Audin K, Margison FR, Clark JM, Barkham M. Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment service. Br J Psychiatry 2001; 178: 561-566.

Bebbington P, Brugha T, Hill T, Marsden L, Window S. Validation of the Health of the Nation Outcome Scales. Br J Psychiatr 1999; 174: 389-394.

Beljouw IMJ van, Verhaak PFM. Geschikte uitkomstmaten voor routinematige registratie door eerstelijnspsychologen. Utrecht: NIVEL 2010.

Berk M, Ng F, Dodd S, Callaly T, Campbell S, Bernardo M, e.a. The validity of the CGI Severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. J Eval Clin Pract 2008; 14: 979-983.

Broersma TW, Sytema S. Implementatie van het meetinstrument HoNOS 65+ Onderzoek op een afdeling voor ouderenpsychiatrie. Tijdschr Psychiatri 2008; 50: 77-82.

Broersma TW, Sytema S. De Health of the Nation Outcome Scales als effectmaat in de ouderenpsychiatrie. Tijdschr Gerontol Geriatri 2010; 41: 13-18.

Bringhurst MDL, Watson CW, Miller SD, Duncan BL. The Reliability and Validity of the Outcome Rating Scale: A Replication Study of a Brief Clinical Measure. J Brief Therapy 2006; 5: 23-28.

Buwalda VJA, Draisma S, Smit JH, Swinkels JA, van Tilburg W. Validering van twee meetinstrumenten voor Routine Outcome Monitoring in de psychiatrie: de HORVAN studie. Tijdschr Psychiatri 2011; 53: 715-726.

De Beurs E, Hollander-Gijsman M den, Buwalda VJA, Trijsburg W, Zitman, F. De Outcome Questionnaire (OQ-45): psychodiagnostisch gereedschap. De Psycholoog. 2005; 40: 393-400.

De Beurs E de, Zitman F. De Brief Symptom Inventory (BSI). De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90. Maandblad Geestelijke Volksgezondheid 2006; 61: 120-141.

De Jong K, Nugter MA, Polak MG, Wagenborg JEA, Spinhoven P, Heiser WJ. The Outcome Questionnaire (OQ-45) in a Dutch Population: A Cross-Cultural Validation. Clin Psychol Psychother 2007; 14: 288-301.

De Jong K, Nugter AM. De Outcome Questionnaire: psychometrische kenmerken van de Nederlandse vertaling. Nederlands tijdschrift voor Psychologie 2004; 59: 76-79.

Derogatis LR, Lipman RS, Covi L. SSCL·90: an outpatient psychiatric rating scale-preliminary report. Psychopharmacol Bul 1973; 9: 13-27.

Evers A, Lucassen W, Meijer R, Sijtsema K, COTAN Beoordelingssysteem voor de kwaliteit van tests. Geheel herziene versie. Amsterdam: NIP 2010.

Hafkenscheid A, Duncan BL, Miller SD. The Outcome and Session Rating Scales: A Cross-Cultural Examination of the Psychometric Properties of the Dutch Translation. J Brief Psychother 2010.

Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991; 59: 12-19.

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

Lambert MJ, Burlingame GM, Umpress V, Hansen NB, Vermeersch DA, Clouse GC, and Yanchar SC. The Reliability and Validity of the Outcome Questionnaire. Clin Psychol Psychother 1996; 3(4): 249-258

Lambert MJ, Morton JJ, Hatfield DR et al. Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire) (3rd ed.). Wilmington; American Professional Credentialing Services LLC 2004.

Miller SD, Duncan BL, Brown J, Sparks, JS, Claud DA. The Outcome Rating Scale: A Preliminary Study on the Reliability, Validity and Feasibility of a Brief Visual Analog Measure. J Brief Therapy 2003; 2: 91-100.

Mulder CL, Staring ABP, Loos J, Buwalda VJA, Kuijpers D, Sytema S, e.a. De health of the nation outcome scale (HoNOS) als 'routine outcome assessment'. Tijdschr Psychiatr 2004; 46: 274-284.

Nugter MA, Buwalda VJA. Achtergronden en gebruiksmogelijkheden van ROM in de ggz. Tijdschr Psychiatr 2012; 54: 111-120.

Nugter MA, Buwalda VJA, Dhondt ADF, Draisma S. Using HoNOS in treating Dutch Patients. Tijdschr Psychiatr 2012; 54: 153-159.

Sharma VK, Wilkinson G, Fear S. Health of the nation outcome scales: a case study in general psychiatry. Br J Psychiatr 1999; 174: 395-398.

Sapyta J, Riemer M, Bickman L. Feedback to clinicians: Theory, research, and practice. J of Clin Psychol 2005; 61(2): 145-153.

Dissertation Series