

# Core Symptoms That Discriminate Premenstrual Syndrome

Ellen W. Freeman, Ph.D.,<sup>1,2</sup> Steffanie M. Halberstadt, M.S.,<sup>3</sup> Karl Rickels, M.D.,<sup>2</sup>  
Julie M. Legler, Sc.D.,<sup>4</sup> Hui Lin, M.S.,<sup>5</sup> and Mary D. Sammel, Sc.D.<sup>3</sup>

## Abstract

**Objective:** To identify core symptoms that discriminate premenstrual syndrome (PMS) in prospective daily diary ratings and determine the association of these symptoms with functional impairment.

**Methods:** The study analyzed prospective daily symptom ratings and functional impairment data provided by 1081 women who requested PMS treatment at an academic medical center. The data were obtained before any treatment procedures. A random-split sample design provided separate developmental and validation datasets. Logistic regression was used to identify a reduced set of symptoms that best discriminated PMS. The results were validated in a separate dataset. Optimal cutoff points in the symptom scores were identified for clinical use.

**Results:** Statistical modeling identified 6 symptoms that discriminated PMS and not PMS as well as 17 symptoms in daily diary ratings. The identified core symptoms included anxiety/tension, mood swings, aches, appetite/food cravings, cramps, and decreased interest in activities. The area under the curve (AUC) was 0.84 in both models. The sums of the premenstrual symptom scores also discriminated PMS and not PMS and correctly classified 84%–86% of the cases.

**Conclusions:** Six symptoms rated in daily diaries discriminate between PMS and not PMS among women seeking treatment and are significantly associated with functional impairment. The findings suggest that the burden of daily diaries to confirm PMS can be reduced to a smaller number of symptoms that distinguish the patients who meet this requirement. Results also support the concept that a clinical diagnosis of PMS can be developed around a core symptom group.

## Introduction

MORE THAN 20% OF MENSTRUATING WOMEN experience premenstrual syndrome (PMS) to a degree that warrants clinical treatment,<sup>1,2</sup> yet there is no widely accepted diagnosis of the disorder. There are ongoing efforts to develop uniform diagnostic criteria,<sup>3–6</sup> but primary care clinicians who typically manage PMS patients are faced with assessing a broad array of nonspecific symptoms that have been associated with PMS. Diverse diagnostic guidelines and limited clinician time frequently result in a poor diagnosis, which in turn leads to inadequate or inappropriate treatment of a chronic problem that extends over women's reproductive years.<sup>7</sup>

The numerous symptoms that are linked with PMS are a major impediment to its diagnosis. Considerable evidence shows that many symptoms are entrained to the menstrual cycle,<sup>8</sup> but their broad inclusion in diagnostic assessments produces noise that reduces diagnostic accuracy. Further-

more, the inclusion of many symptoms in daily diaries, which are the primary tool in diagnosing PMS, is a considerable burden to both clinicians and patients. However, it is not known which of the many symptoms that are associated with PMS best discriminate the likelihood of the disorder.

There are appreciable differences among the current diagnostic approaches. The guidelines for PMS offered by the American College of Obstetricians and Gynecologists (ACOG) list 6 affective and 4 somatic symptoms.<sup>9</sup> At least 1 symptom must meet the defined cyclic pattern linked to the menstrual cycle, be confirmed by prospective daily ratings for one or more menstrual cycles, and be associated with identifiable dysfunction or impairment. Whether or not these criteria discriminate a clinically significant condition has not been demonstrated.

The criteria most frequently used in recent studies of premenstrual symptoms are those of the American Psychiatric Association (APA). These criteria define a severe form of PMS termed premenstrual dysphoric disorder (PMDD) and are

Departments of <sup>1</sup>Obstetrics/Gynecology and <sup>2</sup>Psychiatry, <sup>3</sup>Center for Clinical Epidemiology and Statistics, and <sup>5</sup>Center for Research in Reproduction and Women's Health, University of Pennsylvania, Philadelphia, Pennsylvania.

<sup>4</sup>Department of Mathematics, Statistics and Computer Science, St. Olaf College, Northfield, Minnesota.

described in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).<sup>10</sup> PMDD requires 5 of 11 listed symptoms, including at least 1 mood symptom, that are severe premenstrually and remit after menses; severe functional impairment; symptoms not the result of another disorder; and prospective confirmation of the symptoms for at least two consecutive menstrual cycles. These stringent criteria identify a severe, dysphoric form of PMS and are met by approximately 5%–8% of reproductive-age women, a much smaller number than those who seek treatment for PMS.<sup>1,7</sup> In contrast to the specified criteria for PMDD, the World Health Organization's (WHO) *International Classification of Diseases* (ICD-10) broadly classifies premenstrual tension syndrome (PMTS) as a gynecological disorder but lists no symptoms and no specific criteria to guide a diagnosis.<sup>11</sup>

The purpose of this study was to identify a small group of symptoms that discriminate between women who meet prospectively confirmed PMS from those who do not meet this requirement. The study participants were women who believed they had PMS and sought treatment for the symptoms. We hypothesized that a small number of symptoms could predict the likelihood of PMS. We also hypothesized that the identified core symptoms are strongly associated with impairment in family relationships, work, or social activities.

## Materials and Methods

### Sample

We reviewed the study records of all women who enrolled for PMS treatment in three clinical trials that were supported by the National Institutes of Health (NIH) at an academic medical center between 1998 and 2007.<sup>12–14</sup> A total of 1400 women enrolled for PMS treatment. Of these women, 319 women did not have a complete daily diary; 1081 women completed a daily diary for at least 1 month in the untreated screen period and were included in the study. Comparisons of demographic and clinical characteristics between the study sample and the group with no daily diary showed no statistically significant differences. The study was approved by the Institutional Review Board of the University of Pennsylvania.

Characteristics of the participants included ages 18–45 years, regular menstrual cycles of 22–35 days for at least 6 months, and persistent premenstrual symptoms for at least 1 year. At enrollment in the screen period, the participants reported no serious medical problems or current mental disorders. Exclusions at enrollment included major axis 1 psychiatric diagnosis; alcohol or substance abuse within the past year; history of psychosis or bipolar disorder; current use of psychotropic medications or any current prescription, over-the-counter (OTC), herbal, or nonmedical therapies for PMS; pregnancy; breastfeeding; hysterectomy; symptomatic endometriosis; irregular menstrual cycles; and any serious or unstable medical illness.

### Study design

The study sample included all women with a complete daily diary for at least one untreated menstrual cycle ( $n = 1081$ ). A random-split sample design of the data provided a developmental dataset ( $n = 541$ ) and a validation dataset ( $n = 540$ ). Modeling was performed using the developmental dataset, and selected models were fit to the

validation dataset in order to objectively evaluate model performance.

Power calculations before the study assumed that the prevalence of true PMS in the population of women who seek clinical treatment was approximately 37%. Given that we evaluated 17 predictors (Penn Daily Symptom Report [DSR] symptoms), we required a minimum of 170 cases of true PMS based on the rule of 10 events per candidate predictor to insure sufficient data for precise estimates of model specificity.<sup>15</sup> Assuming that 50% of the women would meet criteria for PMS at enrollment and, of these, 25% would be false positive, the calculations indicated that 460 subjects were required in the developmental sample to provide statistical power of 90% with alpha at 0.05. Another 460 subjects were required for the validation sample, for a total of 920 subjects in the study.

### Study variables

**PMS symptom scores.** PMS symptom scores for the analysis were obtained from the Penn DSR, a validated daily diary that lists 17 PMS symptoms.<sup>16</sup> Each symptom was rated daily on a 5-point scale from 0 (none) to 4 (severe): irritability/anger, mood swings, depression, anxiety/tension, feeling out of control, feeling worthless/guilty, decreased interest in usual activities, poor coordination, insomnia, difficulty concentrating/confusion, fatigue, aches, headache, cramps, breast tenderness, swelling/bloating, food cravings/increased appetite.

Premenstrual scores were obtained for each menstrual cycle in the same manner as reported elsewhere.<sup>12–14,17,18</sup> Each symptom score was summed for the 6 days before menses, and the symptom scores were summed for a total premenstrual score.<sup>12–14,19</sup> Postmenstrual scores were calculated by summing each symptom score for cycle days 5–10 (day 1 was the first day of menstrual bleeding) and summing the symptom scores for a total postmenstrual score. Missing days in a symptom score were treated by carrying forward the data from the previous day. Only 11% of the sample had a missing value in the symptom ratings, and the majority of these were missing only 1 day. The second screen cycle DSR was selected for the analysis because it was the first complete and untreated menstrual cycle after enrollment in the study; the DSR in the first screen cycle was used if it was the only available diary.

Functional impairment was rated in each menstrual cycle on a 5-point scale (0–4: none, mild, moderate, a lot, severe) in the domains of work, family life, and social activities. Clinical and demographic descriptors were obtained from the self-report questionnaire completed at enrollment.

**Classification of PMS.** We divided the cases in the developmental sample into two groups, termed PMS and not PMS, to establish a PMS group for the analysis. Because there is no consensus definition of PMS, we employed our previously reported criteria for PMS that have demonstrated reliability and validity in clinical trials.<sup>12–15,20</sup> These criteria were applied to the DSR scores in an untreated menstrual cycle as follows: a total premenstrual DSR score  $>80$ , a total postmenstrual score  $>40$ , a 50% or greater difference between the postmenstrual and premenstrual scores, and a score  $\geq 2$  (moderate to severe) on at least one functional impairment item.

### Statistical analysis

All cases with a complete DSR ( $n = 1081$ ) were divided by a random-split to provide the developmental dataset ( $n = 541$ ) and a validation dataset ( $n = 540$ ). The cases were then identified as having or not having PMS as defined. In the developmental dataset, there were 353 women with PMS (65.2%) and 188 women who did not meet the PMS criteria.

Logistic regression models were used to identify the symptoms that were independently associated with PMS. We first fit the full model of 17 DSR symptoms to provide a basis for comparison with a reduced set of core symptoms. To reduce the total number of symptoms in the 17-symptom model, a backward elimination strategy removed symptoms with large  $p$  values until the remaining symptoms had  $p$  values  $< 0.10$ . To reduce the potential for overfitting the model, a bootstrap resampling procedure of the model selection process was conducted to reduce potential bias.<sup>21</sup> It has been shown that this procedure also provides an effective measure of internal validation for predictive logistic models.<sup>22</sup> The procedure was conducted by drawing 1000 samples with replacement from the development half of the original dataset. In each of these bootstrap samples, the same logistic regression and variable selection methods were repeated as described above. The symptoms with  $p < 0.10$  were identified. The percentage of times each symptom met the criteria to remain in the model was accumulated over the 1000 bootstrap analyses. The results show the 6 items that remained in the model over 50% of the times. The remaining 11 items remained in  $< 50\%$  of the bootstrap models. Further models were then assessed using standard model fit criteria and compared with the fit characteristics of the full 17-symptom model.

We also fit models with summed symptom scores to provide a single unweighted predictor score as a more clinically relevant approach. The models obtained from summed symptom scores were compared to a weighted summary based on logistic model predictions that were obtained from single symptom scores using standard model criteria. External validation was conducted. Models fit to the developmental data were applied to the validation dataset to assess the true performance of the derived prognostic models in the population. Model criteria were computed to determine the degree to which the model based on the developmental data discriminated cases of PMS in the validation data.

Optimal cutoff points for predicting the probability of PMS were computed from the summed symptom score models. The selection of the cutoff point was determined by the maximum percent of subjects correctly classified as having PMS in the developmental dataset. Given a particular cutoff score, sensitivity and specificity were evaluated. The same cutoff points were applied to the validation data, and validation sensitivity and specificity were compared to the corresponding values in the developmental data.

Demographic characteristics were compared between groups with chi-square tests. Symptom scores were initially examined using Pearson correlation coefficients to identify pairwise correlations and Cronbach's alpha coefficient to estimate individual item associations and the overall internal consistency of the symptom scores. Statistical tests were 2-sided, with  $p \leq 0.05$  considered significant, with the exception of the bootstrapping validation procedures as described.

### Results

The mean age in the developmental sample was 33.2 years (standard deviation [SD] 6.86); 78% were employed; 83% had education or training beyond high school; 59% were white, 23% were African American, and 18% were of other or unknown racial status. Comparisons of the demographic variables between the PMS and not PMS groups showed no significant differences, with the exception of race, which was more likely to be white in the PMS group. Comparisons of the variables between the developmental dataset and the validation dataset showed no significant differences.

### PMS symptoms

Table 1 shows the mean premenstrual scores for the 17 DSR symptoms. All symptom scores were significantly higher in the PMS group than in the not PMS group. Affective symptoms (mood swings, anxiety/tension, irritability, feeling out of control) were the most severe symptoms on average and had the largest mean difference between PMS and not PMS groups. Physical symptoms, such as breast tenderness, headaches, and cramps, had the lowest mean scores on average and the least difference between the PMS and not PMS groups. The symptoms had high internal consistency (Cronbach's alpha = 0.95). Pairwise correlations of the symptoms were moderate to high. Among the highest correlations were mood swings, irritability, depression, and hopelessness; each

TABLE 1. DAILY SYMPTOM REPORT SYMPTOM SCORES AND IMPAIRMENT RATINGS IN DEVELOPMENTAL DATASET

| DSR symptom                     | PMS<br>n = 353 | Not PMS<br>n = 188 |
|---------------------------------|----------------|--------------------|
|                                 | Mean (SD)*     | Mean (SD)          |
| Fatigue                         | 13.06 (5.77)   | 7.43 (6.11)        |
| Poor coordination               | 6.32 (6.57)    | 2.73 (4.65)        |
| Out of control/overwhelmed      | 11.89 (6.60)   | 5.41 (6.52)        |
| Guilty                          | 8.71 (7.18)    | 4.18 (6.10)        |
| Headache                        | 6.90 (6.25)    | 3.88 (5.12)        |
| Anxiety/tension/on edge         | 13.80 (5.87)   | 7.14 (6.60)        |
| Aches                           | 8.93 (6.67)    | 4.59 (5.93)        |
| Irritability/persistent anger   | 13.71 (5.86)   | 7.31 (6.32)        |
| Mood swings                     | 13.94 (6.00)   | 6.98 (6.73)        |
| Weight gain                     | 13.02 (6.66)   | 7.29 (6.74)        |
| Food cravings                   | 12.95 (6.47)   | 6.68 (6.29)        |
| No interest in usual activities | 10.53 (6.65)   | 4.43 (6.12)        |
| Cramps                          | 5.63 (6.46)    | 3.51 (5.34)        |
| Sad/depressed/blue              | 11.58 (6.71)   | 6.45 (6.49)        |
| Breast tenderness               | 9.17 (7.44)    | 5.69 (6.32)        |
| Sleep problems                  | 8.80 (7.20)    | 4.39 (6.27)        |
| Difficulty concentrating        | 9.71 (6.72)    | 4.39 (6.11)        |
| Impairment <sup>a</sup>         | n (%)          | n (%)              |
| Family                          | 225 (64)       | 76 (40)**          |
| Work                            | 181 (51)       | 73 (39)**          |
| Social                          | 200 (57)       | 75 (40)**          |

\*Student's  $t$  test  $p < 0.001$  for all symptoms.

<sup>a</sup>Rated 3 or 4.

\*\*Chi-square test  $p < 0.001$ .

\*\*\*Chi-square test  $p < 0.01$ .

PMS, premenstrual syndrome; SD, standard deviation; DSR, daily symptom report.

was correlated with feeling out of control at  $r > 0.70$ . Other high correlations included mood swings with anxiety and irritability, depression with hopelessness, and poor concentration with poor coordination, all at  $r > 0.70$ .

### Symptom predictors of PMS

Six of the 17 daily-recorded symptoms were statistically significant predictors of PMS after adjusting for all other symptoms in the model. The significant predictors of PMS were appetite/food cravings ( $p = 0.003$ ), decreased interest in activities ( $p = 0.013$ ), mood swings ( $p = 0.016$ ), cramps ( $p = 0.010$ ), aches ( $p = 0.051$ ), and increased anxiety/tension ( $p = 0.055$ ). The odds ratios (ORs) for these symptoms (Table 2) indicate that the likelihood of PMS increased 5%–8% with each unit increase in the symptom score after adjustment for all other variables in the model.

The internal validation procedures (bootstrap analysis) were applied to the 17-symptom model, and the same 6 symptoms were identified as statistically significant predictors of PMS. Table 2 shows the 6 significant symptoms as confirmed in the validation analysis: anxiety, aches, mood swings, food cravings, no interest in activities, and cramps (each at  $p < 0.01$ ). These 6 symptoms identified PMS as well as 17 symptoms, as indicated by the model criteria scores, which were nearly identical in the 17-symptom and 6-symptom models (Table 2).

Sadness/depression was not a significant predictor of PMS in either the 17-symptom or the 6-symptom model. However, because depression is a clinically important symptom, we examined a 7-symptom model that retained the depression symptom. The results of the 7-symptom model were nearly identical to the results for the 6-symptom model (Table 2), and inclusion of the depression symptom added nothing further to the discrimination of PMS.

Irritability was reported by most participants regardless of their PMS status and did not discriminate between PMS and not PMS (OR 1.10, confidence interval [CI] 0.94–1.08,  $p = 0.83$ ). Irritability was also highly correlated with mood swings ( $r = 0.84$ ) and anxiety/tension ( $r = 0.84$ ), symptoms that proved to be stronger independent predictors of PMS among these treatment-seeking women.

We then compared the performance of a summary score (the total premenstrual symptom score), rather than the logistic regression of single symptom score, as a more efficient approach for clinical practice. The results showed that the summary premenstrual score for 6 symptoms identified PMS as well as the logistic regression scores for single symptoms that are shown in Table 2 (area under the curve [AUC] = 0.83). The summary premenstrual score for the 17 symptoms also performed as well as the 17-item logistic model for single symptoms (AUC = 0.84). These results indicate that little information was lost by summing the individual symptom

TABLE 2. FULL MODEL AND CORE SYMPTOM MODELS PREDICTING SEVERE PREMENSTRUAL SYNDROME

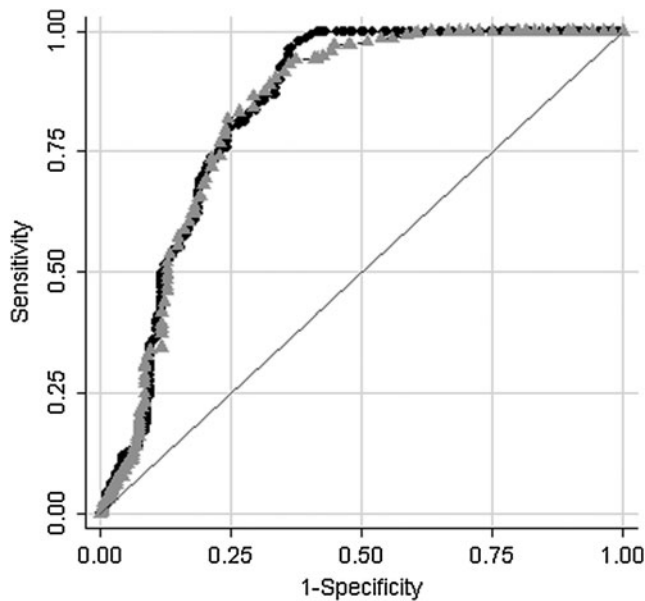
| Method for fitting                           | 17-symptom model    |         | 6-symptom model     |         |
|--|---------------------|---------|---------------------|---------|
|  | Logistic regression |         | Bootstrap stepwise  |         |
| Symptom                                      | Odds ratio (95% CI) | p value | Odds ratio (95% CI) | p value |
| Fatigue                                      | 1.03 (0.97-1.08)    | 0.30    |                     |         |
| Poor coordination                            | 0.95 (0.9-1.02)     | 0.15    |                     |         |
| Out of control                               | 1.03 (0.97-1.10)    | 0.32    |                     |         |
| Guilty                                       | 0.98 (0.92-1.04)    | 0.61    |                     |         |
| Headache                                     | 1.01 (0.96-1.07)    | 0.75    |                     |         |
| Anxiety/tension                              | 1.07 (1.00-1.14)    | 0.06    | 1.07 (1.02-1.12)    | 0.01    |
| Aches  | 1.05 (0.99-1.10)    | 0.05    | 1.06 (1.01-1.10)    | 0.01    |
| Irritability                                 | 1.01 (0.94-1.08)    | 0.83    |                     |         |
| Mood swings                                  | 1.08 (1.01-1.14)    | 0.02    | 1.07 (1.02-1.13)    | 0.01    |
| Weight gain                                  | 1.02 (0.98-1.07)    | 0.47    |                     |         |
| Food cravings                                | 1.07 (1.02-1.12)    | 0.003   | 1.08 (1.04-1.12)    | <0.001  |
| No interest in usual activities              | 1.07 (1.01-1.12)    | 0.01    | 1.06 (1.01-1.11)    | 0.01    |
| Cramps                                       | 0.93 (0.89-0.99)    | 0.01    | 0.93 (0.89-0.98)    | 0.01    |
| Sad/depressed/blue                           | 0.96 (0.90-1.02)    | 0.17    |                     |         |
| Breast tenderness                            | 0.98 (0.94-1.02)    | 0.30    |                     |         |
| Sleep problems                               | 1.01 (0.97-1.06)    | 0.54    |                     |         |
| Difficulty concentrating                     | 1.01 (0.96-1.07)    | 0.86    |                     |         |
| <i>Model criteria</i>                        |                     |         |                     |         |
| c-statistics (AUC) <sup>a</sup>              | 0.84                |         | 0.84                |         |
| c-statistics ( <i>v</i> ) (AUC) <sup>b</sup> | 0.84                |         | 0.84                |         |
| Brier score <sup>c</sup>                     | 0.29                |         | 0.29                |         |
| Brier score ( <i>v</i> ) <sup>b</sup>        | 0.28                |         | 0.28                |         |

<sup>a</sup>The c-statistic, also termed the area under the receiver operating curve (AUC), assesses model discrimination, i.e., how well the model discriminates between severe PMS and not PMS groups. Models with larger c-statistics are preferred.

<sup>b</sup>Results in italics (*v*) are from the validation models.

<sup>c</sup>Brier scores are based on the average squared differences between each outcome and its predicted outcome. Small Brier scores are preferred.

CI, confidence interval.



**FIG. 1.** Comparison of the receiver operator characteristic (ROC) curves for a summary score derived from all 17 symptoms (sum score 17) to one using the 6 core symptoms (sum score 6). As is illustrated in the graph, the discriminate ability of the 6-item score is similar to that obtained using all 17 items,  $p = 0.557$ .

scores directly instead of using weights derived via logistic regression.

Figure 1 depicts the curves for the summary score derived from all 17 symptoms and the summary score derived from the 6 symptoms (the receiver operator characteristic, [ROC]). As Figure 1 shows, the discriminate ability of the 6-symptom score is nearly identical to that obtained using 17 symptoms ( $p = 0.557$ ).

#### Cutoff points to predict PMS

Table 3 shows cutoff points in DSR scores that can be used to discriminate PMS in prospective symptom ratings. We reviewed the full range of the total premenstrual DSR scores and show selected cutoff points for the greatest number of cases correctly classified. Supporting the hypothesis of the study, the results show that classifications for 6 symptoms consistently perform as well as the classifications for 17

symptoms. We have used the cutoff point of 80 in the 17-symptom model in many PMS clinical trials.<sup>12-14,16</sup> The present study shows that this cutoff point correctly classifies 86% of the cases, whereas 84% of the cases are correctly classified with a cutoff point of 33 in the 6-symptom model. Selecting a classification cutoff point balances the specificity and sensitivity of the measure and varies with the objectives of its use (higher specificity may underidentify PMS, and higher sensitivity may overidentify PMS). The cutoff points provided in Table 3 are in the approximate midrange of specificity for this measure.

#### Validation of core symptoms

The models created in the developmental dataset were fit to the validation dataset. The results indicate that these models are replicable and not simply overfitting the idiosyncrasies of these particular data. The nearly identical AUC and Brier scores in the two datasets indicate there was no degradation in discrimination of PMS. (Validation results are shown in Table 2 in italics [*v*]).

#### Associations of symptoms with functional impairment

Approximately 95% of all participants reported some level of impairment in at least one of the assessed domains. PMS was reported to interfere most often in the family domain, followed by the social and work domains. Ratings of severe impairment (ratings of 3 or 4) were reported by 56% for the family domain, 51% for the social domain, and 47% for the work domain among all participants in the developmental dataset (Table 1). Estimations of the association of symptoms with impairment indicated that the likelihood of severe impairment was increased with symptoms approximately 2% with each unit increase in the summed 17-symptom scores (OR 1.02, CI 1.01-1.02,  $p < 0.001$ ) and approximately 3% with each unit increase in the summed 6-symptom model (OR 1.03, 95% CI 1.03-1.04,  $p < 0.001$ ).

#### Discussion

Three major findings deserve emphasis. First, 6 symptoms discriminated PMS as well as 17 symptoms rated in daily diaries by women who were seeking treatment for PMS. These parsimonious results clearly indicate that a small number of prospectively rated symptoms can identify the likelihood of PMS in women who seek treatment. With >90%

**TABLE 3. CUTOFF POINTS IN TOTAL PREMENSTRUAL DAILY SYMPTOM REPORT SCORES FOR PREDICTION OF SEVERE PREMENSTRUAL SYNDROME**

| DSR score | 17 symptoms              |                          |                      | DSR score | 6 symptoms   |              |                      |
|-----------|--------------------------|--------------------------|----------------------|-----------|--------------|--------------|----------------------|
|           | Sensitivity <sup>a</sup> | Specificity <sup>b</sup> | Correctly classified |           | Sensitivity  | Specificity  | Correctly classified |
| 60        | 100%                     | 42.5%                    | 80.2%                | 25        | 97.8%        | 49.5%        | 81.2%                |
| 70        | 100%                     | 52.7%                    | 83.7%                | 30        | 94.9%        | 58.1%        | 82.3%                |
| 80        | <b>100%</b> <sup>c</sup> | <b>59.1%</b>             | <b>86.0%</b>         | <b>33</b> | <b>94.1%</b> | <b>63.4%</b> | <b>83.6%</b>         |
| 120       | 75.5%                    | 77.4%                    | 76.2 %               | 40        | 84.2%        | 71.5%        | 80.0%                |
| 140       | 63.4%                    | 81.7%                    | 69.7%                | 50        | 67.9%        | 80.7%        | 72.3%                |

<sup>a</sup>Sensitivity maximizes the number correctly classified.

<sup>b</sup>Specificity minimizes the number falsely classified.

<sup>c</sup>The bold cutoff points identify the largest number correctly classified in each model.

power to detect the core symptoms, the likelihood that the excluded symptoms were false negative (type II error) results is low. Identifying a small group of symptoms that discriminate PMS among women seeking treatment also supports the concept that a clinical diagnosis for PMS might be developed around a core symptom group.

Second, the symptoms that discriminated PMS represent three domains that are widely believed to describe the syndrome and provide further evidence that PMS is a multifactorial disorder that encompasses both emotional and physical symptoms. The strongest independent predictors of PMS were mood swings and anxiety/tension, both in the emotional symptom domain and long considered predominant PMS symptoms.<sup>23–25</sup> A recent community-based survey of premenstrual symptoms reported by women in Europe and Latin America also reported that mood swings was one of the most prevalent and severe symptoms and the leading emotional symptom experienced by these women.<sup>23</sup> In the behavioral domain, decreased interest in activities and appetite changes/food cravings were independent predictors of PMS; aches and cramps were the independent predictors of PMS in the physical domain.

Although irritability was among the most frequently reported and severe symptoms in this study, it was not a discriminator of PMS. Irritability has long been considered a cardinal PMS symptom and is among the most responsive symptoms to selective serotonin reuptake inhibitor (SSRI) treatment for the disorder.<sup>23,24,26</sup> Nearly all women reported irritability, however, and it was also highly correlated with half of the other PMS symptoms, which precluded its ability to discriminate PMS. These findings clearly show that irritability is likely to be part of the condition, but it is a common symptom that does not specifically define PMS.

The depression symptom also did not discriminate PMS in these data. We further examined depression in a 7-symptom model, but its addition did not alter the results of the 6-symptom model and did not add to the prediction of PMS. This suggests that depressed mood does not have a primary role in pure PMS. Depressive symptoms are among the core symptoms of PMDD as listed in the DSM-IV, although whether depressed mood is a core component of PMDD is also not clearly demonstrated.

Although the present study was not designed to compare PMS and PMDD, observations indicated a large overlap in the daily symptom ratings: 47% (255 of 541) of the PMS group also met criteria for PMDD as defined in a single menstrual cycle. As expected, the total symptom scores were higher in the PMDD group (more symptoms were required). Mood swings and decreased interest were even more predominant in the PMDD group, whereas aches were slightly higher in the PMS group. Other studies are needed to determine if there are differences in the primary symptoms of PMS and PMDD.

Third, nearly all participants reported some level of impairment in at least one domain when they sought treatment, and impairment ratings were strongly associated with the DSR scores. A previous study that evaluated criteria to identify PMS indicated that the respondents experienced reduced work productivity and quality of life regardless of the severity criteria of PMS.<sup>5</sup> The present findings demonstrate the increase in impairment with increasing symptom severity and add further support to the evidence that PMS is strongly

associated with diminished functioning in relationships and the normal activities of daily life.<sup>5,27</sup>

The sensitivity analysis showed that use of 6 symptoms classified the PMS cases nearly identically to use of 17 symptoms and further supported the hypothesis that a small number of symptoms rated daily can discriminate PMS as well as a longer symptom list. The classification results provide a balance between specificity and sensitivity and allow the clinician or researcher to select more or less restrictive cutoff points that best meet the intended objectives in their use. The classifications in the present study appear to be strong, although we know of no other studies that have identified the specificity and sensitivity of symptom scores to discriminate either PMS or PMDD. Whether correct classifications can be increased to an even higher level than the 84%–86% achieved in this study remains an important question.

Several other limitations can be considered. There is no demonstrated gold standard definition of PMS. The criteria that were used to define PMS in this study have previously demonstrated reliability and validity, but the use of other criteria might yield different results.<sup>5</sup> The data represent generally healthy women who seek medical treatment for premenstrual symptoms and may not encompass the entire heterogeneous PMS population, particularly women with other physical or psychiatric disorders that commonly have premenstrual exacerbations. The study identified a parsimonious number of symptoms that discriminate between PMS and not PMS that can be used to evaluate women who believe they have the disorder. However, the study was not designed to provide a validated daily diary or a diagnosis of PMS, and further studies that address these objectives are needed. It is also remains for other studies to identify the associations of the identified core symptoms with treatment response, which is the essential measure of their clinical utility.

The strengths of the study include prospective DSRs that were completed before any treatment interventions, appropriate statistical power, and rigorous analysis that demonstrated notable consistency in the findings and evidence that the results are not idiosyncratic to the study sample. The findings indicate that 6 symptoms can discriminate PMS as well as 17 symptoms when prospectively rated in daily diaries to confirm PMS. The findings suggest that the burden of daily diaries could be reduced by using a smaller number of symptoms and also suggest that a clinical diagnosis for PMS might be developed around a core symptom group. Further studies are needed to construct and validate a brief daily diary and to develop the criteria for a widely accepted diagnosis of PMS.

### Acknowledgments

This work was supported by grants from the National Institutes of Health; the Eunice Kennedy Shriver National Institute of Child Health and Human Development: RO1 HD018633, and the National Cancer Institute: T32 CA93283.

### Disclosure Statement

During the last 3 years, E.W.F. received research support (issued to the University of Pennsylvania) from Wyeth, Xanodyne Pharmaceuticals, and Forest Research Institute and honoraria from or served as a consultant to Wyeth, Pherin

Pharmaceuticals, Bayer Health Care, and Forest Research Institute. During the past 12 months, K.R. received honoraria and served as a consultant or on an advisory board to Pfizer, Inc., and received research grants (issued to the University of Pennsylvania) from Bristol-Myers Squibb, Epix Pharmaceuticals, Renaissance Pharmaceuticals, Inc. (PgxHealth LLC), Pamlab, Pfizer, Inc., and Wyeth. M.D.S., S.M.H., J.M.L., and H.L. have no competing financial interests.

## References

- Potter J, Bouyer J, Trussell J, Moreau C. Premenstrual syndrome prevalence and fluctuation over time: Results from a French population-based survey. *J Womens Health* 2009;18:31–39.
- Wittchen HU, Becker E, Lieb R, Krause P. Prevalence, incidence and stability of premenstrual dysphoric disorder in the community. *Psychol Med* 2002;32:119–132.
- Halbreich U, Backstrom T, Eriksson E, et al. Clinical diagnostic criteria for premenstrual syndrome and guidelines for their quantification for research studies. *Gynecol Endocrinol* 2007;23:123–130.
- Halbreich U. Algorithm for treatment of premenstrual syndromes (PMS): Experts' recommendations and limitations. *Gynecol Endocrinol* 2005;20:48–56.
- Dean BB, Borenstein JE, Knight K, Yonkers, K. Evaluating the criteria used for identification of PMS. *J Womens Health* 2006;15:546–555.
- Steiner M, Born L. Advances in the diagnosis and treatment of premenstrual syndrome. *CNS Drugs* 2000;13:287–304.
- Halbreich U, Borenstein J, Pearlstein T, Kahn LS. The prevalence, impairment, impact, and burden of premenstrual dysphoric disorder (PMS/PMDD). *Psychoneuroendocrinology* 2003;28 (Suppl 3):1–23.
- Halbreich U. Menstrually related disorders—Towards interdisciplinary international diagnostic criteria. *Cephalalgia* 1997;17 (Suppl 20):1–4.
- American College of Obstetricians and Gynecologists. Practice Bulletin: Premenstrual syndrome. *Int J Gynecol Obstet* 2001;73:183.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th ed. Washington, DC: American Psychiatric Association, 1994:350.
- World Health Organization. The ICD-10 classification of mental, behavioral and developmental disorders, 10th revision. Geneva, Switzerland:WHO, 1987.
- Freeman EW, Rickels K, Sammel MD, Lin H, Sondheimer SJ. Time to relapse after short- or long-term treatment of severe premenstrual syndrome with sertraline. *Arch Gen Psychiatry* 2009;66:537–544.
- Freeman EW, Rickels K, Sondheimer SJ, Polansky M, Ziao S. Continuous or intermittent dosing with sertraline for patients with severe premenstrual syndrome or premenstrual dysphoric disorder. *Am J Psychiatry* 2004;161:343–351.
- Freeman EW, Rickels K, Sondheimer SJ, Polansky M. Differential response to antidepressants in women with premenstrual syndrome/premenstrual dysphoric disorder: A randomized controlled trial. *Arch Gen Psychiatry* 1999;56:932–939.
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–1379.
- Freeman EW, DeRubeis RJ, Rickels K. Reliability and validity of a daily diary for premenstrual syndrome. *Psychiatry Res* 1996;65:97–106.
- Mortola JF, Girton L, Beck L, Yen SS. Diagnosis of premenstrual syndrome by a simple, prospective, and reliable instrument: The calendar of premenstrual experiences. *Obstet Gynecol* 1990;76:302–307.
- Endicott J, Nee J, Harrison W. Daily Record of Severity of Problems (DRSP): Reliability and validity. *Arch Womens Ment Health* 2006;9:41–49.
- Cohen LS, Miner C, Brown E, Freeman EW, Halbreich U, Sundell K, McCray S. Premenstrual daily fluoxetine for premenstrual dysphoric disorder: A placebo-controlled, clinical trial using computerized diaries. *Obstet Gynecol* 2002;100:435–444.
- Shah NR, Jones JB, Aperi J, Shemtov R, Karne A, Borenstein J. *Obstet Gynecol* 2008;111:1175–1182.
- Harrell FE. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Med* 1996; 15:361–387.
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–781.
- Dennerstein L, Lehert P, Bäckström TC, Heinemann K. Premenstrual symptoms—Severity, duration and typology: An international cross-sectional study. *Menopause Int* 2009;15:120–126.
- Landén M, Eriksson E. How does premenstrual dysphoric disorder relate to depression and anxiety disorders? *Depress Anxiety* 2003;17:122–129.
- Hartlage SA, Arduino KE. Toward the content validity of premenstrual dysphoric disorder: Do anger and irritability more than depressed mood represent treatment-seekers' experiences? *Psychol Rep* 2002;90:189–202.
- Landén M, Nissbrandt H, Allgulander C, Sorvik K, Ysander C, Ericsson E. Placebo-controlled trial comparing intermittent and continuous paroxetine in premenstrual dysphoric disorder. *Neuropsychopharmacology* 2007;32:153–161.
- Dean BB, Borenstein JE. A prospective assessment investigating the relationship between work productivity and impairment with premenstrual syndrome. *J Occup Environ Med* 2004;46:649–656.

Address correspondence to:

*Ellen W. Freeman, Ph.D.*

*Department of Obstetrics/Gynecology  
3701 Market Street, Suite 820 (Mudd)*

*Philadelphia, PA 19104*

*E-mail: freemane@mail.med.upenn.edu*

