Menopause Rating Scale as Outcome Measure for Hormone Treatment

Juergen Dinger and Lothar A.J. Heinemann, ZEG – Centre for Epidemiology and Health Research Berlin, Germany

Background & Objectives

The Menopause Rating Scale (MRS) was developed in the early 1990s in response to the lack of standardized, validated scales for measuring the severity of symptoms associated with Menopause, and the impact of these symptoms on health-related quality of life (HRQoL). The objective was to develop a scale that would be easy to complete by the women themselves. Designed and standardized as a self-administered scale (patient reported outcome), the MRS:

- assesses symptoms/complaints of aging women under different conditions and in different countries
- evaluates the severity of symptoms over time
- assesses changes during treatment.

The scale is easy to complete and can be evaluated quickly. As is the case with other QoL scales, the test instrument reflects a compromise between clinical utility – which requires a short simple test – and high sensitivity and specificity - which requires a large number of sophisticated questions. Given these limitations, the MRS can be considered a reliable and valid test instrument according to standard psychometric requirements.

This poster presents sensitivity and specificity of the instrument when used for the assessment of the efficacy of hormone replacement therapy provided to meno-pausal women.

International Use

The MRS measures 11 items (complaints or symptoms), each on a scale from 0 (no complaints) to 4 points (very severe symptoms). The 11 items are grouped into three independent dimensions:

- psychological domain
- somato-vegetative domain
- urogenital domain

The MRS is used on a widespread basis. It is available in 22 language versions that are cross-cultural equivalent:

Bulgarian, Chinese, Croatian, English (Australian, South African, UK and US version), Flemish, French (French and Belgium version), German, Indonesian, Polish, Portuguese (Brazilian version), Rumanian, Russian (Russian and Ukrainian version), Spanish (Spanish and Argentinean/Mexican version), Swedish, Turkish, Ukrainian.

Reliability & Validity

Reliability

Reliability of a test instrument refers to the extent to which the instrument yields consistent, stable and uniform results over repeated measurements, i.e. consistency of results and test-retest reliability. All reported reliability parameters were similarly good for countries in Europe, the Americas, and Asia): internal consistency (Cronbach's Alpha) and test-retest correlation coefficients (see www.hqlo.com/content/2/1/45).

Validity

Validity assesses whether a scale really measures what it *intends* to measure. This is a process of accumulating evidence. Many data are available that suggest a good level of validity in numerous respects: similar internal structure of the scale across countries, compatibility of MRS scores among countries, good correlation between the MRS and other scales (SF36, Kupperman Index) – see www.hqlo.com/content/2/1/45.

Conflict of Interest Statement

None declared. Analysis and publication were funded from own sources. The analysis was based on datasets of a large postmarketing study which had been performed by Schering Deutschland GmbH. These datasets were unconditionally given to the authors for methodological analyses.

Results

For many clinicians, validity means that a scale can be successfully used to measure treatment effects under routine clinical conditions. We analyzed pre- and post-treatment data for 9,300 women using the MRS scale. The data came from an open post-marketing study of a product for hormone replacement therapy, and the MRS scale was used as the outcome measure for the treatment effect (before hormone treatment and 6 months after it). Gynaecologists from throughout Germany participated in the study on a voluntary basis. The study data were used to critically evaluate whether the scale is sufficiently sensitive to measure the effect of hormone treatment independently of the severity of complaints. Furthermore, the study data were used to perform a ROC (receiver operating characteristics) analysis of the MRS as a diagnostic tool for treatment effects.

The absolute improvement in symptoms following treatment averaged 9.3 points of the MRS total score. This is equivalent to 36% of the baseline score, and was also similar for all three subscales. The degree of improvement in QoL (i.e. reduction in number of complaints) showed a correlation with the degree of symptom severity at baseline. The number of menopause-related complaints declined during/after treatment, with an average decrease of 36% relative to the baseline score. As can be seen in Figure 1, patients with few complaints before therapy improved by 11%, those with mild complaints at entry by 32%, those with moderate complaints by 44%, and those with severe symptoms by 55% – compared to the baseline score. The scale is even sensitive enough to detect positive treatment effects in women with few complaints.

This can also be demonstrated by a comparison with the normal range for the general population (Table 1). In the general population, the percentage of women with no/few symptoms is much higher than those with severe complaints. By contrast, women eligible for therapy show a much lower proportion of few or mild complaints and a much higher proportion of moderate or severe complaints, as compared to the general population. After 6 months of hormone treatment, however, the proportion of patients with severe symptoms were clearly lower than in the general population. Overall, the MRS scale showed a convincing ability to measure the effect of hormone replacement therapy on quality of life across the full range of severity of menopausal complaints.

The MRS scale was also tested to determine whether it predicts the clinical judgment of the attending physicians. The gynecologists condensed their clinical judgment into two categories - *"successful"* and *"not successful"*. ROC analysis was used to identify the optimal cut-off point for the decrease of the 'total score' that represents successful treatment. Figure 2 shows that cut-off points for the minimal score improvement are in the range between 4 and 7. The exact sensitivity and specificity values for these cut-off points are given in Table 2.

Overall, the sensitivity range of 60 to 76% and the specificity range of 68 to 82% can be considered to be sufficient. Compared to clinical judgment the MRS has a better test-retest reliability and the low negative predictive values (approx. 10%) suggest that it is less susceptible to "therapeutic optimism" (placebo effect).

Based on these results we recommend to use a cut-off point of 7 to establish clinical relevant efficacy of a treatment (uncontrolled or placebo controlled). On the other hand a difference of 4 score points between two efficacious treatment makes a difference for the patient. Therefore, a statistical significant difference of at least 4 score points seems to be a reasonable goal for a direct head to head comparison of two treatments.

Table 1: Effects of hormone replacement therapy on MRS scores

Domain	Severity of symptoms	Proportion of Patients		
		Reference [%]	Baseline [%]	6 months [%]
Total Score	No/few	48	26.4	86.8
	Mild	25	18.2	8.4
	Moderate	20	27.8	4.0
	Severe	8	27.6	0.8
Psychological score	No/few	48	30.7	82.5
	Mild	23	17.0	10.4
	Moderate	20	22.2	5.5
	Severe	9	30.1	1.6
Somatic Score	No/few	53	35.7	93.0
	Mild	24	22.7	5.1
	Moderate	15	25.5	1.6
	Severe	8	16.1	0.3
Urogenital Score	No/few	64	37.3	76.8
	Mild	13	13.3	11.6
	Moderate	16	22.2	8.7
	Severe	7	27.3	2.9

Conclusion

The MRS scale is a suitable test instrument to measure the effect of hormone treatment across the full range of severity of symptoms in menopausal women. Compared to clinical judgement the MRS yield more consistent, stable and uniform results. A score improvement of at least 4 to 7 points is recommended to establish clinical meaningful efficacy.

Table 2: Sensitivity and specificity of potential cut-off points for the assessment of treatment Success

Cut-off Point [score improvement]	Sensitivity [%]	Specificity [%]
≥ 4	76.3	67.9
≥ 5	70.8	73.5
≥ 6	65.5	79.2
≥ 7	60.2	82.3

Fig. 1: Symptom improvement [% of baseline score] vs. baseline symptom status

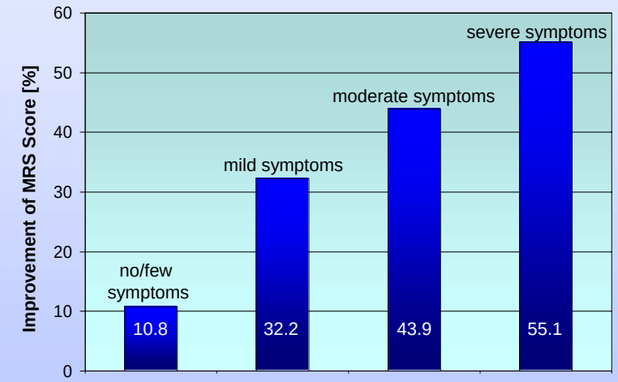

Fig. 2 ROC Analysis: Sensitivity and specificity of the MRS using differnet cut-off points for successful treatment