Arch Womens Ment Health (2006) 9: 41–49
DOI 10.1007/s00737-005-0103-y

Archives of
Women's
Mental Health
Printed in Austria

Original contribution

Daily Record of Severity of Problems (DRSP): reliability and validity

J. Endicott[1], **J. Nee**[1], and **W. Harrison**[2]

[1] Department of Research Assessment and Training, New York State Psychiatric Institute, New York, U.S.A.
[2] Pfizer Inc., New York, U.S.A.

Received November 29, 2004; accepted August 22, 2005
Published online September 20, 2005

Summary

The Daily Record of Severity of Problems (DRSP) form was developed to aid in the diagnosis and evaluation of DSM-IV Premenstrual Dysphoric Disorder (PMDD). The reliability and validity of the procedure was tested in two studies. Study A included 27 subjects who ranged from having few or no premenstrual problems to those who met criteria for PMDD. Study B included 243 subjects, all of whom met criteria for PMDD. Individual items and Summary Scores had high test–retest reliability in both studies. Internal consistency of Summary Scores was also high in both studies. Summary Scores had moderate to high correlations with other measures of severity of illness. In addition, items and Summary Scores have been shown to be sensitive to change and to treatment differences in Study B. The DRSP provides sensitive, reliable, and valid measures of the symptoms and impairment criteria for PMDD.

Keywords: Premenstrual Dysphoric Disorder; Daily Record of Severity of Problems; reliability; validity; diagnosis.

Introduction

Daily ratings are essential to confirm a preliminary diagnosis of Premenstrual Dysphoric Disorder (PMDD) as defined by the criteria in the Fourth Edition of the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-IV) (1994). (Fig. 1) In addition, given the evidence that daily ratings often fail to support the retrospective reports of severity of premenstrual problems (Endicott & Halbreich, 1982), they are also needed to assess the results of treatment or other changes over time. The prior experience of our group (Endicott et al., 1986; Schechter et al., 1989) and that of others (Moos et al., 1969) has indicated that women will readily complete daily ratings of the severity of many individual items which describe specific types of symptoms and impairment. Such specificity of ratings can also be of value in determining correlations of premenstrual changes (Halbreich et al., 1986).

The Daily Record of Severity of Problems form (DRSP) was developed to aid clinicians in the assessment of the DSM-IV criteria for PMDD as well as to assess severity of symptoms and impairment at various phases of the menstrual cycle. In order to enhance the assessment of the specific DSM-IV criteria for PMDD, the eleven psychological and physical symptoms of criteria A were described in 21 separate items. An additional three items described specific types of impairment in functioning caused by the symptoms (criterion B). The item content is shown in Table 1. The ratings on the DRSP are to be made daily by the subject throughout her menstrual cycle, on items with 6-point severity scales, ''to indicate the degree to which the problems had been experienced.'' The levels of severity on the DRSP are: 1 – Not at all, 2 – Minimal, 3 – Mild, 4 – Moderate, 5 – Severe, 6 – Extreme. The women are also instructed to indicate the days of ''spotting'' or ''full flow of menses.''

Investigators and clinicians who use the DRSP are urged to examine the patterns of ratings on the individual items. In addition, a Summary Scoring System was developed which includes a Total Score made up of the 21 items tapping the 11 DSM IV symptoms and three

A. In most menstrual cycles during the past year, at least five of the following symptoms (which markedly interfered with functioning) were present for most of the time during the last week of the luteal phase, began to remit within a few days after the onset of the follicular phase, and were absent in the week post-menses, with at least one of the symptoms being either (1), (2), (3) or (4):

   (1) markedly depressed mood, feelings of hopelessness, or self-deprecating thoughts

   (2) marked anxiety, tension, feelings of being "keyed up" or "on edge"

   (3) marked affective lability (e.g., feeling suddenly sad or tearful or increased sensitivity to rejection)

   (4) persistent and marked anger or irritability or increased interpersonal conflicts

   (5) decreased interest in usual activities (e.g., work, school, friends, hobbies)

   (6) subjective sense of difficulty in concentrating

   (7) lethargy, easy fatigability, or marked lack of energy

   (8) marked change in appetite, overeating, or specific food cravings

   (9) hypersomnia or insomnia

   (10) a subjective sense of being overwhelmed or out of control

   (11) other physical symptoms, such as breast tenderness or swelling, headaches, joint or muscle pain, a sensation of "bloating", weight gain

B. The disturbance markedly interferes with work or school or with social activities and relationships with others (e.g., avoidance of social activities, decreased productivity and efficiency at home and at work or school).

C. The disturbance is not an exacerbation of the symptoms of another disorder, such as Major Depressive, Panic, Dysthymic, or Personality Disorders (although it may be superimposed on any of these disorders).

D. Criteria A, B, and C must be confirmed by prospective daily ratings during at least two consecutive symptomatic cycles. (The diagnosis may be made provisionally prior to this confirmation.)

Fig. 1. DSM-IV depressive disorder not otherwise specified: Premenstrual Dysphoric Disorder (Criteria listed in the DSM-IV Appendix)

clinically determined subscales: Depressive Symptoms (felt depressed, felt hopeless, felt worthless or guilty, slept more, trouble sleeping, felt overwhelmed); Physical Symptoms (breast tenderness, bloating, headache, joint or muscle pain): and Anger/Irritability (anger/irritability and conflicts with people).

A DSM-IV PMDD Worksheet was designed to aid clinicians in the systematic, step-wise, application of DSM-IV criteria to the DRSP ratings. As can be seen in Fig. 2, the Worksheet is designed to help clinicians to efficiently assess the DSM-IV criteria. The most common reasons for failure to meet criteria are noted early in the process. This Worksheet can also be used to identify women with PMS as well as those with premenstrual worsening of ongoing symptoms if the "STOP" instructions are ignored.

In applying the PMDD criteria, some investigators prefer to specify the amount of mid-follicular to late-luteal change needed (e.g., 50%, 75%) (Freeman et al., 2000; Yonkers et al., 1997). This further specification is in recognition that the levels of change in severity may vary greatly among a group of women, all of whom meet criteria for PMDD, and that greater change in severity may be desirable for some types of studies.

Subjects

The psychometric characteristics of a procedure may vary considerably depending upon the samples of subjects used as well as the methods of assessment employed. Therefore the reliability and validity of the DRSP were tested using data from two very different sets of subjects and using several methods of assessment.

Study A subjects

The data in Study A was from a sample of 27 women, all of whom applied to participate in one of a series of

Table 1. Item content of the Daily Record of Severity of Problems (DRSP)

1a. Felt depressed, sad, "down," or "blue"
1b. Felt hopeless
1c. Felt worthless, or guilty
2. Felt anxious, tense, "keyed up" or "on edge"
3a. Had mood swings (e.g., suddenly felt sad or tearful)
3b. Was more sensitive to rejection or my feelings were easily hurt
4a. Felt angry, irritable
4b. Had conflicts or problems with people
5. Had less interest in usual activities (e.g., work, school, friends, hobbies)
6. Had difficulty concentrating
7. Felt lethargic, tired, fatigued, or had a lack of energy
8a. Had increased appetite or overate
8b. Had cravings for specific foods
9a. Slept more, took naps, found it hard to get up when intended
9b. Had trouble getting to sleep or staying asleep
10a. Felt overwhelmed or that I could not cope
10b. Felt out of control
11a. Had breast tenderness
11b. Had breast swelling, felt "bloated", or had weight gain
11c. Had headache
11d. Had joint or muscle pain
At work, at school, at home, or in daily routine, at least one of the problems noted above caused reduction of productivity or inefficiency
At least one of the problems noted above interfered with hobbies or social activities (e.g., avoid or do less)
At least one of the problems noted above interfered with relationships with others

ongoing studies. The recruitment notice was for screening and called for women with or without premenstrual problems. All of the women made daily ratings for a minimum of two menstrual cycles. They varied greatly in their types and levels of severity of premenstrual changes. They were also screened to exclude women with any current (within the past year) mental disorder through use of the Structured Clinical Interview for DSM-IV (SCID) (First et al., 1997) and were also free of other current medical disorders. None were taking oral contraceptives or other medications. Their mean age was 35.9 (S.E. 4.2, range 22–44), the mean years of education was 14 (S.E. 1.2, range 12–18), one was Hispanic and the remainder were Caucasian. The subjects signed informed consent for the screening evaluations and the protocol was approved by the New York State Psychiatric Institute Institutional Review Board.

Study B subjects

The Study B sample consisted of 243 women, all of whom had applied for treatment, had met DSM-IV criteria for PMDD for at least two screening cycles as well as one placebo cycle. They also had at least a moderate level of severity for at least two of the late-luteal days in at least one of the three DRSP depressed mood items. In addition the level of severity as reflected in one of the three impairment items had to be of at least moderate severity. These criteria resulted in a relatively homogeneous sample of women who had premenstrual depressive symptoms, were symptomatic and impaired during the late-luteal phase of the cycle (days -5 to -1) and free of clinically significant symptoms during the mid-follicular phase (days $+6$ to $+10$). The degree of change required for a symptom to be counted was at least 75% from the mid-follicular to the late-luteal phase of the cycle. The subjects were also evaluated diagnostically with the SCID and those who had met criteria for any other mental disorder during the past year were excluded. Characteristics of the sample are shown in Table 2. The selection of the sample and the treatment study in which they eventually participated is described in more detail elsewhere. (Yonkers et al., 1997)

Subjects in Study B gave informed consent at each of the 12 participating facilities and the Institutional Review boards at each of the 12 sites approved the protocol.

The analyses from Study B provide data for a much more stringent test of reliability i.e., the reliability with which the DRSP items and summary scores discriminate among a relatively homogeneous group of women, all of whom met criteria for PMDD. Furthermore, data from this sample was used to evaluate the validity of the DRSP measures to assess severity of illness and sensitivity to change over time in women who meet criteria for PMDD as well as the ability of the DRSP to detect differences in treatment responses.

Methods

Statistics: Although most prior studies have assessed the reliability of daily ratings using only a measure of the internal consistency of summary scores (Endicott et al., 1986; Freeman et al., 1996; Steiner et al., 1999), we assessed both the internal consistency and the test–retest reliability of the DRSP measures. Cronbach's (1951) coefficient of internal consistency, alpha, was calculated for each of the four Summary Scores at different phases of the cycle for both studies. Alpha reflects the degree to which different components of a summary score measure a single unidimensional latent construct. The test–retest reliability (or stability over time) of the ratings of the items and Summary Scores was assessed in both studies with an intraclass correlation coefficient of reliability using procedures suitable for situations in which the same set of raters make two sets of ratings as described below (Ebel, 1951). In computing the test–retest intraclass correlations, all within-subject variation was counted as error and the reliability index was based on single absolute

NAME _____ DATE _____ ID#_____

This worksheet is designed to assist a clinician in reviewing daily ratings for a subject whose reports suggest that she will meet criteria for DSM-IV Premenstrual Dysphoric Disorder.

1. During the mid-follicular phase (day 6–10 after onset of menses) does the subject have an average daily symptom rating score greater than 3 (mild) on any of the symptoms, i.e., is there any evidence of an ongoing disorder?

Some clinicians choose to allow increased appetite (8a) for obese patients, or insomnia (9a) for those with good reasons (e.g., infants or ill children), or pain from a physical illness (e.g., 11c and 11d). Note if this is the case for this patient.

_____If has greater than mild symptoms during the mid-follicular phase (and they are not "excused") does not meet criteria **- STOP**

_____If essentially symptom-free or only has "excused symptoms" during the mid-follicular phase, proceed to step 2.

2. During the week prior to menses does the subject have scores of at least 4 (moderate) for at least 2 days on one or more of the items that assess the symptoms of (1) depression, (2) anxiety, (3) affective lability, or (4)anger/irritability?

Depression..... 1a ____, 1b ____, 1c____, _____ _____, If NO, does not meet criteria **- STOP**
Anxiety..... 2 ____, ____, If YES, proceed to step 3.
Lability..... 3a ____, 3b ____,
Anger ..... 4a ____, 4b ____,

3. During the week prior to menses does the subject have scores that reach a level of 4 (moderate) for at least two days on at least FIVE of the symptoms (1a through 11d) listed?

_____ If NO, does not have sufficient symptoms to meet criteria **- STOP**

_____ If YES, has sufficient symptom severity – proceed to step 4.

4. During the week prior to menses does the subject have scores of at least 4 (moderate) for at least 2 days on at least one of the three impairment items?

Work, school, home, daily routine_ _____ If NO, does not meet impairment criteria **- STOP**
Hobbies, social activities _____If YES, meets impairment criteria – proceed to step 5
Relationships with others

5. Does your clinical judgement agree with the assessment of the daily ratings (i.e., does the patient appear to meet criteria for Premenstrual Dysphoric Disorder* during the cycles rated)?

_____If NO, specify reason(s)_____

_____If YES, note if this is the first or second cycle of ratings that have met criteria.

_____ **First Cycle** _____ **Second Cycle**

_____
Name of Clinician

*Under copyright (Developed by Jean Endicott, Ph.D.)

Fig. 2. Worksheet for use of Daily Record of Severity of problems to assess diagnostic criteria for DSM-IV Premenstrual Dysphoric Disorder*

rating values. The kappa coefficient was used to assess the reliability with which the DRSP Worksheet was used to determine if the ratings met DSM-IV criteria for PMDD (Spitzer et al., 1961). Product moment correlations were used to assess the relationship of other concurrent measures of severity to the DRSP measures of severity.

Study A reliability of DRSP ratings: The internal consistency of the four DRSP Summary Scores was calculated for the late-luteal, mid-follicular, and late-luteal minus mid-follicular changes of the two cycles. The DRSP ratings made by the 27 women for the two cycles were used to assess the cycle to cycle test–retest reliability for three sets of scores: (1) the late-luteal phase, (2) the mid-follicular phase and (3) the late-luteal minus mid-follicular change scores. The sums of the ratings for the selected 5 days of the late-luteal and mid-follicular phases were used in the analyses.

Study A reliability of use of DRSP worksheet: The reliability with which clinicians can use the DRSP Worksheet to apply the

Table 2. Study B characteristics of subjects with Premenstrual Dysphoric Disorder (N = 243)

Age	
Mean (±SD)	36.7 (4.9)
Range	23–45
Race	
Caucasian	95%
Black	5%
Level of education	
Graduate school	21%
College graduate	26%
Some college	38%
High school graduate	14%
Marital status	
Married/cohabiting	68%
Separated/divorced	18%
Single	14%
Duration of premenstrual problems	
Mean number of years (±SD)	10.3 (6.5)
Number of pregnancies	
Mean (±SD)	1.5 (1.2)
History of major depressive disorder	27.7%
Late-luteal phase Hamilton Depression Rating Scale	
Mean (SD)	13.32 (5.1)

DSM-IV criteria for PMDD to the ratings was assessed for the two cycles of ratings. Two social workers independently examined the daily ratings of the 27 women and used the DRSP Worksheet to determine whether or not the subject met criteria for PMDD.

Study A concurrent validity: Late-luteal phase ratings on the Hamilton Depression Rating Scale, 21 item version (HDRS) (Hamilton, 1960), were used to assess the validity of the DRSP ratings as measures of severity.

Study B reliability of DRSP ratings: Cronbach's coefficient of internal consistency, alpha, was calculated for each of the four Summary Scores for the late-luteal and mid-follicular phases of the placebo-treatment baseline ratings. Since all 243 subjects in Study B had to meet PMDD criteria for three consecutive cycles, a test of reliability using a cycle to cycle comparison would have been relatively meaningless. The data from this study was used to assess the reliability with which day to day ratings reflected cycle phase differences in severity. This procedure indexed the test–retest stability of measurement over repeated occasions during periods when little or no change was expected (i.e. within specific phases of their cycle). Adjacent days of DRSP ratings were sampled within the different phases of the placebo-treatment baseline cycle. Intraclass correlation coefficients of reliability were calculated for the Summary Scores and individual items for: (1) mid-follicular (based upon day +9 versus day +10); (2) late-luteal (based upon day −1 versus day −2 prior to the onset of menses), (3) late-luteal minus mid-follicular change scores (based upon the two sets of ratings for each phase).

Study B correlations with other concurrent measures of severity: Product moment correlations were calculated between the DRSP Summary Scores, the three DRSP impairment items, and other measures of late-luteal levels of severity. These included the Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960), the Social Adjustment Scales (SAS) (Weissman et al., 1978), and the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) (Endicott et al., 1993). Although the coverage of the content of these measures varies somewhat from that of the DRSP, their total scores were used to assess the validity of the DRSP measures of severity of illness. Data from the placebo-treatment baseline cycle were used for this purpose.

Sensitivity of DRSP summary scores and items to change and differences in response to treatments: The 243 women in Study B were randomly assigned to receive either Sertraline or placebo treatment for three cycles. The ability of the DRSP to detect changes in severity over time, and to differentiate among treatments, is reflected in the analyses reported in Yonkers et al. (1997) and will be summarized here.

Results

Study A

Study A reliability of summary scores and individual items in a heterogeneous sample of women: The cycle

Table 3. Study A. Cycle to cycle Daily Record of Severity of Problems test–retest reliability (N = 27)

	Follicular phase days 6 to 10	Luteal phase days −5 to −1	Change luteal-follicular
Summary scores			
Total score	0.99	0.98	0.98
Depressive symptoms	0.98	0.97	0.97
Physical symptoms	0.98	0.96	0.96
Anger/irritability	0.86	0.97	0.97
Individual items			
Depressed/sad/blue	0.97	0.94	0.94
Hopeless	0.96	0.93	0.93
Worthless/guilty	0.93	0.96	0.95
Anxious/tense/ on edge	0.95	0.94	0.94
Mood swings	0.96	0.96	0.96
Sensitive to rejection	0.96	0.94	0.93
Anger/irritability	0.91	0.97	0.96
Conflict/problems w/people	0.86	0.97	0.97
Less interest	0.98	0.97	0.95
Difficulty concentrating	0.97	0.94	0.93
Lethargic/tired/fatigued	0.96	0.92	0.91
Increased appetite	0.94	0.82	0.96
Crave specific foods	0.92	0.84	0.96
Sleep more	0.96	0.78	0.94
Trouble sleeping	0.96	0.77	0.92
Overwhelmed, can't cope	0.98	0.77	0.96
Out of control	0.98	0.79	0.97
Breast tenderness	0.91	0.90	0.97
Breast swelling/"bloated"	0.97	0.84	0.87
Headache	0.97	0.67	0.96
Joint/muscle pain	0.96	0.86	0.96
Impaired work/daily routine	0.99	0.97	0.96
Impaired hobbies/social	0.98	0.96	0.96
Impaired relationships	0.98	0.98	0.97

Table 4. Study A. Internal consistency (Cronbach's Alpha) for summary scores of the Daily Record of Severity of Problems (N = 27)

	Cycle 1		Cycle 2	
	luteal	follicular	luteal	follicular
Total score	0.96	0.93	0.96	0.95
Depressive symptoms	0.94	0.91	0.94	0.93
Physical symptoms	0.57	0.55	0.50	0.53
Anger/irritability	0.88	0.57	0.86	0.56

to cycle test–retest reliability for the Summary Scores and the individual items for the two phases of the cycle and the change scores for Study A are shown in Table 3. They are extremely high (78 of 84 values are over 0.90).

The internal consistency coefficients of reliability are shown in Table 4. With the exception of the Physical Symptoms Summary Score, the late-luteal values are very high, with the mid-follicular phase scores being somewhat lower.

Study A reliability of use of DRSP worksheet: The two clinicians had perfect agreement regarding the diagnosis of the 27 women (kappa = 1.0). They agreed that 13 of the women met DSM-IV criteria for PMDD, five met symptom criteria for PMS only, while four were found to be essentially free of symptoms or impairment during both cycles. The remaining five women were found to have a mean score of more than level 3 (''mild'') for severity on at least one of the DRSP symptoms or impairment items during the mid-follicular phase. The diagnostic conclusions derived through use of the two cycles of DRSP ratings also agreed with those of the clinicians who had reviewed the ratings at the time of screening for various studies.

Study A correlation with Hamilton Depression Rating Scale total scale: The HDRS Total Score was found to be very highly correlated with both the DRSP Total Score and the DRSP Depressive Symptoms score during the late-luteal phase of both cycles (Table 5). The correlations with the other Summary Scores and the individual DRSP items indicative of impairment in functioning were high. The individual DRSP items with the highest correlation (0.80) was ''felt depressed,'' ''sad,'' ''down'' or ''blue'' and the lowest (0.04) was with ''breast tenderness.''

Table 5. Study A. Correlation of Hamilton Depression Rating Scale total score with summary scores of the Daily Record of Severity of Problems (N = 27)

	Cycle 1	Cycle 2
Total score	0.75	0.75
Depressive symptoms	0.73	0.72
Physical symptoms	0.48[a]	0.49[a]
Anger/irritability	0.75	0.74
Impaired work/daily routine	0.67	0.65
Impaired hobbies/social	0.65	0.68
Impaired relationships	0.65	0.67

[a] $p < 0.05$, all others $p < 0.0002$.

Study B

Study B reliability of summary scores and individual items in women with PMDD: The test–retest reliability indices for the PMDD sample are shown in Table 6. The first column lists the mid-follicular phase values for day 9 versus day 10, the second column lists the late-luteal

Table 6. Study B. Test–retest intraclass correlation coefficients of Daily Record of Severity of Problems summary scores and items (N = 243)*

	Mid-follicular phase day 9 vs. day 10	Late-luteal phase day −2 vs. day −1	Change late-luteal-mid-follicular
Total score	0.73	0.86	0.83
Depressive symptoms	0.71	0.86	0.82
Physical symptoms	0.70	0.83	0.80
Anger/irritability	0.69	0.82	0.78
Depressed/sad/blue	0.55	0.76	0.70
Hopeless	0.64	0.79	0.79
Worthless/guilty	0.67	0.79	0.76
Anxious/tense/ on edge	0.63	0.76	0.70
Mood swings	0.64	0.71	0.70
Sensitive to rejection	0.53	0.78	0.70
Anger/irritability	0.58	0.70	0.65
Conflict/problems w/people	0.54	0.70	0.65
Less interest	0.67	0.81	0.78
Difficulty concentrating	0.66	0.80	0.75
Lethargic/tired/ fatigued	0.60	0.72	0.69
Increased appetite	0.59	0.82	0.75
Crave specific foods	0.75	0.84	0.79
Sleep more	0.60	0.78	0.69
Trouble sleeping	0.54	0.77	0.64
Overwhelmed, can't cope	0.66	0.77	0.72
Out of control	0.70	0.79	0.73
Breast tenderness	0.84	0.90	0.88
Breast swelling/ ''bloated''	0.70	0.84	0.81
Headache	0.44	0.67	0.54
Joint/muscle pain	0.69	0.86	0.79
Impaired work/ daily routine	0.74	0.79	0.76
Impaired hobbies/ social	0.61	0.77	0.70
Impaired relationships	0.60	0.70	0.65

* $p < 0.0001$.

phase values for days −2 and −1 prior to menses onset, and column three lists the values for the two sets of late-luteal minus mid-follicular change scores. For the most part, the test–retest values are within the very good to excellent range, particularly for both the late luteal and change scores. Twelve of the 28 sets of late-luteal scores had values of 0.80 or higher and 27 of the 28 were above 0.70. The only value less than 0.70 was that for Headache. The greatly restricted variability of the mid-follicular phase scores (no subject was to have more than a mild level of severity on any item) resulted in somewhat lower values. In spite of the restricted range of the late luteal minus mid-follicular change scores they were found to have good to excellent test–retest reliability with 21 of the 28 values being equal to or greater than 0.70.

The internal consistency coefficients of reliability for each of the late-luteal phase minus mid-follicular phase change Summary Scores are shown in Table 7. All are very high, with that for Physical Symptoms being somewhat lower than those for the other three.

Study B correlations with other concurrent measures of severity: Table 8 lists the correlations of the total scores of the HDRS, the SAS, and the QLESQ with the DRSP Summary Scores and the three items descriptive of impaired functioning due to the symptoms experienced. The correlations are all within the moderate range, indicating that while there is shared variance for measures of severity within this sample of subjects (all of whom had PMDD), the DRSP measures are not simply redundant with those of the other evaluation procedures commonly used to assess severity.

Table 7. Study B. Internal consistency coefficient of Daily Record of Severity of Problems summary scores for late-luteal minus mid-follicular change scores

	Cronbach's alpha
Total score	0.95
Depressive mood	0.88
Depressive symptoms	0.90
Physical symptoms	0.76
Anger/conflicts	0.90

Table 8. Study B. Late-luteal phase correlation of Daily Record of Severity of Problems summary scores with other measures of severity (N = 243)

	HDRS Total	SAS Total	Q-LES-Q Total*
Total score	0.38	0.45	−0.44
Depressive symptoms	0.38	0.45	−0.44
Physical symptoms	0.37	0.44	−0.44
Anger/irritability	0.36	0.44	−0.44
Impaired work/daily routine	0.34	0.37	−0.34
Impaired hobbies/social	0.39	0.42	−0.39
Impaired relationships	0.36	0.39	−0.36

HDRS: Hamilton Depression Rating Scale.
SAS: Social Adjustment Scales.
Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire.
* Higher scores indicate greater satisfaction and enjoyment.

Study B evaluation of change and ability to detect drug-placebo differences: As noted in greater detail in Yonkers et al. (1997), the 243 randomized women showed significantly greater ($p < 0.001$) improvement on all Summary Scores with Sertraline than with placebo treatment. The DRSP Total Score decreased by 32% by study end point in the Sertraline group and 11% in the placebo group. The Depressive Symptoms Score also decreased 32% in the active treatment group versus 13% in the placebo group. The Physical Symptoms Score decreased 32% and 8% in the Sertraline versus placebo treated women. The Anger/Irritability score decreased 31% and 12% in the active and placebo treated groups, respectively.

Individual DRSP items also showed statistically significant differences in level of improvement, favoring Sertraline in all instances at endpoint on the following items: hopeless ($p < 0.001$), mood swings ($p < 0.001$), anger/Irritability ($p < 0.001$), conflicts ($p < 0.001$), depressed mood ($p < 0.001$), rejection sensitivity ($p < 0.001$), feeling out of control ($p < 0.003$), loss of interest ($p < 0.001$), and food cravings ($p < 0.004$).

Six additional items were significantly superior for Sertraline but lost significance after correction for multiple comparison testing: anxious, overwhelmed, trouble sleeping, guilt, decreased concentration, and lethargy. The following items were not significantly different: joint and muscle pain, increased appetite, oversleeping, breast tenderness, bloating, and headache.

Discussion

It is relatively easy to demonstrate that daily ratings can be used to reliably discriminate levels of severity among a ''mixed'' group of women such as those in Study A, some with PMDD and some with few or no problems. This has been done in studies with different daily rating procedures although there have been only a few reports of cycle to cycle test–retest reliability for daily rating measures (Moos et al., 1969; Block et al., 1997). The results reported for Study A are, for the most part, higher than those reported elsewhere. This may be due in part to the use of clinical interviews to screen out women

with current mental or other medical disorders and the use of consecutive cycles. As reported for other daily rating measures, the internal consistency of most DRSP Summary Scores were also found to be quite high.

A much more stringent test of the reliability of a procedure is to assess the various indices among a homogeneous sample such as that used in Study B. The homogeneity of the Study B sample restricted the possible range of scores during both the late-luteal phase and the mid-follicular phase of the cycle, and thereby the degree of mid-follicular to late-luteal change as well. The sample in Study B is similar to that usually entered into treatment trials, and the results are more relevant for treatment studies or biological studies of women with moderate to severe premenstrual problems.

In spite of the relative homogeneity of the Study B sample, the test–retest reliability values reported are, for the most part, high, and quite satisfactory for the purposes for which the DRSP has been designed: differential diagnosis and evaluation of severity of problems and impairment. The results indicate that the level of agreement of ratings made on consecutive days within specific phases of the menstrual cycle was quite high. Furthermore, the DRSP measures were correlated to a moderate degree with other concurrent measures of severity of premenstrual problems within this homogeneous sample. At the same time, the DRSP obviously made an independent contribution to the evaluation of differences in severity among the patients at baseline, given the level of shared variance with the other measures.

Finally, and very importantly for clinical purposes, the DRSP items and Summary Scores have been shown to be sensitive to change with treatment and to be capable of detecting differential effects of at least two treatment modalities. As expected, the Summary Scores were somewhat more sensitive than the individual items. Other investigators may wish to combine the items into different sets to form Summary Scores (e.g., mood changes or ''atypical'' depressive features). Given the reliability and validity of the individual items, such Summary Scores would be expected to perform equally as well.

Although initially designed to reflect DSM-IV criteria for PMDD, the DRSP can also be used to assess lesser degrees of severity of premenstrual changes or syndromes. It can also be used to track daily levels of severity of symptoms and impairment in patients with other conditions in which premenstrual exacerbation of symptoms and impairment have been noted (Endicott and Halbreich, 1988).

Limitations

Unfortunately, neither of the two studies provides the kind of data suitable for addressing the important issue of how much follicular to luteal phase change is clinically significant and sufficient to warrant a diagnosis. The sample in Study A is much too small for that purpose and the Study B subjects were selected to meet criteria for PMDD using a 75% change in the criterion symptoms in each of three cycles. Furthermore all were at least moderately impaired during each of the three cycles. A study involving women seeking treatment for PMS, daily ratings of the criterion items, and judgments made by clinicians of clinical significance would better address this issue.

The lack of a concurrent measure of irritability/anger is a limitation of both studies. Problems with irritability/anger are among the most frequent complaints of women seeking treatment. The HAM-D 21 was used in both studies as the primary clinician evaluation. The focus of the projects for which Study A was used to screen subjects was on depressive symptoms. Study B involved patients who were to be treated with an antidepressant (Sertraline) or placebo and the subject had to have at least one of the three DRSP depressive mood symptoms. Additional studies of the concurrent validity of the DRSP irritability/anger items would be of value.

Acknowledgments

This work was supported in part by the New York State Department of Mental Health and Pfizer, Inc., New York, New York. The data from Study B were from the investigators in the Sertraline Premenstrual Dysphoric Collaborative Study Group noted in the reference of Yonkers, et al., 1997.

References

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders, Fourth edition. American Psychiatric Association, Washington, DC.

Bloch M, Schmidt PJ, Rubinow DR (1997) Premenstrual syndrome: evidence for symptom stability across cycles. Am J Psychiatry 154: 1741–1746.

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16: 297–334.

Ebel RL (1951) Estimation of the reliability of ratings. Psychometrika 16: 407–424.

Endicott J, Halbreich U (1982) Retrospective report of premenstrual depressive changes: factors affecting confirmation by daily ratings. Psychopharm Bull 18: 109–112.

Endicott J, Halbreich U (1988) Clinical significance of premenstrual dysphoric changes. J Clin Psychiatry 49: 486–489.

Endicott J, Nee J, Cohen J, Halbreich U (1986) Premenstrual changes: Patterns and correlates of daily ratings. J Affect Disord 6: 127–135.

Endicott J, Nee J, Harrison W, Blumenthal R (1993) Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharm Bull 29: 321–326.

First MG, Spitzer RL, Gibbon, Williams JBW (1997) Structured Clinical Interview of DSM-IV Axis I Disorders, Research Version, Patient Edition with Psychotic Screen (SCID-I/D w/Psy SCREEN). Biometrics Research, New York State Psychiatric Institute, New York.

Freeman EW, DeRubeis RJ, Rickels K (1996) Reliability and validity of a daily diary for premenstrual syndrome. Psychiatry Res 65: 97–106.

Freeman EW, Sondheimer SJ, Polansky M, Garcia-Espagna B (2000) Predictors of response to sertraline treatment of severe premenstrual syndromes. J Clin Psychiatry 61: 579–584.

Halbreich U, Endicott J, Goldstein S, Nee J (1986) Premenstrual changes and changes in gonadal hormones. Acta Psychiatr Scand 74: 576–586.

Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg 23: 56–62.

Moos RH, Kopell BS, Melges FT (1969) Fluctuations in symptoms and moods during the menstrual cycle. J Psychosom Res 13: 37–44.

Schechter D, Bachmann GA, Vaitukaitis J, Phillips D, Saperstein D (1989) Premenstrual symptoms: time course of symptom intensity in relation to endocrinologically defined segments of the menstrual cycle. Psychosom Med 51: 173–194.

Spitzer RL, Cohen J, Fleiss JL, Endicott J (1967) Quantification of agreement in psychiatric diagnosis: a new approach. Arch Gen Psychiatry 17: 83–87.

Steiner M, Streiner DL, Steinberg S, Stewart D, Carter D, Berger C, Reid R, Grover D (1999) The measurement of premenstrual mood symptoms. J Affect Disord 53: 269–273.

Weissman MM, Prusoff BA, Thompson WD, Harding PS, Myers JK (1978) Social adjustment by self-report in a community sample and in psychiatric outpatients. J Nerv Ment Dis 166: 317–326.

Yonkers KA, Halbreich U, Freeman E, Brown C, Endicott J, Frank E, Parry B, Pearlstein T, Severino S, Stout A, Stone A, Harrison (1997) For the Sertraline Premenstrual Dysphoric Collaborative Study Group. Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment: A randomized controlled trial. JAMA 278: 983–988.

Correspondence: Jean Endicott, Ph.D., Department of Research Assessment and Training, New York State Psychiatric Institute, Unit 123, 1051 Riverside Drive, New York, NY 10032, U.S.A.