Chronic pelvic pain and menorrhagia: Assessing treatment effectiveness

Daniels, J.P.

Citation for published version (APA):
Chapter 11

The measurement properties of the menorrhagia multi-attribute quality-of-life scale: a psychometric analysis
Chapter 11

Introduction

Menorrhagia (heavy blood loss during menstruation) is a common gynaecological condition in the UK that has a significant impact on the wellbeing of many women. The number and cost of consultations and treatments impose substantial demands on the NHS. It has been estimated that 6.5% of women aged 12–51 years experience heavy menstrual bleeding (HMB), with 5% of women aged 30–49 years consulting their GP each year because of the condition. Women are mainly treated on the basis of symptoms, and therefore assessment of the condition and the effectiveness of treatment has a large subjective component. There is poor correlation between quasi-objective measures of menorrhagia, e.g. self-reports of blood loss and patient-based health status measures. As the aim of treatment is to improve women’s wellbeing and quality of life (QoL), it is necessary to have valid and reliable instruments to measure this.

A systematic review of published research assessed the quality of QoL instruments in 19 studies of menorrhagia, using a checklist of items for clinical face validity (issues relevant to patients’ expectations and concerns) and items for measurement properties (including reliability, responsiveness, criterion or construct validity and acceptability). The generic Short Form 36 Health Survey Questionnaire (SF-36) was the most commonly used (63% of studies). Only two studies used a condition-specific QoL instrument for menorrhagia, and although 90% complied with more than half of the criteria for measurement properties, only 37% of studies complied with more than half the criteria for face validity. The conclusion of this review, and a review conducted by the UK’s National Institute for Clinical Excellence (NICE), was that there is a need for methodologically sound, condition-specific QoL instruments in menorrhagia with clinical face validity to assess treatment outcomes.

One disease-specific instrument identified as having high face validity is the menorrhagia multi-attribute scale (MMAS) developed by Shaw et al. This was developed in collaboration with women with menorrhagia using multi-attribute utility method. Although the MMAS was designed to have good face and content validity, other psychometric properties of this instrument, i.e. reliability, convergent and discriminant validity, and floor and ceiling effects, could not be evaluated in the original study. One study reported a small advantage for the MMAS over SF-36 in predicting the management outcome for HMB, but otherwise the measurement properties of the questionnaire when used in a treatment or observational study have not been reported. The aim of this study was to assess these properties of the MMAS when used within a treatment trial.
Methods

Participants
All women who took part in this study were enrolled in the ECLIPSE (Effectiveness and Cost-effectiveness of Levonorgestrel-containing Intrauterine system in Primary care against Standard trEatment for menorrhagia) trial, a randomised controlled trial designed to assess the effectiveness of a levonorgestrel-containing intrauterine system in the treatment of menorrhagia [National Institute for Health Research (NIHR) Health Technology Assessment Programme (HTA) project reference 02/06/02]. In order to be invited into the trial, women had to be between the ages of 25 and 50 years, presenting to GPs with menorrhagia, not intending to become pregnant in the next 5 years and agreeing to be randomised. Data from the first 431 women enrolled in the trial were analysed for this study.

Procedure
The ECLIPSE trial, including the study reported here, received multicentre research ethics committee approval. Participants completed the MMAS and a battery of other tests as part of the baseline measures for the trial after randomisation, but before treatment. The other tests used were the SF-36 v2, EuroQol EQ-5D and the Sexual Activity Questionnaire. Fifty women were also invited to participate in a test–retest study of the MMAS, and complete the test again within 2 weeks of the baseline measurement.

MMAS
The MMAS questionnaire captures the subjective consequences of menorrhagia on six domains: practical difficulties; social life; psychological wellbeing; physical health; work routine; and family life. Each of the six domains has four statements that represent four levels of response. Respondents indicate the statement that best matches their feelings for each domain. The statement scores derive from a weighting of the domains and a weighting of the statements in level of severity by women in the original study. Scores range from 0 (worst possible state in all domains) to 100 (best possible state in all domains).

Instruments used in the validation of the MMAS
1 - Short Form 36. The SF-36 v2 is a 36-item short-form survey that measures general health-related QoL. The SF-36 is a practical and reliable way to obtain important health outcomes data in a variety of settings, measuring eight domains of health: physical functioning; role limitations because of physical health; bodily pain; general health perceptions; vitality; social functioning; role limitations because of emotional problems; and mental health. Respondents are asked to recall how they have felt over the previous 4 weeks. It is commonly used in studies of menorrhagia.⁵
2 - EuroQol EQ-5D. EQ-5D is a standardised instrument for use as a measure of health outcome, and is widely used in economic evaluations of medical interventions. Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status.

3 - The Sexual Activity Questionnaire. Sexual activity is an important dimension of QoL, which may be specifically affected by menorrhagia. The Sexual Activity Questionnaire (SAQ) was developed as a self-report questionnaire for use in gynaecological clinical trials, designed to be quick to complete and acceptable to the majority of women.6 Three dimensions of perceptions of sexual activity are measured: pleasure; discomfort; habit.

Psychometric analysis
The measurement properties of the MMAS were assessed using traditional psychometric procedures. Estimates of reliability and validity were made using intra- and inter-test correlations, with an additional comparison of means for the subscales of the questionnaire. Inspection of the data supported the use of Pearson’s product-moment correlation coefficient (r) for bivariate correlations and analysis of variance (anova) for comparison of means. Floor and ceiling effects were also assessed using frequency data on high and low scores.

Results
The mean age of the participants was 41.25 years (SD 5.3 years). The ethnic mix of the sample was representative of the patient population for the area, with 93% of women categorising themselves as from the two main ethnic groups: white British (83%) and South Asian (10%).

Reliability
Forty-nine of the 50 women asked to take the questionnaire again within 2 weeks returned completed questionnaires. The test-retest correlation for the total MMAS score was acceptable (r = 0.836), suggesting that it is stable over time. Cronbach’s alpha was calculated as a measure of the internal consistency of the six items of the MMAS. This was 0.82, which, being within the range 0.7 and 0.9, suggests that the items are all measuring aspects of the same construct. Evidence of the homogeneity of the MMAS was also provided by the item–total correlations (the correlation between each item and the total score minus that item), which range between 0.44 and 0.68. Only one such correlation fell below 0.6: that for ‘practical difficulties’ caused by menorrhagia.
Validity

The MMAS has already been shown to have good face validity, so this analysis concentrated on construct validity: i.e. the extent to which the instrument measures what it is intended to measure. An important method of demonstrating construct validity is by showing that an instrument correlates with other instruments measuring similar constructs, e.g. generic QoL scales (convergent validity), and does not correlate with measures of different constructs, e.g. age or body mass index (discriminant validity).

Convergent validity

To show convergent validity, correlations between test scores should be moderate to high: i.e. equal to or above 0.4. Table 1 shows the correlations between the MMAS total score, the subscales of the SF-36 and the EQ-5D (note that whereas all participants completed all items of the MMAS at baseline, some missed items in the SF-36). These show moderate correlations for most components of the SF-36, but weaker correlations for physical functioning, general health perception and for the EQ-5D standardised score.

Table 1. Correlations between the MMAS total score and the EQ-5D summary and SF-36 subscales.

<table>
<thead>
<tr>
<th>MMAS total</th>
<th>r</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36 subscales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>0.33</td>
<td>420</td>
</tr>
<tr>
<td>Physical role limitations</td>
<td>0.45</td>
<td>422</td>
</tr>
<tr>
<td>Emotional role limitations</td>
<td>0.49</td>
<td>421</td>
</tr>
<tr>
<td>Social functioning</td>
<td>0.54</td>
<td>427</td>
</tr>
<tr>
<td>Mental health</td>
<td>0.40</td>
<td>426</td>
</tr>
<tr>
<td>Energy/vitality</td>
<td>0.42</td>
<td>426</td>
</tr>
<tr>
<td>Pain</td>
<td>0.38</td>
<td>428</td>
</tr>
<tr>
<td>General health</td>
<td>0.32</td>
<td>419</td>
</tr>
<tr>
<td>EQ-5D summary score</td>
<td>0.32</td>
<td>428</td>
</tr>
</tbody>
</table>

The individual domain scores of the MMAS were compared with the SF-36 subscales. With the exception of the practical difficulties domain on the MMAS, which did not significantly correlate with any SF-36 subscale, the highest correlations were between each of the MMAS domains and
the SF-36 social functioning scale. The correlation between the psychological health scales of the two instruments is 0.39, and that between the MMAS relationships scale and the SF-36 emotional role limitations is 0.47. However, these correlations of subscales should be interpreted with caution as there are only four points on the MMAS subscales. So, to investigate these relationships further, respondents were divided into four groups for each MMAS domain, and their scores on the SF-36 were compared using anova. The results for most subscales and domains were highly significant for each comparison, with high scores on the MMAS corresponding to high scores on the SF-36. The exception again was for comparisons in the practical difficulties domain. Here, three comparisons were not significant: mental health ($F_{3,422} = 1.66; p = 0.175$); pain ($F_{3,424} = 0.62; p = 0.60$); and general health ($F_{3,415} = 2.21; p = 0.09$).

There was a small, non-significant difference in the total MMAS scores of women reporting that they were not engaging in sexual activity on the SAQ and those who were (38.86 and 41.49, respectively). It may be that the MMAS relationship domain reflects a particular problem with sexual relationships. To test this, comparisons were made between the scores on the pleasure, discomfort and habit scales of the SAQ for women scoring at the four different levels of the relationship domain. This revealed a significant effect of level on the discomfort scale ($F_{3,321} = 4.05; p = 0.008$), but insignificant effects of pleasure ($F_{3,321} = 1.71; p = 0.165$) and habit ($F_{3,326} = 2.54; p = 0.057$).

**Discriminant validity**

The MMAS is a condition-specific measure, and so should not be affected by variables that are not directly related to variation within the condition. So the total MMAS score would not be expected to correlate with age or body mass index (BMI). As expected, these correlations are very low: age $r = 0.086$; BMI $r = 0.057$. As well as providing evidence for the validity of the MMAS, it suggests that it can be used with all age groups.

**Acceptability**

All items in the MMAS were completed by all participants, in contrast to the other instruments, e.g. SF-36, where participants missed various sections of the questionnaire. We would suggest that this indicates that the MMAS was acceptable to participants, and that the questions were seen as relevant to their condition.

**Floor and ceiling effects**

In psychometric instruments these have implications for the precision of the instrument and its responsiveness to change, as they reduce the likelihood that the instrument will measure further improvement or deterioration. In this sample five (1.2%) women’s scores were the lowest possible and three (0.7%) were the highest possible. The overall distribution of scores was slightly positively skewed (skew = 0.41; SE = 0.12), which is what might be expected in a sample of
women just about to embark on treatment for menorrhagia. These statistics indicate that there are no floor or ceiling effects.

**Discussion**

This study investigated further the psychometric properties of the MMAS, a menorrhagia-specific QoL instrument. It has shown that in addition to having high face validity, the MMAS has good convergent and discriminant validity and test–retest reliability. It also has high internal consistency, with a Cronbach’s alpha score of 0.82, which is further evidence of reliability and indicates that all the items are measuring different aspects of the same construct.

The overall MMAS showed moderate correlations with most of the subscales of the SF-36. Lower correlations were found for physical functioning and general health perception. Rather than being a negative feature these low correlations may reflect the inappropriateness of the SF-36 for measuring QoL in women with heavy menstrual bleeding. The physical functioning subscale is heavily weighted towards mobility and self-care, which are not likely to be affected by HMB. The general health subscale and the EQ-5D, which were also relatively poorly correlated with the MMAS, may not be precise enough to measure differences in this patient group, especially when women are asked to consider their general condition rather than that pertaining to their menstrual cycle.

Conversely, scores on the practical difficulties subscale of the MMAS were not related to other QoL measures. This domain of the MMAS is very specific to menorrhagia, relating to sanitary protection, flooding, etc., and may pick up specific important issues not measured by more general scales.

All women completed the whole of MMAS, which suggests that it is acceptable to women and seen to be relevant. It could be argued that this sample was more highly motivated than women with HMB more generally, because they had already agreed to take part in the ECLIPSE trial. However, there were missing data in other questionnaires administered at the same time. We have not at this stage been able to assess the sensitivity of the MMAS to changes in treatment. However, the findings of the ECLIPSE trial will enable us to report on this in due course.

**Conclusion**

These results suggest that the MMAS has good measurement properties, and is therefore an appropriate condition-specific instrument to measure the outcome of treatment for HMB.
Acknowledgements
Firstly we would thank all the women who agreed to participate in this study. We thank all staff in primary and secondary care involved in the recruitment of participants to the ECLIPSE study and delivery of baseline questionnaires. We also thank Laura Gerrard, the trial co-ordinator, Lisa Leighton, the trial administrator, and Lee Middleton of the Birmingham Clinical Trials Unit for preparing the data for analysis.
References


